Probability Theory

Francisco Moreira Machado

Fall 2023

Contents

1 $\sigma-$fields and Measures

1.1 $\sigma-$fields

Definition. Let Ω be a set. A <u>$\sigma-$field</u> $\mathcal{A}$ is a collection of subsets of Ω ($\mathcal{A} \subset \mathcal{P}(\Omega)$) such that.

1. $\Omega \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, $A^c \in \mathcal{A}$ (Stability by complement)
3. If $(A_n)_{n\geq 1}$ is a sequence of elements of $\mathcal{A}$, then $\cup_{n\geq 1} A_n$ (Stability by countable union).

$(\Omega, \mathcal{A})$ is called a measurable space. Elements of $\mathcal{A}$ are called measurable sets or events.

Example 1.1. Take a set Ω,

- $A_1 = \{\emptyset, \Omega\}$.
- $A_2 = \mathcal{P}(\Omega)$.
- $A_3 = \{A \subset \Omega \colon A \text{ or } A^c \text{ are countable}\}$.
- Exercise $\rightarrow$ $A_4 = \{A \subset \mathbb{N} \colon A \text{ or } A^c \text{ are finite}\}$ is <u>not</u> a $\sigma-$field.

Remark 1.2 (Trivial properties of $\sigma-$fields)**.**

- We can easily derive from 1. and 2. that $\emptyset \in \mathcal{A}$.
- We can also derive from 2. and 3. that $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Now to understand the intuition behind this definition, let us show a possible interpretation in Probability. Ω represents everything that can happen in a model, while elements in $\mathcal{A}$ are the sets an *observer* is able to detect.

Definition 1.3 (Limsup and Liminf)**.** Let $(A_n)_{n\geq 1}$ be events of $(\Omega, \mathcal{A})$. We define

- $\limsup_{n\to\infty} A_n = \bigcap_{N\geq 0} \bigcup_{n\geq N} A_n$.
- $\liminf_{n\to\infty} A_n = \bigcup_{N\geq 0} \bigcap_{n\geq N} A_n$.

Exercise $\rightarrow$ **Remark 1.4.** For $\omega \in \Omega$ we have $\omega \in \limsup_{n\to\infty} A_n \iff \{n \geq 1 \colon \omega \in A_n\}$ is infinite. Moreover $\omega \in \liminf_{n\to\infty} A_n \iff \exists n(\omega)$ s.t. $n \geq n(\omega) \implies \omega \in A_n$.

WARNING: This should <u>not</u> be confused with the usual notion of lim sup and lim inf for sequences of real numbers.

Proposition 1.5. Let $(A_i)_{i\in I}$ be a collection of $\sigma-$fields on Ω (I not necessarily countable). Then, $\bigcap_{i\in I} A_i$ is itself a $\sigma-$field.

Proof. It suffices to check the three properties of $\sigma-$fields.

1. $\Omega \in \mathcal{A}_i \ \forall i \in I$, thus it is in $\bigcap_{i\in I} \mathcal{A}_i$.
2. If $A \bigcap_{i\in I} \mathcal{A}_i$, then $A \in \mathcal{A}_i \ \forall i \in I$, hence $A^c \in \mathcal{A}_i \ \forall i \in I$, hence $A^c \in \bigcap_{i\in I} \mathcal{A}_i$.

Exercise $\rightarrow$ 3. Similar reasoning

1.1.1 Generated $\sigma-$field

Definition 1.6. If $\mathcal{C} \subset \mathcal{P}(\Omega)$ is a collection of subsets of Ω. We define

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{A} \text{ is a } \sigma-\text{field} \\ \mathcal{C}\subset\mathcal{A}}} \mathcal{A}$$

which is called the $\sigma-$field generated by $\mathcal{C}$.

Notice that the generated $\sigma-$field by $\mathcal{C}$ is indeed a $\sigma-$field by proposition 1.2. Moreover, this is an intersection of at least one element, as $\mathcal{P}(\mathcal{C})$ satisfies the conditions.

Finally, this is the **smallest** $\sigma-$field containing $\mathcal{C}$. This construction is particularly useful as it is hard to explicitly construct such a field due to the possible uncountability.

Remark 1.7. If $\mathcal{C}$ is a $\sigma-$field, then $\sigma(\mathcal{C}) = \mathcal{C}$.

Proposition 1.8. If $\mathcal{C} \subset \mathcal{C}'$ then $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$.

Example 1.9 (σ-field)**.** Take $\Omega = \{0,1\}^{\{1,2,\ldots\}} = \{(x_n)_{n\geq 1} \colon x_i \in \{0,1\} \forall i \geq 1\}$ which can model the outcomes of throwing infinitely manay times a coin.

Definition 1.10 (Cylinder Set)**.** We say that asubset of Ω is a **cylinder set** (or, in short, a cylinder) if it is of the form

$$\mathcal{C}_{a_1,\ldots,a_k} = \{(x_n)_{n\geq 1} \colon x_1 = a_1, \ldots, x_k = a_k\}, \text{ with } a_i \in \{0,1\}$$

It represents outcomes where the first k results are fixed.

The cylinder $\sigma-$algebra $\mathcal{C}_{cyl}$ is defined to be the $\sigma - field$ generated by the cylinders.

Example 1.11. $\{(1,1,\ldots)\} \in \mathcal{C}_{cyl}$ because it is the same set as $\bigcap_{n\geq 1} \mathcal{C}_{\underbrace{1,\ldots,1}_{n \text{ times}}}$

Example 1.12. Take $\Omega = \mathbb{R}$ and $\mathcal{A} = \sigma(\{x\}, x \in \mathbb{R})$, one can check that $\mathcal{A} = \{A \subset \mathbb{R} \colon A \text{ or } A^c \text{ is countable}\}$.

Warning In general elements of generated $\sigma-$fields are not "explicit".

Definition 1.13. Borel $\sigma-$field If (E,d) is a metric space (take $E=\mathbb{R}$), the **Borel $\sigma-$field** is $\sigma(\{U : U\subset E, U \text{ open set}\})$. It is denoted by $\mathcal{B}(E)$ or $\mathcal{B}_E$. It is also the $\sigma-$field generated by closed sets.

Example 1.14. for $E=\mathbb{R}$ one can check that

$$\begin{aligned}\mathcal{B}(E) &= \sigma(]a,b[, a<b, a,b\in\mathbb{R})\\ &= \sigma(]-\infty,b[, b\in\mathbb{R})\\ &= \sigma(]-\infty,b), b\in\mathbb{R})\end{aligned}$$

For this, the key property is that any open set of $\mathbb{R}$ is a countable disjoint union of open intervals.

Definition 1.15 (Product $\sigma-$field)**.** Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be two mesurable spaces. The **product $\sigma-$field** $\mathcal{E}\times\mathcal{F}$ is

$$\mathcal{E}\times\mathcal{F} = \sigma(A\times B : A\in\mathcal{E}, B\in\mathcal{F}).$$

It is the smallest $\sigma-$field on $E\times F$ containing elements $A\times B$ with $A\in\mathcal{E}, B\in\mathcal{F}$.

1.2 Measures

Definition 1.16. A measure on a measurable space $(\Omega,\mathcal{A})$ is a function $\mu : \mathcal{A}\to\mathbb{R}_+\cup\{\infty\}$ with

1. $\mu(\emptyset)=0$.
2. If $(A_n)_{n\geq1}$ is a (countable) sequence of pairwise disjoint elements of $\mathcal{A}$, then $\mu\left(\bigcup_{n\geq1}A_n\right)=\sum_{n\geq1}\mu(A_n)$

When $\mu(\Omega)$ is finite, we say that μ is a finite measure. Moreover, when $\mu(\Omega)=1$ we say that μ is a **probability measure**, we usually write $\mathbb{P},\mathbb{Q}$ instead of μ. Then $(\Omega,\mathcal{A},\mu)$ is called a probability space.

Proposition 1.17. Let μ be a measure on $(\Omega,\mathcal{A})$

1. For $A,B\in\mathcal{A}$, if $A\subset B$ then $\mu(B\setminus A)+\mu(A)=\mu(B)$. If $\mu(A)<\infty$, then $\mu(B\setminus A)=\mu(B)-\mu(A)$.
2. If $(A_i)_{i\geq1}$ are measurable and $A_1\subset A_2\dots$ then $\mu(\bigcup_{n\geq1}A_n)=\lim_{n\to\infty}\mu(A_n)$.
3. If $(A_i)_{i\geq1}$ are measurable and $A_1\supset A_2\dots$ and $\mu(A_1)<\infty$ then $\mu(\bigcap_{n\geq1}A_n)=\lim_{n\to\infty}\mu(A_n)$.
4. If $(A_i)_{i\geq1}$ are measurable, $\mu(\bigcup_{n\geq1}A_n)\leq\sum_{n\geq1}^{\infty}\mu(A_n)$.

Proof. 1. Comes from the second property on the definition by taking $A_1 = B \backslash A, A_2 = A, A_3 = \emptyset = A_4 \dots$.

2. Set $B_1 = A_1$ and $B_{i+1} = A_{i+1} \setminus A_i$ for $i \geq 1$, they are pariwise disjoint and $B_1 \cup B_2 \dots B_k = A_k$. Hence $\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} B_n$ thus $\mu\left(\bigcup_{n\geq 1} A_n\right) = \mu\left(\bigcup_{n\geq 1} B_n\right) = \sum_{n\geq 1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) = \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to\infty} \mu(A_n)$.

Exercise $\rightarrow$ 3. **Complementation Trick** apply 2. with $(A_i^c)_{i\geq 1}$

4. Since $B \backslash A \cap B \subset B$, we have $\mu(A \cup B) = \mu(A) + \mu(B \backslash A) \leq \mu(A) + \mu(B)$. Hence by induction $\mu\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mu(A_i)$. As we apply limits we get by 2. $\mu\left(\bigcup_{n\geq 1} A_i\right) \leq \sum_{n\geq 1} A_n$.

Example 1.18 (The Counting Measure)**.** The cardinality on a set E is defined by $Card(B)$ and can be used when E is finite or countable

Example 1.19 (The Dirac Mass)**.** is a measure fo the form δ_a for $a \in \Omega$ defined by $\delta_a(A) = \mathbb{1}_{a\in A}$.

Example 1.20 (Lebesgue Measure)**.** The Lesbegue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ satisfies $\lambda([a, b]) = b - a$ for $a < b$.

Observe that any positive linear combination of measures is a measure on $(\Omega, \mathcal{A})$.

Remark 1.21. Recall $\Omega = \{0, 1\}^{\{1,2,\dots\}}$ and $\mathcal{C}_{a_1,\dots,a_k}$. One can show that there does not exist a *probability measure* μ on $(\Omega, \mathcal{P}(\Omega))$ such that $\mu(\mathcal{C}_{a_1,\dots,a_k}) = 2^{-k}$. This is due to the $\mathcal{P}(\Omega)$ being "too large".

BUT there is one on $(\Omega, \mathcal{C}_{cyc})$.

Notation. μ measure on $(\Omega, \mathcal{A})$

- μ is $\sigma-$finite if $\exists (A_n)_{n\geq 1)}$ sequence of $\mathcal{A}$ such that $\mu(A_n) < \infty$ for all $n \geq 1$ and $\Omega = \bigcup_{n\geq 1} A_n$
- $x \in \Omega$ is an atom if $\mu(\{x\}) > 0$.

If μ has no atoms, we say that μ is non-atomic. If μ is a (weighted) sum of Dirac masses, we say that μ is atomic.

Example 1.22.
- λ (Lebesgue) is atomic
- $\delta_3/3 + 5\delta_{\frac{\sqrt{17}-1}{2}}$ is atomic
- $\lambda + \delta_2$ is neither.

1.3 The Dynkin Lemma

Definition 1.23. Let $\mathcal{D} \subset \mathcal{P}(\Omega)$ be a collection of subsets of Ω. We say that $\mathcal{D}$ is a Dynkin system (or λ-system) if

1. $\Omega \in \mathcal{D}$.
2. If $A \in \mathcal{D}$, then $A^c \in \mathcal{D}$.
3. If $(A_n)_{n\geq 1}$ is a countable sequence in $\mathcal{D}$ of *pairwise disjoint* sets, then $\bigcup_{n\geq 1} A_i \in \mathcal{D}$.

In particular, a $\sigma-$field is a *Dynkin system*, but the converse is false on $\Omega = \{0,1,2,3\}$ take $\mathcal{D} = \{\emptyset, \Omega, \{0,1\}, \{2,3\}, \{0,2\}, \{1,3\}\}$ and check that it is a Dynkin system but not a $\sigma-$field

Lemma 1.24. Assume that $\mathcal{D} \subset \mathcal{P}(\Omega)$ is a Dynkin system. Assume that it is stable by finite intersections, then $\mathcal{D}$ is a $\sigma-$field.

Proof. It suffices to prove the last condition of a $\sigma-$algebra. Let $(A_n)_{n\geq 1}$ be in $\mathcal{A}$ we show that $\bigcup_{n\geq 1} A_n \subset \mathcal{D}$. Let $B_1 = A_1$ and for $j \geq 2$ set $B_j = A_j \setminus (A_1 \cup \ldots A_{j-1})$. By construction $B_1 \cup \ldots \cup B_j = A_1 \cup \ldots \cup A_j$ and the (B_j) are disjoint. We show by strong induction that $\forall j \geq 1$, $B_j \in \mathcal{D}$.

It is direct for $j = 1$, and now if we assume $B_1, \ldots, B_j \in \mathcal{D}$ then

$$\begin{aligned} B_{j+1} &= A_{j+1} \setminus (A_1 \cup \ldots A_j) \\ &= A_{j+1} \setminus (B_1 \cup \ldots B_j) \\ &= A_{j+1} \cap (\Omega \setminus (B_1 \cup \ldots B_j)) \in \mathcal{D} \end{aligned}$$

as $\mathcal{D}$ is closed under interscetion. Moreover, as each $B_j \in \mathcal{D}$, we have that their union also does, finishing the proof.

We sat that **Dynkin system** stable by finite intersecions is a $\sigma-$field.

Exercise $\rightarrow$ As for $\sigma-$fields, one can show that any intersections of Dynkin systems is a Dynkin system. This allows us to define

Definition 1.25. If $\mathcal{C} \subset \mathcal{P}(\Omega)$ is a class of subsets of Ω, we set

$$\lambda(\mathcal{C}) = \bigcap_{\substack{\mathcal{A} \text{ Dynkin Sys} \\ \mathcal{C} \subset \mathcal{A}}} \mathcal{A}$$

It is called the Dynkin system generated by $\mathcal{C}$.

Theorem (Dynkin Lemma)

Let Ω be a set. Let $\mathcal{C} \subset \mathcal{P}(\Omega)$ be a class of subsets of Ω. Assume that $\mathcal{C}$ is stable by finite intersections then

$$\lambda(\mathcal{C}) = \sigma(\mathcal{C}).$$

In words, the Dynkin system generated by $\mathcal{C}$ is equal to the $\sigma-$field generated by $\mathcal{C}$.

Proof. By double inclusion.

First, since $\sigma(\mathcal{C})$ is a Dynkin system, it must hold that $\lambda(\mathcal{C}) \subset \sigma(\mathcal{C})$.

To show that $\sigma(\mathcal{C}) \subset \lambda(\mathcal{C})$ we show that $\lambda(\mathcal{C})$ is stable under finite intersections. Indeed, then it would hold that $\lambda(\mathcal{C})$ is a $\sigma-$field, but $\sigma(\mathcal{C})$ is the smallest one containing all others, which would finish the proof.

Goal: $\forall A \in \lambda(\mathcal{C}), \forall B \in \lambda(\mathcal{C}), A \cap B \in \lambda(\mathcal{C})$.

First: Fix $A \in \mathcal{C}$. We show that $\forall B \in \lambda(\mathcal{C})$ it holds that $A \cap B \in \lambda(\mathcal{C})$.
Idea: Define $\lambda_A = \{B \subset \Omega \colon A \cap B \in \lambda(\mathcal{C})\}$
Goal: $\lambda(\mathcal{C}) \subset \lambda_A$. We show that λ_A is a Dynkin system containing $\mathcal{C}$, which will imply the desired goal.

- $\mathcal{C} \in \lambda_A$: If $B \in \mathcal{C}$, we have $A \cap B \in \lambda(\mathcal{C})$ due to stability under finite intersection.
- Dynkin system
  - $\Omega \in \lambda_A$ as $A \cap \Omega = A \in \mathcal{C} = \lambda(\mathcal{C})$
  - Take $B \in \lambda_A$, then $B^c \in \lambda_A$ iff $A \cap B^c = \Omega \setminus ((A \cap B) \cup A^c)$. Moreover, $A \in \mathcal{C}$, so $A^c \in \lambda(\mathcal{C})$ and $A \cap B \in \lambda(\mathcal{C})$ and they are disjoint sets, hence their union must be part of the Dynkin system, after which we conclude by stability under complementation.
  - Take $(B_n)_{n\geq 1}$ pairwise disjoint sequence in λ_A. Then $\left(\bigcup_{n\geq 1} B_n\right) \cap A = \bigcup_{n\geq 1} B_n \cap A$, but the elements of this union are pairwise disjoint in $\lambda(\mathcal{C})$. Hence their union must be in $\lambda(\mathcal{C})$ because it is a Dynkin system.

We then conclude $\lambda(\mathcal{C}) \subset \lambda_A$ and so $\forall A \in \mathcal{C}$, $\forall B \in \lambda(\mathcal{C})$, $A \cap B \in \lambda(\mathcal{C})$.

Exercise $\rightarrow$ Second: Now we fix $A \in \lambda(\mathcal{C})$ and check that λ_A and check that λ_A is a Dynkin system containing $\mathcal{C}$. Then $\lambda(\mathcal{C}) \subset \lambda_A$ and we get $\forall A \in \lambda(\mathcal{C})$, $\forall B \in \lambda(\mathcal{C}), A \cap B \in \lambda(\mathcal{C})$

In life, Dynkin lemma is often used as follows:
If $\mathcal{D}$ is a Dynkin system containing a collection $\mathcal{C}$, stable by finite intersection, then $\sigma(\mathcal{C}) \subset \mathcal{D}$. (Notice that if $\mathcal{D}$ is a $\sigma-$field, $\mathcal{C} \subset \mathcal{D} \implies \sigma(\mathcal{C}) \subset \mathcal{D}$). Indeed, by the Dynkin Lemma, $\sigma(\mathcal{C}) = \lambda(\mathcal{C}) \subset \lambda(\mathcal{D})$. This justifies the following definition:

Definition 1.26. Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathcal{C} \subset \mathcal{A}$ a collection of measurable sets. We say that $\mathcal{C}$ is a $\pi-$system if it is stable by finite intersections.
We say that $\mathcal{C}$ is a generating $\pi-$system if $\sigma(\mathcal{C}) = \mathcal{A}$.

Example 1.27. $\{(-\infty, a) \colon a \in \mathbb{R}\}$ is generating of $\mathbb{B}(\mathbb{R})$.

Example 1.28. For $\Omega = \{0,1\}^{\mathbb{N}}$ cylinder sets are generating $\pi-$system of the cylinder $\sigma-$field.

Corollary 1.29. Let $(\Omega, \mathcal{A})$ be a measurable space, $\mathcal{C}$ a generating $\pi-$system.

1. Let μ, ν be two finite measures on $(\Omega, \mathcal{A})$ such that $\mu(\Omega) = \nu(\Omega)$ and $\forall A \in \mathcal{C}, \mu(A) = \nu(A)$, then $\mu = \nu$.

2. More generally, if there exists subsets $E_n \in \mathcal{A}$ such that $\mu(E_n) = \nu(E_n) < \infty$ $\forall n \geq 1$and $\mu(E_n \cap A) = \nu(E_n \cap A)$ $\forall A \in \mathcal{C}$ and $\bigcup E_n = \Omega$, then $\mu = \nu$

Example 1.30 (Application to Lebesgue)**.** There is at most one measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ such that $\lambda([a,b]) = b - a \forall a < b$. This comes from 2. above with $E_n = [-n, n]$.

Probability measures are thus characterized by their values on a generating $\pi-$system.

Exercise $\rightarrow$

Proof (Corollary)**.** We show 1. and leave 2. for exercise.
Goal: $\mu(A) = \nu(A) \forall A \in \mathcal{A}$.
To do that, take

$$\mathcal{G} = \{A \in \mathcal{A} \colon \mu(A) = \nu(A)\}.$$

We check (exercise) that $\mathcal{G}$ is a Dynkin system containing $\mathcal{C}$, generating $\pi-$system, therefore $\mathcal{A} \subset \mathcal{G}$ hence $\forall A \in \mathcal{A}$, $\mu(A) = \nu(A)$.

1.4 Independence

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Two events A, B are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Interpretation: If $\mathbb{P}(B) > 0$, this is equivalent to $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$, which intuitively means that B does not influence the likelihood of A hapenning.

Example 1.31. Throw two dice at random $\Omega = [6]^2, \mathbb{P}(\{\omega\}) = 1/36 \forall \omega \in \Omega$, then $A = \{6\} \times [6]$ and $B = [6] \times \{6\}$ are independent.

Example 1.32. Throw one die $\Omega = [6]$ with even probabilities. Then $A = \{1, 2\}$ and $B = \{1, 3, 5\}$ are independent.

Definition 1.33. Events $A_1, \ldots, A_n$ are mutually independent if for every non-empty subset $\{j_1, j_2, \ldots, j_k\}$ of $[n]$ we have

$$\mathbb{P}\left(A_{j_1} \cap \ldots A_{j_k}\right) = \mathbb{P}(A_{j_1}) \ldots \mathbb{P}(A_{j_k}).$$

Notation. $(A_i)_{i \in [n]}$ are $\perp\!\!\!\perp$.

Remark 1.34. Independence is relative to $\mathbb{P}$. Moreover in general pairwise independence does not imply independence.

Proposition 1.35. Events $A_1, \dots, A_n$ are $\perp\!\!\!\perp$ iff $\mathbb{P}(B_1 \cap \dots, B_n) = \mathbb{P}(B_1)\dots\mathbb{P}(A_n)$, where $B_i \in \sigma(\{A_i\}) = \{\emptyset, A_i, A_i^c, \Omega\}$.

This naturally leads to the notion of independent $\sigma-$fields, which is the "good" setting to define independence.

Definition 1.36. Let $\mathcal{B}_1, \dots, \mathcal{B}_n \subset \mathcal{A}$ be $\sigma-$fields. They are independent ($\perp\!\!\!\perp$) if $\forall B_1 \in \mathcal{B}_1, \dots, \forall B_n \in \mathcal{B}_n$,

$$\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1)\dots\mathbb{P}(B_n).$$

By the proposition just above, a set of events are $\perp\!\!\!\perp$ iff $\sigma-$fields are $\perp\!\!\!\perp$.

To show independence, the following result is very useful:

Proposition 1.37. Let $\mathcal{B}_1, \dots, \mathcal{B}_n \subset \mathcal{A}$ be $\sigma-$fields. For $1 \leq i \leq n$, let $\mathcal{C}_i$ be a generating $\pi-$system of $\mathcal{B}_i$ such that $\Omega \in \mathcal{C}_i$, then

$$\mathcal{B}_1, \dots, \mathcal{B}_n \perp\!\!\!\perp \iff \forall C_1 \in \mathcal{C}, \dots, C_n \in \mathcal{C}_n, \mathbb{P}(C_1 \cap \dots \cap C_n) = \mathbb{P}(C_1)\dots\mathbb{P}(C_n).$$

Proof. The proof is based on Dynkin lemma. See the exercise sheet.

Application 1.38 (Coalition Principle)**.** Let $\mathcal{B}_1, \dots, \mathcal{B}_n \subset \mathcal{A}$ independent $\sigma-$fields. Fix $1 \leq n_1 < n_2 \dots \leq n_p = n$, then $\mathcal{D}_1 = \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1})$, $\mathcal{D}_{i+1} = \sigma(\mathcal{B}_{n_i+1}, \dots, \mathcal{B}_{n_{i+1}})$ for $i < p$ are all $\perp\!\!\!\perp$.

Proof. Find a nice generating $\pi-$system of $\mathcal{D}_1, \dots, \mathcal{D}_p$.

Claim. $\mathcal{C}_1 = \{B_1 \cap \dots \cap B_{n_1} : B_1 \in \mathcal{B}_1, \dots, B_{n_1} \in \mathcal{B}_{n_1}\}$ is a generating $\pi-$system of $\mathcal{D}_1$.

Indeed, we show that $\sigma(\mathcal{C}_1) = \sigma(\mathcal{B}, \dots, \mathcal{B}_{n_1})$ by double inclusion.

First, all elements of $\mathcal{C}_1$ are a finite intersection of $\mathcal{B}_i$, therefore $\mathcal{C}_1 \subset \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1})$, which gives us $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1})$. Moreover, each $\mathcal{B}_i \subset \mathcal{C}_1$ for $1 \leq i \leq n_1$, hence $\sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1}) \subset \sigma(\mathcal{C})$

$\mathcal{C}_1$ is clearly stable by finite intersections, so it is a generating $\pi-$system for $\mathcal{D}_1$. Similarly, we can construct $\mathcal{C}_j$ for $1 < j \leq p$ which is a generating $\pi-$system of $\mathcal{D}_j$.

Then by definition of the $\mathcal{C}_j$'s and by assumption $\forall C_1 \in \mathcal{C}_1, \dots, \forall C_p \in \mathcal{C}_p$,

$$\mathbb{P}(C_1 \cap \dots \cap C_p) = \mathbb{P}(C_1)\dots\mathbb{P}(C_p),$$

as we can split any C_i into an intersection of $B_{n_{i-1}+1}, \dots B_{n_i}$ with $B_j \in \mathcal{B}_j$ for $n_{i-1}+1 \leq j \leq n_i$.

Definition 1.39 (Independence of ANY family of $\sigma-$fields)**.** Let $(\mathcal{B}_i)_{i \in I}$ be a family of $\sigma-$fields. They are independent if any finite collection is independent.

The follow result is VERY useful to show that events have probability 0 or 1.

Lemma 1.40 (Borel-Cantelli)**.** There are two lemmas:

1. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$
2. if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $(A_n)_{n\geq 1}$ are $\perp\!\!\!\perp$ then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Now we can interpret the previous lemma.

We can read 1. as almost surely A_n only happens a finite number of times.

We can read 2. as almost surely A_n happens infinitely often.

Proof. We saw that $\limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_{n\to\infty} A_n)$

Let us start with 2. Fix $l \geq 1$, $n \geq l$, write

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k=l}^{n} A_k^c\right) &= \prod_{k=l}^{n} \mathbb{P}(A_k^c) = \prod_{k=l}^{n} (1 - \mathbb{P}(A_k)) \\
&= \exp\left(\sum_{k=l}^{n} \ln(1 - \mathbb{P}(A_k))\right) \\
&\leq \exp\left(-\sum_{k=l}^{n} \mathbb{P}(A_k)\right) \underset{n\to\infty}{\to} 0.
\end{aligned}$$

Notice that $\bigcap_{k=l}^{n} A_k^c$ is decreasing in n, hence $\mathbb{P}(\bigcap_{k=l}^{\infty} A_k^c) = 0$. This gives us that $\mathbb{P}(\liminf_{n\to\infty} A_n^c) = 0$, which is equivalent to what we wanted to prove.

Now we can go for 1. Fix $n \geq 0$.

Since $\limsup_{n\to\infty} A_n \subset \bigcup_{m\geq n} A_m$. Hence

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \leq \mathbb{P}\left(\bigcup_{m\geq n} A_m\right) \leq \sum_{m\geq n} \mathbb{P}(A_m) \underset{n\to\infty}{\to} 0.$$

2 Random Variables

2.1 Measurable Function

Definition 2.1. Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A function $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ is said to be measurable if $\forall B \in \mathcal{F}$, $f^{-1}(B) = \{x \in E \colon f(x) \in B\} \in \mathcal{E}$.

Interpretation in Probability: A measurable function $X : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ is called a random variable. Intuitively this means that $X(w)$ is "observable" in the sense that one can "observe" wheter $X(w) \in B$ for $B \in \mathcal{F}$.

Exercise $\to$

Proposition 2.2. To check that $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ is measurable, one often finds a class $\mathcal{C} \subset \mathcal{F}$ such that $\sigma(\mathcal{C}) = \mathcal{F}$ and $\forall B \in \mathcal{C}, f^{-1}(B) \in \mathcal{E}$. Indeed, $\{B \in \mathcal{F} \colon f^{-1}(B) \in \mathcal{E}\}$ is then a $\sigma-$field , containing $\mathcal{C}$ thus $\sigma(\mathcal{C})$.

Exercise $\to$

Definition 2.3 (Image Measure)**.** Let $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ be a measurable function and μ a measure on $(E, \mathcal{E})$ on $(E, \mathcal{E})$, then $\forall B \in \mathcal{F}$, $\mu_f(B) = \mu(f^{-1}(B))$ defines a measure on $(F, \mathcal{F})$ called the *image measure* of μ by f. (exercise: check that it is a measure)

In probability, if $X : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ is a random variable and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{A})$, then $\mathbb{P}_X$, the image measure of $\mathbb{P}$ by X, is called the law of X.

Remark 2.4. If $(E, \mathcal{E}, \mu)$ is a probability space, there exists a random variable with law μ. Indeed just take $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$. Therefore, it makes sense to take a random variable following a prescribed law, such as the Normal Distribution.

If X and Y are two r.v., how ca we check if they have the same law, i.e. if $\mathbb{P}_X = \mathbb{P}_Y$? How can one characterize a probability measure.

Nice Case E is countable. Indeed if $X \colon (\Omega, \mathcal{A}) \to \mathcal{P}(E)$ is a r.v. with E countable, its law is characterized by the values

$$\mathbb{P}_X(x) = \mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) \text{ with } x \in E$$

with this, for $A \subset E$, $\mathbb{P}_X(A) = \sum_{x \in A} \mathbb{P}(\{x\})$. In particular, $\mathbb{P}(X = z) = \mathbb{P}(Y = z)\ \forall z \in E$ implies $\mathbb{P}_X = \mathbb{P}_Y$.

When $E = \mathbb{R}$, cumulative distribution functions (cdf) are useful.

Definition 2.5 (cdf)**.** If $X \colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a r.v., its cdf is the function $F_X \colon \mathbb{R}[0, 1]$ defined by

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega \colon X(\Omega) \leq x\}) = \mathbb{P}_X(] - \infty, x]).$$

Example 2.6 (Bernoulli Distribution). Bernoulli random variable $\mathbb{P}(X = 0) = 1/4$, $\mathbb{P}(X = 1) = 3/4$.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

Example 2.7 (Uniform Distribution). Assume that the law of X is the Lebesgue measure on $[0, 1]$

Proposition 2.8. The following characterize a random variable.

1. Let X be a $\mathbb{R}-$valued r.v. Then F_X is non-decreasing, $\lim_{-\infty} F_X = 0$, $\lim_{\infty} F_X = 1$, F_X is right-continuous
2. If $F_Y = F_X$ then $\mathbb{P}_X = \mathbb{P}_Y$
3. (Lebesgue-Stieltjes) If $F\colon \mathbb{R} \to [0, 1]$ satisfies the properties of 1., then there exiss a $\mathbb{R}-$valued r.v. X s.t. $F_X = F$

Proof. First, it is clear that a cdf must be non-decreasing. Due to that, we know that F_x is monotone and bounded, and thus it has its limits well defined.

We can define $A_n = \bigcap_{k=1}^{n}]-\infty, -k]$, which is a decreasing sequence, thus $\mathbb{P}_X(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}_X(A_n) = \lim_{n\to\infty} F_x(-n)$, from which we can conclude. The other limit is analogous.

Now for right continuity, we define very similar sets $A_n = \bigcap_{k=1}^{n}] - \infty, x + 1/k]$ and proceed similarly.

To prove 2., notice that $\{] - \infty, x] \colon x \in \mathbb{R}\}$ is a generating $\pi-$system of $\mathcal{B}(\mathbb{R})$, thus by the corollary of the Dynkin lemma, if $\mathbb{P}_X, \mathbb{P}_Y$ coincide in this set, they are equal.

Take $\Omega =]0, 1[$ equiped with $\mathcal{A} = \mathcal{B}(]0, 1[)$. For $\omega \in]0, 1[$, and $\mathbb{P} = \lambda$ set $X(\omega) = \inf\{t \in \mathbb{R} \colon F(t) \ge \omega\}$ (called the right-continuous inverse of F).

Ex. $\to$ Then X is measurable and $X(\omega) \le x \iff x \le F(X)$

Then $F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(\omega \le F(X)) = \mathbb{P}(\{\omega \in \Omega \colon \omega \le F(x)\}) = F(x)$

Ex. $\to$ **Remark 2.9.** Similarly, one can show that

$$F_X(x) - F_X(x-) = \mathbb{P}(X = x)$$

In particular, if F_X is continuous, $\mathbb{P}(X = x) = 0\ \forall x \in \mathbb{R}$.

Notation. If $f\colon E \to (F, \mathcal{F})$ is a function we set $\sigma(f) = \{f^{-1}(B) \colon B \in \mathcal{F}\}$. It is a $\sigma-$field

Ex. $\to$ (exercise) called the $\sigma-$field generated by f.

Similarly if $(f_i)_{i\in I}$ is a collection of functions $f_i \colon E \to (F_i, \mathcal{F}_i)$ we define $\sigma(f_i, i \in I) = \sigma(\{f_i^{-1}(B_i) \colon B_i \in \mathcal{F}_i, i \in I\})$ to be the $\sigma-$field generated by $(f_i)_{i\in I}$.

Interpretation in Probability: $\sigma(X)$ represents the "information" / "observable sets" one has access to by looking at the the values of X.

Example 2.10. $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Then $\sigma(f) = \{A \in \mathcal{B}(\mathbb{R})\colon A = -A\}$.

Proposition 2.11.

1. Let $f\colon E \to (F, \mathcal{F})$ be a function. Then $\sigma(f)$ is the smallest $\sigma-$field on E such that f is measurable.

2. Let $(f_i)_{i\in I}$ with $f_i\colon E \to (F_i, \mathcal{F}_i)$ be a collection of functions, then its sigma field is the smallest $\sigma-$field on E such that all the f_i are measurable.

Proof. We check that $f\colon (E, \sigma(f)) \to (F, \mathcal{F})$ is measurable. This is indeed true by definition of $\sigma(f)$. Assume now that $f\colon (E, \mathcal{E}) \to (F, \mathcal{F})$ is measurable. We now show $\sigma(f) \subset \mathcal{E}$. Indeed, since f is measurable, $\forall B \in \mathcal{F}, f^{-1}(B) \in \mathcal{E}$, thus $\sigma(f) \subset \mathcal{E}$.

Ex. $\to$ The second part is left as exercise.

Proposition 2.12. Let E, F be metric spaces. Let $f\colon E \to F$ be continuous, then $f\colon (E, \mathcal{B}(E)) \to (F, \mathcal{B}(F))$ is measurable.

Proof. $\forall O \subset F$ open, we have that $f^{-1}(O)$ is open by continuity of f, thus $f^{-1}(O) \in \mathcal{B}(E)$. Thus for $\mathcal{C} = \{O\colon O \subset F,\ \text{open}\}$, which is a generating system of $\mathcal{B}(F)$, we have that $\forall O \in \mathcal{C}, f^{-1}(O) \in \mathcal{B}(E)$. Thus $\forall B \in \sigma(\mathcal{C}) = \mathcal{B}(F), f^{-1}(B) \in \mathcal{B}(E)$.

2.2 Product $\sigma-$fields and families of functions

Product $\sigma-$fields are needed when considering pairs of random variables, and more generally families of r.v.

Idea: View a collection $(X_i)_{i\in I}$ of random variables as ONE random variable.

Definition 2.13 (Product $\sigma-$field)**.** Let $(E_i, \mathcal{E}_i)_{i\in I}$ be a measurable space. Set $E = \prod_{i\in I} E_i$. An element $x \in E$ is written as $(x_i)_{i\in I}$ for $i \in I$ set $\Pi_i\colon E \to E_i$ is the projection onto the $i-$th coordinate called the canonical projections.

Example 2.14. $E = \{0,1\}^{\mathbb{N}}$, then $\Pi_j\colon E \to \{0,1\}, \Pi_j((x_i)_{i\in I}) = x_j$.

Example 2.15. $E = \prod_{i\in[0,1]} \mathbb{R} = \{f\colon [0,1] \to \mathbb{R}\}$ is the space of functions from $[0,1]$ to $\mathbb{R}$.

Definition 2.16 (Product $\sigma-$field or Cylinder $\sigma-$field)**.** We define $\otimes\mathcal{E}_i = \sigma(\Pi_i : i \in I)$ to be the smallest $\sigma-$field on $\prod_{i\in I} E_i$ for which the canonical projections are measurable.

Definition 2.17 (Cylinder Sets)**.** Sets of the form $\Pi_{i_1}^{-1}(A_1)\cap\ldots\Pi_{i_k}^{-1}(A_k)$ with $i_1,\ldots,i_k \in I$, $A_1 \in \mathcal{E}_{i_1},\ldots,A_k \in \mathcal{E}_{i_k}$ are called cylinders. They are a generating $\pi-$system of $\otimes_{i\in I}\mathcal{E}_i$

Proposition 2.18. If $|I| = n$ then $\otimes_{i=1}^n\mathcal{E} = \sigma(\{A_1 \times \ldots \times A_n : A_i \in (E)_i\}$

Proof. Set $\mathcal{E} = \sigma(A_1 \times \ldots \times A_n : A_i \in \mathcal{E}_i)$. We show that $\mathcal{E}$ is the smallest $\sigma-$field on $E_1 \times \ldots \times E_n$ for which the Π_i's are measurable.

$\Pi_i : (E,\mathcal{E}) \to E_i$ is measurable because for $B \in \mathcal{E}_i$ $\pi_i^{-1}(B) = E_1\times\ldots E_{i-1}\times B\times E_{i+1}\times E_n \in \mathcal{E}$. So Π_i is measurable $\forall i$, then for $A_i \in \mathcal{E}_\rangle$ $A_1\times\ldots A_n = \Pi_1^{-1}(A_1)\cap\ldots\Pi_n^{-1}(A_n) \in \mathcal{E}$ by measurability. Hence $\sigma(\{A_1 \times \ldots \times A_n : A_i \in \mathcal{E}\}$ is in the $\sigma-$field.

Definition 2.19. The product measure on $(\prod_{i\in I} E_i, \bigotimes_{i\in I}\mathcal{E}_i)$, given probability measures μ_i on $(E_i,\mathcal{E}_i)$ is the unique probability measure $\bigotimes_{i\in I}\mu_i$ on $\prod_{i\in I} E_i$ such that

$$\bigotimes_{i\in I}\mu_i\left(\{(x_i)_{i\in I} : x_{i_1\in A_1},\ldots,x_{i_k} \in A_k\}\right) = \mu_{i_1}(A_i)\ldots\mu_{i_k}(A_k).$$

Uniqueness follows from the fact that cylinders generate the product $\sigma-$field.
Existence we admit.

Particular case: If I is finite. If $\mathbb{P}_i$ is a probability measure on E_i, $\mathbb{P}_1\otimes\ldots\otimes\mathbb{P}_n$ is the unique probability measure on $E_1\times\ldots\times E_n$ such that $\mathbb{P}_1\otimes\ldots\otimes\mathbb{P}_n(A_1\times\ldots\times A_n) = \mathbb{P}_1(A_1)\ldots\mathbb{P}_n(A_n)$ for $A_i \in \mathcal{E}_i$.

Example 2.20. The Lebesgue measure on $\mathbb{R}^n$.

Remark 2.21. If $\mathcal{C}_i$ is a generating $\pi-$system of $\mathcal{E}_i$, then $\{A_1 \times \ldots \times \times A_n : A_i \in \mathcal{C}_i\}$ is a generating $\pi-$system of $\otimes\mathcal{E}_i$.

In probability, if one considers several random variables, product spaces naturally appear:

Example 2.22. Let X, Y be real-valued random variables, then

$$\mathbb{P}(XY \leq 1) = \mathbb{P}_{XY}(]-\infty,1]) = \mathbb{P}_{(X,Y)}(\{(x,y) \in \mathbb{R}^2 : xy \leq 1\}).$$

More generally, if $(X_1,\ldots,X_n)$ is a random variable in $(E_1,\ldots,E_n)$ its law $\mathbb{P}_{(X_1,\ldots,X_n)}$ on $E_1 \times \ldots \times E_n$ is characterized by the quantities

$$\mathbb{P}_{(X_1,\ldots,X_n)}(A_1\times\ldots\times A_n) = \mathbb{P}((X_1,\ldots,X_n) \in A_1,\ldots,A_n) = \mathbb{P}(X_1 \in A_1 \text{ and } \ldots \text{ and } X_n \in A_n).$$

Proposition 2.23.

1. Let $(E_1, \mathcal{E}_i)$ be a measurable space. A function $f\colon (\Omega, A) \to (\prod_{i\in I} E_i, \bigotimes_{i\in I} \mathcal{E}_i)$ given by $f(\omega) = (f_i(\omega))_{i\in I}$ is measurable iff all the $\Pi_i \circ f$ are measurable, that is iff $\forall i \in I\ \omega \mapsto f_i(\omega)$ is measurable.
   Probabilistic Interpretation: If $(X_i)_{i\in I}$ are a collection of random variables, then $(X_i)_{i\in I}$ can be viewed as ONE random variable in a product space.

2. If $f, g\colon (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are measurable, then $f+g, f-g, \min(f,g), \max(f,g)$ are measurable.

Proof. First, if f is measurable, then $\Pi_i \circ f$ is measurable as it is a composition of measurable functions.

Indeed, if $g\colon (E, \mathcal{E}) \to (F, \mathcal{F})$ and $H\colon (F, \mathcal{F}) \to (G, \mathcal{G})$are measurable, then $h \circ g$ is measurable because for $B \in \mathcal{G}$, $(h \circ g)^{-1}(B) = g^{-1} \circ h^{-1}(B)$ but $h^{-1}B \in \mathcal{F}$ thus $g^{-1}(h^{-1}(B)) \in \mathcal{E}$.

Now for the other direction, since $\bigotimes_{i\in I} \mathcal{E}_i = \sigma\left(\Pi_i^{-1}(B_i)\colon B_i \in \mathcal{E}_i\right)$, it suffices to check that $f^{-1}(\Pi_i^{-1}(B_i)) = (\Pi_i \circ f)^{-1}(B_i) \in \mathcal{E}$ because $\Pi_i \circ f$ is measurable.

Now for part 2 Set

$$\begin{aligned} P\colon (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) &\to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ (x,y) &\mapsto x+y \end{aligned}$$

which is continuous, thus measurable. Additionally, set

$$\begin{aligned} I\colon (\mathbb{R}, \mathcal{B}(\mathbb{R})) &\to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})) \\ x \mapsto (f(x), g(x)) \end{aligned}$$

But $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ (see exercise sheet).

Thus $f+g$ is measurable as the composition $P \circ I$ of measurable functions. For the other operations the proof is similar.

2.3 Independence of Random Variables

For a function $X\colon \Omega \to (E, \mathcal{E})$, $\sigma(X) = \{X^{-1}(A)\colon A \in \mathcal{E}\}$.

Definition 2.24 ($\perp\!\!\!\perp$ for a finite number of r.v.)**.** Random variables $X_1, \ldots, X_n$ with $X_i\colon \Omega \to E_i$ are $\perp\!\!\!\perp$ if $\sigma(X_1), \ldots, \sigma(X_n)$ are $\perp\!\!\!\perp$.

Remark 2.25. by the definition of $\perp\!\!\!\perp$ of $\sigma-$fields this means $X_1, \dots, X_n$ are $\perp\!\!\!\perp$

$$\begin{aligned}
&\iff \forall B_i \in \sigma(X_i)\mathbb{P}(B_1 \cap \dots B_n) = \mathbb{P}(B_1)\dots\mathbb{P}(B_n)\\
&\iff \forall A_i \in \mathcal{E}_i \ \ \mathbb{P}(X_1^{-1}(A_1) \cap \dots \cap X_n^{-1}(A_n) = \mathbb{P}(X_1^{-1}(A_1))\dots\mathbb{P}(X_n^{-1}(A_n))\\
&\iff \forall A_i \in \mathcal{E}_i \ \ \mathbb{P}(X_1 \in A, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1)\dots\mathbb{P}(X_n \in A_n)\\
&\iff \forall \mathbb{P}_{(X_1,\dots,X_n)}(A_1 \times \dots \times A_n) = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}(A_1 \times \dots \times A_n)\\
&\iff \forall \mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}
\end{aligned}$$

Remark 2.26. To show independence one often shows that

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1)\dots\mathbb{P}(X_n \in A_n),$$

for $A_i \in \mathcal{C}_i$ with $\mathcal{C}_i$ a generating $\pi-$system of $\mathcal{E}_i$ containing Ω.

Corollary 2.27.

1. If $X_1, \dots, X_n$ are $\mathbb{Z}-$valued random variables, they are independent iff $\forall i_1, \dots, i_n \in \mathbb{Z}$ $\mathbb{P}(X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_1 = i_1)\dots\mathbb{P}(X_n = i_n)$

2. If $X_1, \dots, X_n$ are $\mathbb{R}-$valued random variables, then $X_1, \dots, X_n \perp\!\!\!\perp$ iff $\forall x_1, \dots, x_n \in \mathbb{R}$ $\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1)\dots\mathbb{P}(X_n \leq x_n)$

Definition 2.28. Let $X = (X_1, \dots, X_n)$ be a random variable in $E_1 \times \dots \times E_n$. The law of $\mathbb{P}_{X_i}$ of X_i, probability measure on E_i is called a marginal law. The law $\mathbb{P}_{(X_1,\dots,X_n)}$ on $E_1 \times \dots \times E_n$ is called the joint law.

Since $\mathbb{P}_{X_i}(A_i) = \mathbb{P}_{(X_1,\dots,X_n)}(E_1 \times \dots E_{i-1} \times A_i \times E_{i+1} \dots \times E_n)$.

The joint law determines the marginal laws, while the converse is false in general but when $X_1, \dots, X_n \perp\!\!\!\perp$.

Lemma 2.29 (Composition Principle)**.** Let X_i be $\perp\!\!\!\perp$ r.v with $X_i \colon \Omega \to E_i$ let $f_i \colon E_i \to F_i$ be measurable, then $(f_i(X_i))_{1 \leq i \leq n}$ are $\perp\!\!\!\perp$.

Proof. This comes from the fact that $\sigma(f_i(X_i)) \subset \sigma(X_i)$, thus $\forall A_i \in \sigma(f_i(X_i))$ we have $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\dots\mathbb{P}(A_n)$.

Now we show the inclusion of $\sigma-$fields above. Notice that $\sigma(f_i(X_i))$ have elements of the form $(f_i \circ X_i)^{-1}(B)$ with $B \in \mathcal{F}_i$, then as $f_i^{-1}(B) \in \mathcal{E}_i$, we have that $(f_i \circ X_i)^{-1}(B) \in \sigma(X_i)$.

Definition 2.30 (Independence of ANY family of Random Variables)**.** If $(X_i)_{i \in I}$ are r..v with $X_i \colon \Omega \to E_i$, they are independent if for any finite subset of indices J, $(X_j)_{j \in J} \perp\!\!\!\perp$.

Lemma 2.31 (Coalition Principle - Countable Family)**.** Let $(X_i)_{i \geq 1} \perp\!\!\!\perp$ r.v. Fix $p \geq 1$. Set $\mathcal{B}_1 = \sigma(X_1, \dots, X_p)$ and $\mathcal{B}_2 = \sigma(X_{p+1}, X_{p+2}, \dots)$, then $B_1 \perp\!\!\!\perp B_2$.

Proof. We use the fact that if $\mathcal{C}_1, \mathcal{C}_2$ are generating $\pi-$systems of $\mathcal{B}_1, \mathcal{B}_2$ respectively with $\forall A_1 \in \mathcal{C}_1, A_2 \in \mathcal{C}_2$ $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$, then $\mathcal{B}_1 \perp\!\!\!\perp \mathcal{B}_2$.

Ex. $\rightarrow$ Take $\mathcal{C}_1 = \sigma(X_1, \ldots, X_p)$ and $\mathcal{C}_2 = \bigcup_{k=p+1}^{\infty} \sigma(X_{p+1}, \ldots, X_k)$. Check that this works.

Application 2.32. Let X, Y, Z, T be $\perp\!\!\!\perp$ random variables, then $X + Z$ and YT are $\perp\!\!\!\perp$

Proof. Indeed, X, Z, Y, T are $\perp\!\!\!\perp$ ($\perp\!\!\!\perp$ is preserved under permutation). then we apply the Coalition Principle to get that (X, Z) and (Y, T) are independent. Moreover, by the Composition Principle, we have that $f_1(X, Z)$ and $f_2(Y, T)$ are independent if we pick two measurable functions $f_1(x, z) = x + z$ and $f_2(y, t) = yt$.

Lemma 2.33. The two random variables $(X_i)_{i\in I}$ and $(Y_i)_{i\in I}$ with values in $\Pi_{i\in I} E_i$ and $\Pi_{i\in I} F_i$ are $\perp\!\!\!\perp$ iff

$$\forall i_1, \ldots, i_k \in I, \forall j_1, \ldots, j_l \in J, (X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \perp\!\!\!\perp (Y_{j_1}, \ldots, Y_{j_l})$$

Definition 2.34. If $(X_i)_{i\geq 1}$ are random variables we set $B_n = \sigma(X_K : k \geq n)$ and $B_\infty = \bigcap_{n\geq 1} B_n$, which is a $\sigma-$field called the <u>tail $\sigma-$field</u>.

Intuitively B_∞ represents information that does not depend on a finite number of random variables.

Example 2.35. If $(X_i)_{i\geq 1}$ are $\mathbb{R}-$valued rv. Set $S_n = X_1 + \ldots + X_n$ then $\{\sup_{n\geq 1} S_n = +\infty\} \in B_\infty$

Theorem 2.36 (Kolmogorov $0 - 1$ law)
Assume that $(X_i)_{i\geq 1}$ are $\perp\!\!\!\perp$ then $\forall A \in B_\infty, \mathbb{P}(A) = 0$ or 1.

Proof. Set $\mathcal{D}_n = \sigma(X_1, \ldots, X_n)$, then $\mathcal{D}_n \perp\!\!\!\perp B_{n+1}$. Hence $\mathcal{D}_n \perp\!\!\!\perp B_\infty$ because $B_\infty \subset B_{n+1}$. Thus $\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n$, $\forall B \in B_\infty$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. But $\bigcup_{n\geq 1} \mathcal{D}_n = \bigcup_{n\geq 1} \sigma(X_1, \ldots, X_n)$ is a generating $\pi-$system of $\sigma(X_i : i \geq 1)$. Thus

$$\forall A \in \sigma(X_i : i \geq 1), \forall B \in B_\infty, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Finally, observe that $B_\infty \subset \sigma(X_n : n \geq 1)$, thus $\forall A, B \in B_\infty$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, from which we conclude that $\mathbb{P}(A) = \mathbb{P}(A)^2$ $\forall A \in B_\infty$, finishing the proof.

2.4 Real-valued random-variables

Proposition 2.37. Let $f_n : (E, \mathcal{E}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})$ be measurable functions where $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ with $d(x,y) = |\arctan x - \arctan y|$. Then $\sup_n f_n$ i.e. the function $x \mapsto \sup_n f_n(x)$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ are all measurable from $(E, \mathcal{E})$ to $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$

Proof. Let us show for $f = \sup f_n$.

$\sup_{n\geq 1} x_n \leq a \iff \forall n \geq 1, x_n \leq a$. Thus $\forall a \in \mathbb{R}$, $f^{-1}([-\infty, a]) = \bigcap_{n\geq 1} f_n^{-1}([-\infty, a]) \in \mathcal{E}$ because f_n is measurable.

Since $([-\infty, a] : a \in \mathbb{R})$ generates $\mathcal{B}(\overline{R})$, this shows that f is measurable.

Definition 2.38 (Simple Function)**.** A simple function $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function which takes a finite number of values. Equivalently f can be written

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$$

with $a_i \in \mathbb{R}$ and $A_i \in \mathcal{E}$. It can be uniquely written if we suppose A_i are pairwise disjoint and we order the a_i.

Theorem 2.39

Let $f : (E, \mathcal{E}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ be measurable. There exists a sequence (f_n) of simple measurable functions $(E, \mathcal{E}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ such that $\forall x \in E$ the sequence $(f_n(x))_{n\geq 1}$ is weakly increasing and converges to $f(x)$.

This is a powerful tool to show properties for general functions. First we check the property for simple functions then conclude by approximations.

Proof.

Step 1 Approximate the identity function. To do so, just take $\phi_n(x) = \min\left(\frac{1}{2^n} \lfloor 2^n x \rfloor, n\right)$, which only takes finitely many values.

Step 2 Just take $f_n = \phi_n \circ f$.

Application 2.40 (Doob-Dynkin Lemma)**.** Let $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ and $g : (E, \sigma(f)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable. Then $\exists h : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $g = h \circ f$

In Probabaility: A $\sigma(X)-$measurable rv is just a function of X.

Remark 2.41. If $g = h \circ f$ then g is $\sigma(f)-$ measurable since

$$g^{-1}(B) = (h \circ f)^{-1}(B) = f^{-1}(h^{-1}(B)) \in \sigma(f).$$

Proof. Assume $g \geq 0$ by decomposing $g = max(g, 0) + max(-g, 0)$.

Now, consider the case $g = \mathbb{1}_A$ with $A \in \sigma(f)$, then $A = f^{-1}(B)$ with $B \in \mathcal{F}$. we then take $h = \mathbb{1}_B$, from which it follows.

By linearity, the statement holds for any simple function, so now we can conclude by using the fact that we can write g as a limit of simple functions $g_n = h_n \circ f$ and build h to be the limit of h_n, then the desired result holds.

2.5 Integration

The notion of expectation is defined in probability theorey using the Lebesgue integration with respect to a probability theory. We recap the main results. We start with non-negative functions. Let $(E, \mathcal{E}, \mu)$ be a measured space.

2.5.1 Definition of the Integral

Definition 2.42 (Integral for simple functions). If $f\colon E \to [0,\infty]$ is a measurable simple function, $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ with $a_i \in \mathbb{R}_+ \cup \{\infty\}$ and $A_i \in \mathcal{E}$. We define

$$\int_E f d\mu = \sum_{i=1}^n a_i \mu(A_i),$$

with the convention $0 \times \infty = 0$.

One checks that if we write f in another simple function representation, the integral does not change.

Elementary Properties: Let $f, g \geq 0$ be simple functions, then

1. for $a, b \geq 0$ it holds $\int (af + bg) d\mu = a(\int f d\mu) + b(\int g d\mu)$
2. If $f \leq g$ then $\int f d\mu \leq \int g d\mu$

Definition 2.43 (Integral for Positive Valued). Let $f\colon E \to [0,\infty]$ be measurable. We define

$$\int f d\mu = \sup_{\substack{0 \leq h \leq f \\ h \text{ simple}}} \int h d\mu.$$

Definition 2.44 (Expectation). In probability, if $X\colon \Omega \to [0,\infty]$ is a rv. we define

$$\mathbb{E}[X] = \int_\Omega X(\omega) \mathbb{P}(d\omega).$$

Proposition 2.45. If $0 \leq f \leq g \leq \infty$

- $\int f d\mu \leq \int g d\mu$
- If $\mu(\{x \in E\colon f(x) > 0\}) = 0$, then $\int f d\mu = 0$.

2.5.2 Monotone Convergence

Theorem 2.46
Let $f_n \colon E \to [0, \infty]$ be measurable functions such that $(f_n)_{n \geq 1}$ is non-decreasing, that is $\forall x \in E, \forall n \geq 1, f_n(x) \leq f_{n+1}(x)$.
Set $f(x) = \lim_{n \to \infty} f_n(x)$, measurable, then

$$\int f d\mu = \lim_{n \to \infty} \int f_n d\mu.$$

(Notice that the RHS is an increasing sequence)

This is very useful combined with the fact that any ≥ 0 function is the pointwise limit of simple functions.

Theorem 2.47 (Probabilistic Version of Monotone Conv)
If $(X_n)_{n \geq 1}$ is a sequence of random variables such that $X_n \leq X_{n+1}$

$$\mathbb{E}[\lim_{n \to \infty} X_n] = \lim_{n \to \infty} \mathbb{E}[X_n].$$

Corollary 2.48.

1. If $f, g \geq 0$, $a, b \geq 0$, $\int (af + bf) d\mu = a \int f d\mu + b \int g d\mu$
2. If $f_k \geq 0$, $\int (\sum_{k \geq 1} f_k) d\mu = \sum_{k \geq 1} (\int f_k d\mu)$

Sketch. Show it for simple functions and conclude by monotone convergence by passing to the limit.

Example 2.49.

- If we use δ_a, the dirac function for $a \in E$, as the measure, then if $\forall f \colon E \to \mathbb{R}_+$ is measurable,
$$\int_E f d\delta_a = f(a).$$
- If $\#$ is the counting on $\mathbb{N}$ ($\# = \sum_{i=0}^{\infty} \delta_i$). Then for $f \colon \mathbb{N} \to \mathbb{R}_+$ measurable
$$\int f d\# = \sum_{i=0}^{\infty} f(i).$$
- If $f \colon \mathbb{R} \to \mathbb{R}_+$ is Riemann-integrable then its Lebesgue integral coincides.

2.5.3 Fatou's Lemma

Theorem 2.50 (Fatou Lemma)
Let $f_n \geq 0$ be measurable functions then

$$\int (\liminf_{n\to\infty} f_n) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Alternatively in probability

$$\mathbb{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

2.5.4 Markov's Inequality

We say that a property is true almost everywhere if the set of $x \in E$ for which it is not true is negligeable meaning has 0 $\mu-$measure. In probability we say almost surely.

Proposition 2.51. Let $f \geq 0$.

1. $\forall a > 0$, $\mu(\{x \in E \colon f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu$
2. $\int f d\mu < \infty \implies f < \infty$ almost everywhere.
3. $\int f d\mu = 0 \implies f = 0$ almost everywhere.
4. If $g \geq 0$ and $f = g$ almost everywhere, then $\int f d\mu = \int g d\mu$.

Equivalently in probability, if we let $X \geq 0$

1. $\forall a > 0$, $\mathbb{P}(X \geq a) \leq \frac{1}{a}\mathbb{E}[X]$.
2. $\mathbb{E}[X] < \infty \implies x < \infty$ a.s.
3. $\mathbb{E}[X] = 0 \implies x = 0$ a.s.
4. $X = Y$ a.s. $\implies E[X] = E[Y]$.

2.5.5 Fubini's Theorem

Recall that μ is $\sigma-$finite if $E = \bigcup_{n\geq 1} E_n$ with $\mu(E_n) < \infty$ $\forall n \geq 1$.

Informally speaking the Fubini-Tonelli theorem says that for non-negative functions of several variables, when $\mu_1, \ldots, \mu_n$ are $\sigma-$finite , then

$$\int \left(\int \left(\ldots \int f(x_1, \ldots, x_n) \mu_1(dx_1) \ldots \mu_n(dx_n) \ldots \right) \right)$$

can be computed by integrating any order. (see lecture notes for full statement). Typically

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

Theorem 2.52 (Fubini-Tonelli)
Let μ, ν be $\sigma-$finite measures on $(E, \mathcal{E}), (F, \mathcal{F})$ respectively. We equip $E \times F$ with the product sigma field $\mathcal{E} \otimes \mathcal{F}$. Let $f \colon E \times F \to \mathbb{R}_+$ be measurable.

1. $x \mapsto \int f(x,y)\nu(dy)$ and $y \mapsto \int f(x,y)\mu(dx)$ are measurable
2. We have
$$\int_{E\times F} f d\mu \otimes \nu = \int_E \left(\int_F f(x,y)\nu(dy)\right)\mu(dx) = \int_F \left(\int_E f(x,y)\mu(dx)\right)\nu(dy).$$

2.5.6 Real-valued functions

If $f \colon E \to \mathbb{R}$ is measurable, when $\int_E |f| d\mu < \infty$, we say that f is integrable (with respect to μ) and write $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ or $f \in \mathcal{L}^1$ in short.

Similarly, for $p > 0$, when $\int_E |f|^p d\mu < \infty$ we write $f \in \mathcal{L}^p$.

Definition 2.53. Let $f \colon E \to \mathbb{R}$ be measurable when $\int |f| d\mu < \infty$, we write $f = f^+ - f^-$ and define
$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

This is well defined because $0 \leq f^+ < |f|$ and $0 \leq f^- \leq |f|$ so the integrals are less than infinity.

Now, as for non-negative functions, we have the usual properties for $f, g \in \mathcal{L}^1$

- $f \leq g$ a.e. $\implies \int f d\mu \leq \int g d\mu$.
- $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$.
- $f = g$ a.e. $\implies \int f d\mu = \int g d\mu$.
- $\left|\int f d\mu\right| \leq \int |f| d\mu$

Theorem 2.54 (Dominated Convergence)
Let $f_n \colon E \to \mathbb{R}$ be integrable functions such that

1. $\exists f \colon E \to \mathbb{R}$ measurable such that for μ, for almost every x the sequence $f_n(x)$ converges to $f(x)$.
2. $\exists g \colon \to \mathbb{R}_+$ such that $\int g d\mu < \infty$ and $\forall n \geq 1$, for almost every x $|f_n(x)| \leq g(x)$

then
$$\int_E |f_n - f| d\mu \to 0$$
which also gives us $\int f_n d\mu \to \int f d\mu$.

Theorem 2.55 (Dominated Convergence in Probabilistic Setting)
Let X_n be a $\mathbb{R}-$valued r.v.

1. $X_n \to X$ a.s.
2. $\exists Z \geq 0$ such that $E[Z] < \infty$ and $\forall n \geq 1$ $|X_n| \leq Z$ as.

then
$$\mathbb{E}[|X_n - X|] \to 0.$$

There is an extension of Fubini's Theorem to $\mathbb{R}-$valued functions, **Fubini-Lebesgue** Theorem.

In short, one may compute

$$\int \dots \int f(x_1, \dots, x_n)\mu(dx_1)\dots\mu(dx_n)$$

for $\sigma-$finite measures in any order of integration as soon as $\int \dots \int |f(x_1, \dots, x_n)|\mu(dx_1)\dots\mu(dx_n) < \infty$

2.6 Classical Laws

2.6.1 Discrete Laws

Definition 2.56 (Uniform Law). If E is a finite set with n elements, X follows the uniform distribution on E if

$$\mathbb{P}(X = x) = \frac{1}{n}\ \forall x \in E$$

Definition 2.57 (Bernoulli). $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$.
Interpretation Rigged coing giving heads with probability p.

Definition 2.58 (Binomial Law $\mathcal{B}(n, p)$). For $0 \leq k \leq n$ $\mathbb{P}(X = l) = \binom{n}{k}p^k(1-p)^{n-k}$
Interpretation number of heads when tossing the previous coin n-times.

Definition 2.59 (Geometric Law). $\mathbb{P}(X = k) = p(1-p)^{k-1}$ for $k \geq 1$
Intepretation Number of trials before a success having probability p.

Definition 2.60 (Poisson Law of parameter $\lambda > 0$). $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k \geq 0$
Interpretation law of rare events.

Remark 2.61 (Law of Total Probability). Let $(A_i)_{i\geq 1}$ be events such that $A_i \cap A_j = \emptyset$ for $i \neq j$ then $\forall A$ an event, $\mathbb{P}(A) = \sum_{i\geq 1}\mathbb{P}(A \cap A_i)$.
Function Extension: If $Y \geq 0$ is a random variable, $\mathbb{E}[Y] = \sum_{i=1}^{\infty}\mathbb{E}[Y\mathbb{1}_{A_i}]$ (Consequence of Fubini-Tonelli).

2.6.2 Continuous Laws

Definition 2.62. Let $p\colon \mathbb{R} \to \mathbb{R}_+$ be a measurable function such that $\int_{\mathbb{R}} p(x)dx = 1$, then $\forall A \in \mathcal{B}(\mathbb{R})$ the formula:

$$\mu(A) = \int_A p(x)dx = \int_{\mathbb{R}} p(x)\mathbb{1}_A(x)dx$$

defines a probability measure on $\mathbb{R}$.

A random variable having htis law is said to have density p.

Warning: a density is not uniquely defined: it is define uniquely up to 0 Lebesgue measure sets.

Moreover, if X has density p then its **cdf** is

$$\mathbb{P}(X \leq t) = \int_{-\infty}^{t} p(x)dx.$$

One then checks that $\forall f\colon \mathbb{R} \to \mathbb{R}_+$ measurable

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)p(x)dx.$$

Indeed, we can show that it holds for simple functions and then we conclude by an approximation and monotone convergence.

Definition 2.63 (Uniform law)**.** $a < b$, $p(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$.

Definition 2.64 (Exponential law of parameter $\lambda > 0$)**.** $p(x) = \lambda e^{-\lambda x}\mathbb{1}_{x \geq 0}$.

Definition 2.65 (Gaussian Law)**.** For parameters $m \in \mathbb{R}, \sigma > 0$ denoted by $\mathcal{N}(m, \sigma^2)$ has density $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}$

Proposition 2.66. If X has density p, its **cdf** is continuous.

Proof. Set $F(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} p(x)dx$.

Fix $t \in \mathbb{R}, t_n \to t$. We show $F(t_n) \to F(t)$. Now define $f_n(x) = p(x)\mathbb{1}_{(-\infty,t_n]}(x)$. Notice that $\forall x \in \mathbb{R} \setminus \{t\}$, $f_n(x) \to p(x)\mathbb{1}_{(-\infty,t]}(x)$, and $0 \leq f_n(x) \leq p(x)$ which is an integrable function respective to dx.

Therefore, by Dominated Convergence

$$F(t_n) \to \int_{-\infty}^{\infty} p(x)\mathbb{1}_{(-\infty,t]}(x)dx = F(t).$$

Proof. Let us now prove that $\mathbb{E}[f(x)] = \int_{\mathbb{R}} f(x)p(x)dx$.

If $f = \mathbb{1}_A$, $\mathbb{E}[f(X)] = \int_{\Omega} \mathbb{1}_A(X(\omega))\mathbb{P}(d\omega) = \mathbb{P}(X \in A) = \int_{\mathbb{R}} \mathbb{1}_A(x)p(x)dx$. Therefore it holds for simple functions.

Now we can take $0 \leq f_n \leq f$ such that f_n converges pointwise and increasingly to f with f_n simple, then

$$\mathbb{E}[f(X)] \leftarrow \mathbb{E}[f_n(X)] = \int_{\mathbb{R}} f_n(x)p(x)dx \to \int_{\mathbb{R}} f(x)p(x)dx$$

by monotone convergence twice.

Now coming back to **cdf**'s, if F is a function, to see if it's a **cdf** of a random variable X with density, it is sufficient to show that F is piecewise $\mathcal{C}^1$ and $F(t) = \int_{-\infty}^t F'(X)dx$ with $\int_{\mathbb{R}} F'(x)dx = 1$.

Definition 2.67 (Density in $\mathbb{R}^n$)**.** Take $p\colon \mathbb{R}^n \to \mathbb{R}^+$ with $\int_{\mathbb{R}^n} p(x)dx = 1$. $X = (X_1, \ldots, X_n)$ with values in $\mathbb{R}^n$ has density p if

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = \int_A p(x_1, \ldots, x_n)dx_1 \ldots dx_n \forall A \in \mathcal{B}(\mathbb{R}^n).$$

Moreove,r notice that $\forall 1 \leq i \leq n$, X_i has density p_i obtained by integrating p with respect to the other variables

$$p_i(x) = \int_{\mathbb{R}^{n-1}} p(x_1, \ldots, x_i, \ldots, x_n)dx_1 \ldots dx_{i-1}dx_{i+1} \ldots d_n.$$

2.7 Independence and Integration

Theorem 2.68 (Transfer Theorem)

Let $X\colon \Omega \to E$ be a random variable and let $f : E \to \mathbb{R}_+$ measurable. Then

$$\mathbb{E}[f(X)] = \int_E f(x)\mathbb{P}_X(dx).$$

Proof. First, let us prove for $f = \mathbb{1}_A$.

$$\mathbb{E}[f(X)] = \int_{\Omega} \mathbb{1}_A(X(\omega))d\omega = \mathbb{P}_X(A) = \int_E \mathbb{1}_A(x)\mathbb{P}_X(dx).$$

By linearity, the theorem holds for simple functions. Then for $f \geq 0$, take $0 \leq f_n$ converging pointwise to f with f_n simple

$$\mathbb{E}[f(X)] \leftarrow \mathbb{E}[f_n(X)] = \int_E f_n(x)\mathbb{P}_X(dx) \to \int_E f(x)\mathbb{P}_X(dx)$$

by monotone convergence twice.

Remark 2.69. The Transfer Theorem is also valid for $f : E \to \mathbb{R}$ bounded and more generally for $f : E \to \mathbb{R}$ such that $\mathbb{E}[|f(X)|] < \infty$.

Application 2.70. Let U be uniform on $[0,1]$, let us find the law of U^2. For $f : \mathbb{R} \to \mathbb{R}_+$ measurable and $g = f \circ (x \mapsto x^2)$, using the transfer theorem we write

$$\mathbb{E}[f(U^2)] = \int_0^1 g(x)dx = \int_0^1 f(x^2)dx = \int_0^1 f(u)\frac{1}{2\sqrt{u}}du.$$

Indeed, this gives us that a candidate function is $\mathbb{P}_{U^2}(dx) = \frac{1}{2\sqrt{x}}\mathbb{1}_{[0,1]}(x)dx$, but as we can choose any f measurable, this has to be unique.

Takeaway: If we obtain $\mathbb{E}[f(X)] = \int_E f(x)\mu(dx)$ for all $f \geq 0$ measurable, then μ is the law of X.

Example 2.71. If X has density $\frac{\alpha+1}{x^\alpha}\mathbb{1}_{[1,+\infty[}(x)dx$ with $\alpha > 0$, let us find all p such that $\mathbb{E}[X^p] < \infty$.
Indeed by the Transfer Theorem

$$\mathbb{E}[X^p] = \int_{\mathbb{R}} x^p \mathbb{P}_X(dx) = (\alpha+1)\int_1^\infty \frac{1}{x^{\alpha-p}}dx < \infty \iff \alpha - p > 1.$$

Corollary 2.72. If $X, T \colon \Omega \to E$ are random varaibles having the same law, then $\forall f \colon \to \mathbb{R}_+$ measurable,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y)].$$

Theorem 2.73

If $X_1, \dots, X_n$ are $\perp\!\!\!\perp$, with X_i having density p_i, then $(X_1, \dots, X_n)$ has density in $\mathbb{R}^n$ which is $p_1(x_1)\dots p_n(x_n)$.

Proof. We use the dummy function method. We take $f \colon \mathbb{R}^n \to \mathbb{R}_+$ measurable and compute $\mathbb{E}[f(X_1, \dots, X_n)]$.

Due to the Transfer Theorem with $(X_1, \dots, X_n)$ and f we get

$$\begin{aligned}
\mathbb{E}[f(X_1, \dots, X_n)] &= \int_{\mathbb{R}^N} f(x_1, \dots, x_n)\mathbb{P}_{(X_1,\dots,X_n)}(dx_1 dx_2 \dots dx_n) \\
&= \int_{\mathbb{R}^N} f(x_1, \dots, x_n)\mathbb{P}_{X_1}(dx_1) \otimes \dots \otimes \mathbb{P}_{X_n}(dx_n) \text{ by } \perp\!\!\!\perp \\
&= \int_{\mathbb{R}^n} f(x_1, \dots, x_n)p_1(x_1)\dots p_n(x_n) \text{ by Fubini-Tonelli}
\end{aligned}$$

Theorem 2.74

If X, Y are $\perp\!\!\!\perp$ random variables and have densities, then $X + Y$ has a density.

Moreover, if X, Y have densities p, q, respectively, the density of $Z = X + Y$ is given by $z \mapsto \int_{\mathbb{R}} p(x)q(z-x)dx$, called the convolution product of p and q.

Remark 2.75. This theorem does not hold true in general. Take $Y = -X$ for example.

Application 2.76. Let X, Y have densitites and be $\perp\!\!\!\perp$. Then $\mathbb{P}(X = Y) = 0$.

Proof. Let p, q be the densities of X, Y respectively. Notice that

$$\begin{aligned}
\mathbb{P}(X = Y) &= \mathbb{E}[\mathbb{1}_{X=Y}] \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{X=Y}(x,y)\mathbb{P}_{(X,Y)}(dxdy) \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{X=Y}(x,y)p(x)q(y)dxdy \text{ by } \perp\!\!\!\perp \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{X=y}(x,y)p(x)dx \right) q(y)dy \text{ by Fubini-Tonelli} \\
&= \int_{\mathbb{R}} (0)q(y)dy = 0.
\end{aligned}$$

Corollary 2.77. If X has density, then (X, X) does not have a density in $\mathbb{R}^2$.

Ex. $\rightarrow$ Indeed, one can show that if (X, Y) has a density in $\mathbb{R}^2$, then $\mathbb{P}(X = Y) = 0$

Theorem 2.78

The following are equivalent for $X_i : \Omega \to E_i$ random variables

1. $X_1, \dots, X_n$ are $\perp\!\!\!\perp$
2. $\forall f_i : E_i \to \mathbb{R}_+$ measurable

$$\mathbb{E}\left[f_1(X_1) \dots f_n(X_n)\right] = \mathbb{E}[f_1(X_1)] \dots \mathbb{E}[f_n(X_n)].$$

In practice, to show that $X \perp\!\!\!\perp Y$ one often computes $\mathbb{E}[f(X)g(Y)]$ and checks the previous statement.

Corollary 2.79. If $(X_1, \dots, X_n)$ has a density of the form $g_1(x_1) \dots g_n(x_n)$, then $X_1, \dots, X_n$ are $\perp\!\!\!\perp$.

If $X_1, \dots, X_n$ are $\perp\!\!\!\perp$ and $f_i : E_i \to \mathbb{R}$ the equality

$$\mathbb{E}[f_1(X_1) \dots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \dots \mathbb{E}[f_n(X_n)]$$

is true under the integrability conditions $\mathbb{E}[|f_i(X_i)|] < \infty$ for all $i \leq n$. This implies in particular that $f_1(X_1) \dots f_n(X_n)$ is integrable.

Application 2.80.

1. Let X be a L^2 random variable. Then $X \in L^1$ and we can define the variance $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
2. (Cauchy-Schwarz) If $X \in L^2$ then $\mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2]$
3. Let $(X_i)_{1 \leq i \leq n}$ be $\perp\!\!\!\perp, L^2$ random variables, then $Var(X_1 + \ldots X_n) = Var(X_1) + \ldots + Var(X_n)$.

3 Sequences and Series of Independent Random Variables

Goal Study limits of $X_1 + \ldots + X_n$ as $n \to \infty$ where $X_1 \ldots, X_n$ are $\perp\!\!\!\perp$.

Recall that a property $P(\omega)$ is said to hold almost surely if $\mathbb{P}(\{w \in \Omega \colon P(\omega) \text{ is true }\}) = 1$.

3.1 The use of Borel-Cantelli

Let $(X_n)_{n\geq 1}$ be a sequence of independent, real valued random variables and let $(a_n)_{n\geq 1}$ be a sequence, then

- $\sum_{i=1}^{\infty} \mathbb{P}(X_n \geq a_n) < \infty$, then almost surely for n sufficiently large, $X_n < a_n$.
- $\sum_{i=1}^{\infty} \mathbb{P}(X_n \geq a_n) = \infty$, then almost surely $X_n \geq a_n$ infinitely many often.

This is very often used in the following way

Lemma 3.1. Assume that $\forall \varepsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty$, then $X_n \to X$ almost surely, i.e. $\mathbb{P}(\{\omega \in \Omega \colon X_n(\omega) \to X(\omega)\} = 1$.

Proof. Fix $\varepsilon > 0$. By Borel Cantelli 1. almost surely for n sufficiently large $|X_n - X| \leq \varepsilon$.

But notice that what we want is $X_n \to X$ almost surely, which is equivalent to a.s $\forall \varepsilon > 0, \forall n > N, |X_n - X| \leq \varepsilon$. In general, we CANNOT interchange the "almost surely for all ε" and "for all ε almost surely".

This comes due to the almost surely for all being an uncountable intersection. So instead of all ε, we can take a countable sequence converging to 0, such as $1/n$.

Corollary 3.2. Let $(X_n)_{n\geq 1}$ be a sequence of real-valued independent and identically distributed (iid) r.v.

1. If $\mathbb{E}[|X_1|] < \infty$, then almost surely $X_n/n \underset{n\to\infty}{\to} 0$.
2. If $\mathbb{E}[|X_1|] = \infty$, then almost surely $X_n/n \underset{n\to\infty}{\not\to} 0$.
3. If $\frac{X_1 + \ldots X_n}{n}$ converges as $n \to \infty$, then $\mathbb{E}[|X_1|] < \infty$.

Proof. We show that $\forall \varepsilon > 0$, $\sum_{n\geq 1} \mathbb{P}\left(\left|\frac{X_n}{n}\right| \geq \varepsilon\right) < \infty$.

Recall that if $Z \geq 0$, $\mathbb{E}[Z] = \int_0^{\infty} \mathbb{P}(Z \geq t)dt$ (Identity from PSet4), thus

$$\infty > \mathbb{E}\left[\frac{|X_n|}{\varepsilon}\right] = \int_0^{\infty} \mathbb{P}\left(\frac{|X_n|}{\varepsilon} \geq t\right) dt \geq \sum_{n=1}^{\infty} \int_n^{n+1} \mathbb{P}(|X_n| \geq t\varepsilon)dt,$$

but notice that for $t \in [n, n+1]$, $\mathbb{P}(|X_n| \geq t\varepsilon) \geq \mathbb{P}(|X_n| \geq (n+1)\varepsilon)$, thus we can conclude that the desired sum converges, and apply the lemma above.

Ex. $\to$ Item 2. goes similarly, thus it stays as an exercise.

For part 3. if we take $S_n = X_1 + \ldots + X_n$ and assume that almost surely $S_n/n \to X$, then it is clear that $S_{n+1}/n - S_n/n \to 0$ almost surely, which in turn give us X_{n+1}/n converges almost surely to 0, and we can apply the contrapositive of 2.

A remark for this contrapositive is that the negation of statement 2. goes by If $\mathbb{P}(X_n/n \not\to 0) \neq 1$, then $\mathbb{E}[|X_1|] < \infty$ and not that if it almost surely converges to 0, then has finite expectation.

Theorem 3.3 (Strong Law of Large Numbers - SLN)
Let $(X_i)_{i\geq 1}$ be iid real-valued r.v. such that $\mathbb{E}[|X_1|] < \infty$, then

$$\frac{X_1 + X_2 \ldots + X_n}{n} \to \mathbb{E}[X_1] \ \ a.s.$$

By the previous corollary 3. the integrability condition cannot be removed.

We will start by proving some variants of this theoorem which are easier to establish.

3.2 L^4 version of SLN

Theorem 3.4 (L^4 version of SLN)
Take $(X_n)_{n\geq 1}$ iid real valued r.v. with $\mathbb{E}[|X_1|^4] < \infty$ then

$$\frac{X_1 + \ldots + X_n}{n} \to \mathbb{E}[X_1].$$

Proof. Without loss of generality, assume $\mathbb{E}[X_1] = 0$. Set $S_n = X_1 + \ldots + X_n, K = \mathbb{E}[X_1^4] < \infty$.

We show that $\sum_{n\geq 1} \mathbb{E}[(S_n/n)^4] < \infty (*)$. Indeed if this holds, then

$$\sum_{n\geq 1} \mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] = \mathbb{E}\left[\sum_{n\geq 1} \frac{S_n}{n}^4\right] < \infty,$$

which in turn gives us $\sum_{n\geq 1}(S_n/n)^4 < \infty$ almost surely, thus almost surely $S_n/n \to 0$ as it is the general term of a convergent series.

Hence, let us show the desired identity with a combinatorial argument. Observe that

$$\mathbb{E}[S_n^4] = \sum_{1\leq j_1, j_2, j_3, j_4 \leq n} \mathbb{E}[X_{j_1} X_{j_2} X_{j_3} X_{j_4}]$$

however, by independence and the fact that $\mathbb{E}[X_{j_i}] = 0$, we have that $\mathbb{E}[X_{j_1} X_{j_2} X_{j_3} X_{j_4}] = 0$ as soon as of one these indices is independent from the others. Thus we can simplify to

$$\mathbb{E}[S_n^4] = \sum_{1\le j\le n} \mathbb{E}[X_j^4] + 6 \sum_{1\le j_1<j_2\le n} \mathbb{E}[X_{j_1}^2 X_{j_2}^2] = n\mathbb{E}[X_1^4] + 3n(n-1)\mathbb{E}[X_1^2]^2.$$

Moreover, by Cauchy-Schwarz, $\mathbb{E}[X_1^2]^2 \le \mathbb{E}[X_1^4] = K$, hence $\mathbb{E}[S_n^4] \le 4Kn^2$ and $\mathbb{E}[(S_n/n)^4] \le 4k/n^2$ and therefore $(*)$ holds, as we wanted.

Application 3.5. Let $(A_i)_{i\ge1}$ be independent events with same probability p, then

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i} \underset{n\to\infty}{\to} p \ \ a.s.$$

This makes a connection between the "historical" definition of probabilities as the frequency of an event happening when repeating an experiment many times and our "modern" axiomatic approach of probability theory.

3.3 Kolmogovor's Two Series Theorem

Kolmogorov's series theorems gives condtions for almost sure convergence of $\perp\!\!\!\perp$ random variables (not identically distributed).

Lemma 3.6 (Kolmogorov's Maximal Inequality)**.** Let $(Z_k)_{1\le k\le n}$ be $\perp\!\!\!\perp$ real-valued r.v. in L^2. Set $S_K = Z_1 + \ldots Z_k$ for $1 \le k \le n$. Assume tat $\mathbb{E}[Z_K] = 0$ for every $1 \le k \le n$. Then $\forall\lambda > 0$

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k| \ge \lambda\right) \le \frac{\mathbb{E}[S_n^2]}{\lambda^2}.$$

Proof. Idea For $1 \le k \le n$, introduce $A_k = \{|S_k| \ge \lambda, |S_i| < \lambda \forall i < k\}$. These events are disjoint and they union is $\{\max_{1\le k\le n}\}$. Since they are disjoint, $0 \le \sum_{i=1}^k \mathbb{1}_{A_i} \le 1$.

Then $S_n^2 \ge S_n^2 \sum_{k=1}^n \mathbb{1}_{A_k}$, so $\mathbb{E}[S_n^2] \ge \sum_{k=1}^n \mathbb{E}[S_n^2 \mathbb{1}_{A_k}]$.

Idea $S_n^2 = S_k^2 + 2(S_k)(S_n - S_k) + (S_n - S_k)^2$. We force the appearence of $S_n - S_k$ because $S_n - S_k \perp\!\!\!\perp (Z_1, \ldots, Z_k)$.

Hence using that $(S_n - S_k)^2 \ge 0$

$$\mathbb{E}[S_n^2] \ge \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbb{1}_{A_k}] + \sum_{k=1}^n \mathbb{E}[2S_k(S_n - S_k)\mathbb{1}_{A_k}]$$

observe that $2S_k\mathbb{1}_{A_k}$ is $\sigma(Z_1, \ldots Z_k)$-measurable and $(S_n - S_k)$ is $\sigma(Z_{k+1}, \ldots, Z_n)$-measurable, thus they are independent.

So $\mathbb{E}[2S_k(S_n - S_k)\mathbb{1}_{A_k}] = 2\mathbb{E}[S_k\mathbb{1}_{A_k}]\mathbb{E}[S_n - S_k] = 0$ as we have $\mathbb{E}[Z_k] = 0$.

Finally, as $S_k^2 \mathbb{1}_{A_k} \geq \lambda^2 \mathbb{1}_{A_k}$ we obtain

$$\mathbb{E}[S_n^2] \geq \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbb{1}_{A_k}] + \sum_{k=1}^n \mathbb{E}[2S_k(S_n - S_k)\mathbb{1}_{A_k}] \geq \lambda^2 \left(\sum_{k=1}^n \mathbb{P}(A_k) \right) = \lambda^2 \mathbb{P}(\max_{1\leq k \leq n} |S_k| \geq \lambda)$$

Theorem 3.7 (Kolmogorov Two Series Theorem)
Let $(Z_k)_{k\geq 1}$ be $\perp\!\!\!\perp$ real valued r.v. in L^2. Assume that

1. $\sum_{n\geq 1} \mathbb{E}[Z_n]$ converges in $\mathbb{R}$.
2. $\sum_{n\geq 1} Var(Z_n) < \infty$.

Then $\sum_{k=1}^n Z_k$ converges almost surely as $n \to \infty$.

Remark 3.8. We do not assume that (Z_k) have the same law. In fact, if this was the case, for any $Var(Z_1) > 0$, then the second condition never holds.

Proof. We show that almost surely $(\sum_{k=1}^n Z_k)_{n\geq 1}$ is a Cauchy Sequence.

Since $Var(Z_n - \mathbb{E}[Z_n]) = Var(Z_n)$, we can assume that $\mathbb{E}[Z_n] = 0$ for $1 \leq k \leq n$ (we then apply the result with $Z_k - \mathbb{E}[Z_k]$).

Set $S_n = Z_1 + \ldots + Z_n$. The idea is to show:

$$\forall k \geq 1, \ a.s. \ \exists m \geq 1 \ s.t. \ \forall n \geq m, |S_n - S_m| \leq \frac{1}{k} \quad (*)$$

Indeed, then we interchange $\forall k \geq 1$ and almost surely to get (as it is a countable set):

$$a.s. \ \forall k \geq 1, \exists m \geq 1 \ s.t. \ n \geq m \implies |S_n - S_m| < \frac{1}{k}.$$

Notice that this gives us $\forall p, q \geq m$, $|S_p - S_q| < 2/k$ due to triangular inequality, which in turn is enough to imply that almost surely (S_n) is a Cauchy sequence.

Now let us go back to proving $(*)$.

Fix $k \geq 1$ and set A_m to be the event that $\forall n \geq m, |S_n - S_m| \leq 1/k$. We want to show that $\mathbb{P}(\bigcup_{m\geq 1} A_m) = 1$, but it is clear by definition that (A_m) is increasing, so $\mathbb{P}(\bigcup_{m\geq 1} A_m) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

But now notice $1 - \mathbb{P}(A_m) = \mathbb{P}(\exists n \geq m \colon |S_n - S_m| > 1/k) = \lim_{l\to\infty} \mathbb{P}(\exists n, m \leq n \leq l \colon |S_n - S_m| > 1/k$.

Finally, we rewrite this more explicitly to

$$\mathbb{P}(\exists n, m \leq n \leq l \colon |Z_{m+1} + \ldots + Z_n| > 1/k) \leq k^2(\mathbb{E}[Z_{m+1}^2] + \ldots + \mathbb{E}[Z_l^2])$$

which holds by Kolmogorov Max Inequality.

Moreover, this yields

$$1 - \mathbb{P}(A_m) \leq \lim_{l\to\infty} k^2 \sum_{i>m} (Var(Z_i)) \underset{m\to\infty}{\longrightarrow} 0$$

which is enough to conclude!

3.4 Three Series Theorem

Theorem 3.9 (Kolmogorov Three Series Theorem)
Let $(X_n)_{n\geq 1}$ be $\perp\!\!\!\perp$ real random variables. Assume that there exists $a > 0$ such that

1. $\sum_{k=1}^{\infty} \mathbb{P}(|X_k| \geq a) < \infty$
2. $\sum_{k=1}^{\infty} \mathbb{E}[X_k \mathbb{1}_{|X_k|<a}]$ converges in $\mathbb{R}$
3. $\sum_{k=1}^{\infty} Var(X_k \mathbb{1}_{|X_k|<a}) < \infty$

then almost surely $\sum_{k=1}^{n} X_k$ converges as $n \to \infty$.

Remark 3.10. $X_k \mathbb{1}_{|X_k|<a}$ is bounded random variable so it is in L^2.

Remark 3.11. It is possible to show that the converse is true, that is if $\sum_{k=1}^{n} X_k$ converges then 1., 2., 3. hold for every $a > 0$.

In other words, if 1., 2. or 3. fails for some $a > 0$, then almost surely $\sum_{k=1}^{n} X_k$ diverges as $n \to \infty$.

Remark 3.12. Strictly speaking the converse gives that if one of the condition fails, then $\mathbb{P}(\sum_{k=1}^{n} X_k \text{ converges}) < 1$, but this implies by Kolmogorov's $0-1$ law that this probability is 0.

Proof. We use Borel Cantelli due to Condition 1. to obtain that almost surely for k sufficiently large, $|X_k| < a$.

Thus, if we set $Z_k = X_k \mathbb{1}_{|X_k|<a}$, almost surely for k sufficiently large $Z_k = X_k$, thus almost surely $\sum Z_k$ converges iff $\sum X_k$ converges. However, by the Two Series Theorem, almost surely $\sum Z_k$ converges as $(Z_k)_{k\geq 1}$ are $\perp\!\!\!\perp$ by the composition principle and 2. and 3. satisfy the conditions of the previous theorem.

3.5 The Strong Law of Large Numbers

Theorem 3.13
Let $(X_i)_{i\geq 1}$ be iid real-valued r.v, $\mathbb{E}[|X_1|] < \infty$, then

$$\frac{X_1 + \ldots + X_n}{n} \underset{n\to\infty}{\longrightarrow} \mathbb{E}[X_1].$$

Lemma 3.14 (Kronecker). Let $(x_n)_{n\geq 1}$ be real numbers such that $\sum_{k=1}^n x_k/k$ converges as $n\to\infty$ then

$$\frac{x_1+\ldots+x_n}{n} \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. Set $w_n = \sum_{k=1}^n \frac{x_k}{k}$, assume $w_n \to w$ as $n\to\infty$. By Cesaro's Theorem, $\frac{1}{N}\sum_{n=1}^N w_n \to w$ as $N\to\infty$.

Now, let us proceed with calculations

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^N w_n &= \frac{1}{N}\sum_{n=1}^N\sum_{k=1}^n \frac{x_k}{k} = \frac{1}{N}\sum_{n=1}^N\sum_{k=1}^N \mathbb{1}_{k\leq n}\frac{x_k}{k}\\
&= \frac{1}{N}\sum_{k=1}^N\sum_{n=1}^N \mathbb{1}_{k\leq n}\frac{x_k}{k} = \frac{1}{N}\sum_{k=1}^N\frac{x_k}{k}\sum_{n=1}^N \mathbb{1}_{k\leq n}\\
&= \frac{1}{N}\sum_{k=1}^N\frac{(N-k+1)x_k}{k} = \frac{N+1}{N}\sum_{k=1}^N\frac{x_k}{k} - \frac{1}{N}\sum_{k=1}^N x_k
\end{aligned}$$

Now notice that both $1/N\sum_{k=1}^N x_k$ is the difference of two series that converge, so it must converge as well.

Proof (Strong Law of Large Numbers).

First let us assume that $\mathbb{E}[X_1]=0$.

If $\sum_{k=1}^n \frac{X_k}{k}$ converges almost surely, then by Kronecker Lemma almost surely $\frac{1}{n}\sum_{k=1}^n X_k \to 0$ as $n\to\infty$. Unfortunately this is not always the case, so we need to move to a cutoff argument.

We check that $\sum_{n=1}^\infty \mathbb{P}(|X_n|>n)<\infty$. Indeed $\sum_{n=1}^\infty \mathbb{P}(|X_n|>n) = \sum_{n=1}^\infty \mathbb{P}(|X_1|>n) < \mathbb{E}[|X_1|]$. This gives us by Borel Cantelli that almost surely for n sufficiently large $|X_n|\leq n$.

Therefore, it is enough to show that $(X_1'+\ldots+X_n')/n$ converges to 0 almost surely if we define $X_i' = X_i\mathbb{1}_{|X_i|\leq i}$

We can check that $\mathbb{E}[X_i'] = \mathbb{E}[X_1\mathbb{1}_{|X_1|\leq i}] \to \mathbb{E}[X_1]$ as $i\to\infty$. Thus, it is enough to show that

$$\frac{Y_1'+\ldots+Y_n'}{n} \xrightarrow{a.s} 0 \quad (*)$$

with $Y_i' = X_i' - \mathbb{E}[X_i']$.

To show $(*)$ we shot that almost surely $\sum_{k=1}^n \frac{Y_k'}{k}$ converges as $n\to\infty$ $(**)$ and the result will follow by Kronecker's Lemma.

To show $(**)$ we use Kolmogorov's Two Series Theorem. We must just check the conditions for the theorem. First, by the composition principle $(Y_k'/k)_{k\geq 1}$ are independent. Second, as $\mathbb{E}[Y_k']=0$, the condition 1. also holds. Finally, for the sum of the variance, write

$$Var\left(\frac{Y'_k}{k}\right) = \frac{1}{k^2}Var(X'_k) \leq \frac{1}{k^2}\mathbb{E}[X'^2_k] = \frac{1}{k^2}\mathbb{E}[X_1^2\mathbb{1}_{|X_1|\leq k}]$$

Moreover, $\mathbb{E}[X_1^2\mathbb{1}_{|X_1|\leq k}] = \sum_{j=1}^n \mathbb{E}[X_1^2\mathbb{1}_{j-1<|X_1|\leq j} \leq \sum_{j=1}^k j^2\mathbb{P}(j-1 < |X_1| \leq j)$.

Thus

$$\begin{aligned}
\sum_{n=1}^{\infty} Var\left(Y'_k/k\right) &\leq \sum_{n=1}^{\infty}\sum_{j=1}^{n}\frac{1}{n^2}j^2\mathbb{P}(j-1<|X_1|\leq j) \\
&= \sum_{n=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{1}_{j\leq n}\frac{1}{n^2}j^2\mathbb{P}(j-1<|X_1|\leq j) \\
&= \sum_{j=1}^{\infty}\left(\sum_{n=j}^{\infty}\frac{1}{n^2}\right)j^2\mathbb{P}(j-1<|X_1|\leq j) \\
&\leq \sum_{j=1}^{\infty}\frac{c}{j}\mathbb{P}(j-1<|X_1|\leq j) \\
&= c\sum_{j=1}^{\infty} j\int \mathbb{1}_{j-1<|X_1|\leq j}\mathbb{P}_{|X_1|}(dx) \\
&= c\int_0^{\infty}\sum_{j=0}^{\infty} j\mathbb{1}_{j-1<|X_1|\leq j}\mathbb{P}_{|X_1|}(dx) \\
&\leq c\int_0^{\infty}\sum_{j=0}^{\infty}(x+1)\mathbb{1}_{j-1<|X_1|\leq j}\mathbb{P}_{|X_1|}(dx) \\
&= c\mathbb{E}[|X_1|+1] < \infty
\end{aligned}$$

so the last condition is also satisfied and we are done.

3.6 Different Notions of Convergence

Let X, X_n be random variables in $\mathbb{R}^k$ (with any norm). We have already seen the notion of almost sure convergence:

$$X_n \xrightarrow{a.s} X \quad \text{if } \mathbb{P}(\{\omega\in\Omega\colon X_n(\omega)\to X(\omega)\}) = 1$$

Definition 3.15. We say that $X_n \to X$ in probability and write $X_n \xrightarrow{\mathbb{P}} X$ if $\forall\varepsilon > 0$, $\mathbb{P}(|X_n - X| \geq \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$. Here the norm $|\cdot|$ is the norm in $\mathbb{R}^k$.

If X_n, X are $\mathbb{R}-$valued, we say that X_n converges to X in L^p if $\mathbb{E}[|X_n - X|^p] \underset{n\to\infty}{\longrightarrow} 0$.

Remark 3.16. Almost sure convergence involves the joint law of $(X, X_1, X_2, \ldots)$ while convergence in probability and L^p only involve the joint law of (X_n, X).

Remark 3.17. By monotonicity, $\varepsilon' > \varepsilon$ then $\mathbb{P}(|X_n - X| \geq \varepsilon') \leq \mathbb{P}(|X_n - X| \geq \varepsilon)$, so $X_n \xrightarrow{\mathbb{P}} X$ if $\forall\varepsilon > 0$ small enough the condition holds.

Proposition 3.18. $X_n \xrightarrow{\mathbb{P}} X$ iff $\mathbb{E}[\min(|X_n - X|, 1)] \to 0$.

Proof. $\boxed{\Rightarrow}$ Take $\varepsilon > 0$ and write

$\mathbb{E}[min(|X_n - X|, 1)] = \mathbb{E}[\min(|X_n - X|, 1)\mathbb{1}_{|X_n - X| < \varepsilon}] + \mathbb{E}[\min(|X_n - X|, 1)\mathbb{1}_{|X_n - X| \geq \varepsilon}]$

Moreover, we have $\mathbb{E}[\min(|X_n - X|, 1)\mathbb{1}_{|X_n - X| < \varepsilon}] \leq \mathbb{E}[\varepsilon] = \varepsilon$ and $\mathbb{E}[\min(|X_n - X|, 1)\mathbb{1}_{|X_n - X| \geq \varepsilon}] \leq \mathbb{E}[\mathbb{1}_{|X_n - X| \geq \varepsilon}] = \mathbb{P}(|X_n - X| \geq \varepsilon)$
Thus $\limsup_{n\to\infty} \mathbb{E}[\min(|X_n - X|, 1)] \leq \varepsilon$, which holds for all ε, so it must be 0.

$\boxed{\Leftarrow}$ Take $\varepsilon \in [0, 1]$ and observe that $|X_n - X| \geq \varepsilon \implies \min(|X_n - X|, 1) \geq \varepsilon$, thus $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \mathbb{P}(\min(|X_n - X|, 1) \geq \varepsilon) \leq \mathbb{E}[min(|X_n - X|, 1)]/\varepsilon \underset{n\to\infty}{\longrightarrow} 0$, by Markov's inequality.

Proposition 3.19. If $X_n \xrightarrow{a.s} X$ or $X_n \xrightarrow{L^p} X$ then $X_n \xrightarrow{\mathbb{P}} X$

Proof. Assume $X_n \xrightarrow{L^p} X$. Fix $\varepsilon > 0$ and write $\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \mathbb{E}[|X_n - X|^p]/\varepsilon^p \underset{n\to\infty}{\longrightarrow} 0$ again by Markov's Inequality.

Assume that $X_n \xrightarrow{a.s.} X$. Now observe that $\min(|X_n - X|, 1) \xrightarrow{a.s.} 0$ and $0 \leq \min(|X_n - X|, 1) \leq 1$, hence by Dominated Convergence we get the result.

Remark 3.20. For $p = 2$ and $\mu = \mathbb{E}[X_n]$ the inequality

$$\mathbb{P}(|X_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{E}[(X_n - \mu)^2]}{\varepsilon^2} = \frac{Var(X_n)}{\varepsilon^2}$$

is know as the Bienaymé-Tchebyshev Inequality.

Example 3.21. Fix $\alpha > 0$ and let $(X_n)_{n\geq 1}$ be $\perp\!\!\!\perp$ r.v. with $\mathbb{P}(X_n = 1) = 1/n^\alpha$ and $\mathbb{P}(X_n = 0) = 1 - 1/n^\alpha$.

For this, we can compute $\mathbb{E}[X_n^p] = 1/n^\alpha \underset{n\to\infty}{\longrightarrow} 0$, hence it converges in L^p and probability to 0.

<u>What about a.s convergence?</u>
For $\alpha > 1$, we have that $\sum_{n=1}^\infty \mathbb{P}(X_n = 1) < \infty$, thus by Borel Cantelli, almost surely $X_n = 1$ happens a finite number of times, thus $X_n \xrightarrow{a.s.} 0$.
For $\alpha \leq 1$, we have that $\sum_{n=1}^\infty \mathbb{P}(X_n = 1) = \infty$, and $\sum_{n=1}^\infty \mathbb{P}(X_n = 0) = \infty$ thus by Borel Cantelli, since $(\{X_n = 1\})_{n\geq 1}$ are independent and $(\{X_n = 0\})_{n\geq 1}$ are as well, we have that almost surely $X_n = 1$ and $X_n = 0$ infinitely often, so almost surely it does not converge.

Lemma 3.22. If $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} Y$, then $X = Y$ almost surely.

Proof. Fix $m \geq 1$, then $\mathbb{P}(|X-Y| \geq 2/m) \leq \mathbb{P}(|X_n - X| \geq 1/m) + \mathbb{P}(|X_n - Y| \geq 1/m)$, but as $n \to \infty$, we have that the terms in the right hand side converge to 0, thus $\mathbb{P}(|X-Y| \geq 2/m) = 0$, fomr which the result follows.

Lemma 3.23 (Subsequence Lemma)**.** We have $X_n \xrightarrow{\mathbb{P}} X$ iff of every subsequence of (X_n) we can extract a subsubsequence which converges $a.s$ to X. (a subsequence of (X_n) is $(X_{\varphi(n)})$ with φ an increasing function mapping the naturals to itself.)

Proof. $\boxed{\Rightarrow}$ Let ϕ be a subsequence. Since $X_n \xrightarrow{\mathbb{P}} X$, we have $X_{\varphi(n)} \xrightarrow{\mathbb{P}} X$ so $\mathbb{E}[\min(|X_{\varphi(k)} - X|, 1)] \underset{k\to\infty}{\longrightarrow} 0$

Therefore we can find a subsequence ψ such that $\forall n \geq 1$ $\mathbb{E}[min(|X_{\varphi(\psi(n))} - X|, 1)] \leq 1/2^n$. Indeed, for k sufficiently large we have $\mathbb{E}[\min(|X_{\varphi(k)} - X|, 1)] \leq 1/2^n$. Then $\sum_{n=1}^{\infty} \mathbb{E}[\min(|X_{\varphi(\psi(n))} - X|, 1)] < \infty$, which then implies $\mathbb{E}[\sum_{n=1}^{\infty} \min(|X_{\varphi(\psi(n))} - X|, 1)] < \infty$, thus almost surely $\sum_{n=1}^{\infty} \min(|X_{\varphi(\psi(n))} - X|, 1) < \infty$, which is enough to conclude that $|X_{\varphi(\psi(n))} - X|$ converges almost surely to 0.

$\boxed{\Leftarrow}$ Assume that $\forall \varphi$, $\exists \psi$ such that $X_{\varphi(\psi(n))} \xrightarrow{a.s.} X$. Argue by contradiction, then $\mathbb{E}[\min(|X_n - X|, 1)] \underset{n\to\infty}{\not\to} 0$.

Thus there exists $\varepsilon > 0$ and a subsequence ϕ such that $\mathbb{E}[\min(|X_{\phi(n)} - X|, 1)] \geq \varepsilon$. But by assumption, there exists a ψ subsequence such that $X_{\varphi(\psi(n))} \xrightarrow{a.s.} X$, thus $X_{\varphi(\psi(n))} \xrightarrow{\mathbb{P}} X$, thus $\mathbb{E}[\min(|X_{\varphi(\psi(n))} - X|, 1)] \underset{n\to\infty}{\longrightarrow} 0$, which contradicts the first identity of this paragraph.

Application 3.24. Assume $X_n \xrightarrow{\mathbb{P}} X$ and f continuous. Then $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Proof. Take any φ a subsequence, then by the subsequence lemma there exists ψ such that $X_{\varphi(\psi(n))} \xrightarrow{a.s} X$, which implies $f(X_{\varphi(\psi(n))}) \xrightarrow{a.s.} f(X)$, which in turn implies by the subsequence lemma the desired identity.

Example 3.25 (Flying Saucepans)**.** Equip $[0,1]$ with the Borel $\sigma-$field, and let λ be the Lebesgue Measure. For $k \geq 0$ and $0 \leq j \leq 2^k - 1$ define

$$X_{2^k+j}(\omega) = \mathbb{1}_{\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]}(\omega).$$

Then $X_n \xrightarrow{\mathbb{P}} 0$ as $\mathbb{P}(|X_n| > \varepsilon) \leq 1/n$.
But $\forall \omega \in [0,1]$, there exists infinitely many $n \geq 1$ such that $X_n(\omega) = 1$, so X_n diverges almost surely.

Example 3.26 (Spiky Cat)**.** Take again $[0,1]$ Set $X_n(\omega) = 2^n \mathbb{1}_{[0,1/2^n]}(\omega)$ for $\omega \in [0,1]$, then $X_n \xrightarrow{a.s} 0$ but $\mathbb{E}[X_n] = 1$, so X_n does not converge to 0 by L^1.

In the example above, the portion of space where $X_n \neq 0$ becomes small, however its contribution to the expected value is constant. We have a probabilistic notion that prevents such spikes, which is uniform integrability.

We saw that for $X \in L^1$ we have $\mathbb{E}[|X|\mathbb{1}_{|X|\geq x}] \underset{x\to\infty}{\longrightarrow} 0$ by dominated convergence. Uniform integrability extends this to a family of random variables.

Definition 3.27 (Uniformly Integrable Family)**.** A family $(X_i)_{i\in I}$ of integrable random variables is uniformly integrable if $\sup_{i\in I} \mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq x}] \underset{x\to\infty}{\longrightarrow} 0$

Equivalently, $\forall \varepsilon > 0, \exists x > 0$ such that $\forall i \in I, \mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq x}] \leq \varepsilon$.

Example 3.28.

- A finite family of L^1 random variables is UI by dominated convergence applied a finite number of times.
- If $Z \geq 0$ is integrable, then $\{X\colon |X| \leq Z\}$ is UI. Indeed if $|X| \leq Z$, the $\mathbb{E}[|X|\mathbb{1}_{|X|\geq x}] \leq \mathbb{E}[Z\mathbb{1}_{Z\geq x}]$.
- If $(X_i)_{i\in I}$ is bounded in L^p for $p > i$ i.e., $\exists c > 0$ such that $\forall i \in I$ $\mathbb{E}[|X_i|^p] \leq C$, then (X_i) is uniformly integrable. Indeed

$$\mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq x}] = \mathbb{E}\left[\frac{|X_i|}{|X_i|^p}|X_i|^p\mathbb{1}_{|X_i|\geq x}\right] \leq \frac{\mathbb{E}[|X_i|^p]}{x^{p-1}} \leq \frac{C}{x^{p-1}}.$$

Remark 3.29. By definition, a sequence $(X_n)_{n\geq 1}$ of L^1 random variables is UI if

$$\sup_{n\geq 1} \mathbb{E}[|X_n|\mathbb{1}_{|X_n|\geq k}] \underset{k\to\infty}{\longrightarrow} 0$$

But since it is a finite family of L^1 random variables, this is equivalent to

$$\limsup_{n\geq 1} \mathbb{E}[|X_n|\mathbb{1}_{|X_n|\geq k}] \underset{k\to\infty}{\longrightarrow} 0$$

Theorem 3.30 ($\varepsilon - \delta$ condition)

A family $(X_i)_{i\in I}$ of L^1 random variables is Uniformly Integrable iff

1. $(X_i)_{i\in I}$ is bounded in L^1 (i.e. $\sup_{i\in I} \mathbb{E}[|X_i|] < A$)
2. $\forall \varepsilon > 0 \exists \delta > 0$ such that $\forall$ event A with $\mathbb{P}(A) \leq \delta$, $\mathbb{E}[|X_i|\mathbb{1}_A] \leq \varepsilon$ for every $i \in I$

Corollary 3.31. If $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ are two families which are UI, then $\{X_i+Y_j : i \in I, j \in J\}$ is UI.

Proof. $\boxed{\Rightarrow}$ Let $K > 0$ be such that $\mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq k}] \leq 1$ for every $i \in I$, then

$$\mathbb{E}[|X_i|] = \mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq k}] + \mathbb{E}[|X_i|\mathbb{1}_{|X_i|\leq k}] \leq 1 + k$$

so (X_i) is bounded in L^1.

Now we proceed for the $\varepsilon - \delta$ condition. Fix $\varepsilon > 0$. Let K_ε be such that $\sup_{i\in I} \mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq K_\varepsilon}] \leq \varepsilon$, then taking $\delta = \varepsilon/K_\varepsilon$ we get for $\mathbb{P}(A) \leq \delta$

$$\mathbb{E}[|X_i|\mathbb{1}_A] = \mathbb{E}[|X_i|\mathbb{1}_A\mathbb{1}_{|X_i|\geq K_\varepsilon}] + \mathbb{E}[|X_i|\mathbb{1}_A\mathbb{1}_{|X_i|\geq K_\varepsilon}] \leq \varepsilon + K_\varepsilon\mathbb{P}(A) \leq 2\varepsilon.$$

$\boxed{\Leftarrow}$ Fix $\varepsilon > 0, \delta > 0$ such that the condition holds. Let $k > 0$ be such that $\sup_{i\in I} \mathbb{E}[|X_i|] \leq K\delta$. Then by Markov's inequality

$$\mathbb{P}(|X_i| \geq k) \leq \frac{\mathbb{E}[|X_i|]}{K} \leq \delta$$

Thus we can just apply the $\varepsilon - \delta$ condition with $A = \{|X_i| \geq k\}$ to get the desired result.

UI bridges the gap between convergence in $\mathbb{P}$ and convergence in L^1

Theorem 3.32 (Super Dominated Convergence)
Let (X_n) be integrable real-valued random variables, X a real valued random variable then the following conditions are equivalent

1. $X \in L^1$ and $X_n \xrightarrow{L^1} X$
2. $X_n \xrightarrow{\mathbb{P}} X$ and $(X_n)_{n\geq 1}$ is UI.

(The name comes from the fact that $\{X : |X| \leq Z\}$ with $Z \geq 0$ integrable is a UI family: it implies dominated convergence).

Proof. $\boxed{1. \Rightarrow 2.}$ We know that $X_n \xrightarrow{L^1} X$ implies $X_n \xrightarrow{\mathbb{P}} X$. To show that $(X_n)_{n\geq 1}$ is UI by the corollary, it suffices to show that $(X_n - X)_{n\geq 1}$ is UI.

To do this, fix $\varepsilon > 0$ and choose n_0 such that $n \geq n_0$ implies $\mathbb{E}[|X_n - X|] \leq \varepsilon$. Let k_0 be such that $k \geq k_0$ implies $\max_{1\leq i\leq n_0} \mathbb{E}[|X_i - X|\mathbb{1}_{|X_i - X|\geq k}] \leq \varepsilon$.

Thus $\forall n \geq 1$, $\mathbb{E}[|X_n - X|\mathbb{1}_{|X_n - X|\geq k}] \leq \varepsilon$ for $k \geq k_0$.

$\boxed{2. \Rightarrow 1.}$ We first show that $X \in L^1$. Since $X_n \xrightarrow{\mathbb{P}} X$, there exists a subsequence ψ such that $X_{\psi(n)} \xrightarrow{a.s.} X$.

Thus by Fatou's Lemma

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf_{n\to\infty} |X_{\psi(n)}|] \leq \liminf_{n\to\infty} \mathbb{E}[|X_{\psi(n)}|] < \infty$$

Now we show that $X_n \xrightarrow{L^1} X$. Since $X \in L^1$, we have $(X_n - X)$ is UI by the corollary. Now fix $\varepsilon > 0$ and let $\delta > 0$ be such that the $\varepsilon - \delta$ condition holds.

Then for n sufficiently large $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \delta$ because $X_n \xrightarrow{\mathbb{P}} X$. Thus

$$\mathbb{E}[|X_n - X|] = \mathbb{E}[|X_n - X|\mathbb{1}_{|X_n - X| < \varepsilon}] + \mathbb{E}[|X_n - X|\mathbb{1}_{|X_n - X| \geq \varepsilon}] \leq \varepsilon + \varepsilon$$

finishing the proof.

Remark 3.33. Existence of a sequence of id random variables. We have implicitly used the following theorem so far

Theorem 3.34

Let μ be a probability distribution on $\mathbb{R}$. There exists a sequence $(X_n)_{n\geq 1}$ of id random variables with law μ.

This is related to the existence of product measures on infinite product spaces (see lecture notes).

4 Conditional Expectation

4.1 Discrete Setting

Goal: see how the knowledge of information modifies probability measures. Here we will "just" define the conditional expectation of random variables given a $\sigma-$field.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Take $B \in \mathcal{A}$ with $\mathbb{P}(B) > 0$. We can define

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

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

$\mathbb{P}(\cdot|B)$ defines a probability measure called the conditional probability given the EVENT B.

Similarly for $X \in L^1$ we define

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}.$$

Interpretation: Average value of X when B occurs. $\mathbb{E}[X|B]$ is the expectation of X in $(\Omega, \mathcal{A}, \mathbb{P}(\cdot|B))$.

Now let $Y : \Omega \to E$ be a random variable with E countable. We want to define $\mathbb{E}[X|Y]$. From before, we have $\mathbb{E}[X|Y = y] = \dfrac{\mathbb{E}[X \mathbb{1}_{Y=y}]}{\mathbb{P}(Y = y)}$ for every y with $\mathbb{P}(Y = y) > 0$.

Thus we naturally set $\phi(y) : E \to \mathbb{R}$ to $y \mapsto \mathbb{E}[X|Y = y]$ if $\mathbb{P}(Y = y) > 0$ and 0 otherwise. Moreover, $\phi(Y)$ is itself a random variable which is $\sigma(Y)-$measurable.

In other words: $\mathbb{E}[X|Y](\omega) = \phi(Y(\omega))$

Example 4.1. Let the space be such that $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathbb{P}(\{\omega\}) = 1/6 \ \forall \omega \in \Omega$
Now let $X(\omega) = \omega$ and $Y(\omega)$ be the indicator of ω being odd.
What is $\mathbb{E}[X|Y]$?

Lemma 4.2. We have

1. $\mathbb{E}[X|Y] \in L^1$
2. $\forall Z$ a bounded random variable, $\sigma(Y)-$measurable, $\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|Y]]$

Proof. For the first statement, we have $\mathbb{E}[|\mathbb{E}[X|Y]|] = \mathbb{E}[|\phi(Y)|] = \sum_{y \in E} \mathbb{P}(Y = y)|\phi(y)|$, so substituting the definition

$$\mathbb{E}[|\mathbb{E}[X|Y]|] = \sum_{y \in E} |\mathbb{E}[X \mathbb{1}_{Y=y}]| \leq \sum_{y \in Y} \mathbb{E}[|X| \mathbb{1}_{Y=y}] = \mathbb{E}[|X|] < \infty.$$

Now for the second statement, we take Z $\sigma(Y)-$measurable and bounded, which ensures that ZX and $Z\mathbb{E}[X|Y]$ are both L^1.

By the Doob-Dynkin Lemma, there exists F measurable such that $Z = F(Y)$. Then

$$\begin{aligned}
\mathbb{E}[Z\mathbb{E}[X|Y]] = \mathbb{E}[F(Y)\mathbb{E}[X|Y]] = \sum_{y \in E} \mathbb{P}(Y = y)F(y)\phi(y) \\
&= \sum_{\substack{y \in E \\ \mathbb{P}(Y=y)>0}} F(Y)\mathbb{E}[X\mathbb{1}_{Y=y}] \\
&= \mathbb{E}[X \sum_{\substack{y \in E \\ \mathbb{P}(Y=y)>0}} F(y)\mathbb{1}_{Y=y}] \quad \text{By Fubini-Lebesgue} \\
&= \mathbb{E}[XF(Y)] \quad \text{as the sum is almost surely } F(Y)
\end{aligned}$$

4.2 Definition and First Properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $\mathcal{A} \subset \mathcal{F}$ is a sub $\sigma-$ field we write $(X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ if

- $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable
- $\mathbb{E}[|X|] < \infty$.

Theorem 4.3

Fix $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{A} \subset \mathcal{F}$ be a $\sigma-$field. There exists a $\mathbb{R}-$valued random variable X' with

- $X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.
- $\forall Z \geq 0$ random variable $\mathcal{A}$-measurable and bounded $\mathbb{E}[ZX] = \mathbb{E}[ZX']$.

Moreover, if X'' is another variable satisfying the theorem above, then $X' = X''$ almost surely.

We denote by $\mathbb{E}[X|\mathcal{A}]$ any such random variable, called a version of the conditional expectation of X given $\mathcal{A}$.

Remark 4.4.

1. 2. is called "characteristic property of conditional expectation"
2. $\mathbb{E}[X|\mathcal{A}]$ is a random variable, $\mathcal{A}$ measurable, defined uniquely up to 0 probability events. In practice this is not a problem because we only consider its expectation or almost sure properties.
3. Interpretation of 2.: "$\langle Z, X - X' \rangle = \mathbb{E}[Z(X - X')] = 0$". Intuitively, $\mathbb{E}[X|\mathcal{A}]$ is the projection of X on $\mathcal{A}-$measurable random variables. We will make this precise for $X \in L^2$.

Notation.

- Take $Y\colon (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ a random variable, we define

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)].$$

- If $B \in \mathcal{F}$ is an event, we define

$$\mathbb{P}(B|\mathcal{A}) = \mathbb{E}[\mathbb{1}_B|\mathcal{A}],$$

  it is an $\mathcal{A}-$ measurable random variable.

Remark 4.5. This definition is consistent with what we saw in the discrete setting. Indeed take $Y\colon (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ a random variable with E countable. Let us find $\mathbb{E}[X|Y]$.

- We know that $\forall Z$ $\mathbb{R}-$valued and $\sigma(Y)-$measurable $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|Y]Z]$
- $\mathbb{E}[X|Y]$ is $\sigma(Y)-$measurable, so by the Doob-Dynkin lemma we can write $\mathbb{E}[X|Y] = \phi(Y)$. Let us find ϕ.

  We take $Z = \mathbb{1}_{Y=y}$ for $y \in E$ and get $\mathbb{E}[X\mathbb{1}_{Y=y}] = \mathbb{E}[\phi(Y)\mathbb{1}_{Y=y}] = \mathbb{E}[\phi(y)\mathbb{1}_{Y=y}] = \phi(y)\mathbb{P}(Y = y)$, from which we get the desired definition of ϕ.

Remark 4.6 (Generalization of Doob-Dynkin)**.** More generally, if Y is $\mathbb{R}^n-$valued, then a $\sigma(Y)-$measurable function is of the form $F(Y)$ with F measurable.

As a consequence, to find $\mathbb{E}[X|Y]$ we often find a function ϕ such that for every f $\mathbb{R}-$valued and bounded $\mathbb{E}[Xf(Y)] = \mathbb{E}[\phi(X)f(Y)]$. Indeed, by Doob-Dynkin this implies that $\mathbb{E}[XZ] = \mathbb{E}[\phi(Y)Z]$ for every Z real valued and bounded (prop 2 of the definition). Since $\phi(Y) \in L^1(\Omega, \sigma(Y), \mathbb{P})$ we conclude that $\mathbb{E}[X|Y] = \phi(Y)$.

Simple properties of conditional expectation Take $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{A} \subset \mathcal{F}$ a $\sigma-$field. Then we have the following almost sure properties:

1. $\mathbb{E}[X|\mathcal{F}] = X$. and $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$.
2. If X is $\mathcal{A}-$measurable, then $\mathbb{E}[X|\mathcal{A}] = X$.
3. $X \mapsto \mathbb{E}[X|\mathcal{A}]$ is linear.
4. $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]] = \mathbb{E}[X]$.
5. $X_1 \geq X_2$ implies $\mathbb{E}[X_1|\mathcal{A}] \geq \mathbb{E}[X_2|\mathcal{A}]$.
6. $|\mathbb{E}[X|\mathcal{A}]| \leq \mathbb{E}[|X||\mathcal{A}]$.

Proposition 4.7. Let Y be a random variable with $X \perp\!\!\!\perp Y$ $(Y\colon \Omega \to E)$. Then $\mathbb{E}[X|Y] = \mathbb{E}[X]$ (almost surely).

Proof. We show that $\mathbb{E}[X]$ satisfies the conditional conditions. First, it is straightforward that $\mathbb{E}[X] \in L^1(\Omega, \sigma(Y), \mathbb{P})$.

Now take Z rea lvalued, $\sigma(Y)-$measurable and bounded. Let us show that $\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X]]$. We write $\mathbb{E}[ZX] = \mathbb{E}[Z]\mathbb{E}[X]$ as Z is $\sigma(Y)-$measurable and $X \perp\!\!\!\perp Y$, then $X \perp\!\!\!\perp Z$, finishing the proof.

Now let us move back and prove the Theorem 4.3.

Proof.
Proof of Uniqueness Assume that X' and X'' satisfy the two contidions of the theorem. Take

Proof of Existence We will use some results from measure theory concerning L^2 spaces.

Assume that $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. We equip $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with a scalar product $\langle Y, Z\rangle = \mathbb{E}[YZ]$ and the norm that comes with it, so that $(L^2(\Omega, \mathcal{F}, \mathbb{P}, \|\cdot\|)$ is a normed vector space which is complete (it is a Hilbert Space).

Also $L^2(\omega, \mathcal{A}, \mathbb{P}) \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a closed subset of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. We can therefore consider the orthogonal projection of X onto $L^2(\Omega, \mathcal{A}, \mathbb{P})$. If we let X' be the orthogonal projection, it satisfies $\langle X - X', Z\rangle = 0$ for every $Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Thus, as X' is bounded, we get the two desired properties.

Now Let us go back to the L^1 case. Assume $X \in L^1(\omega, \mathcal{F}, \mathbb{P})$ with $X \geq 0$. We use a truncation argument: for $n \geq 1$ set $x_n = min(X, n)$ so that $0 \leq X_n \leq X$, set $x'_n = \mathbb{E}[X_n|\mathcal{A}]$. Because $X_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Since $X_n \leq X_{n+1}$ we get $X'_n \leq X'_{n+1}$, so $(X'_n)_{n\geq 1}$ is increasing and bounded, so we can define the limit $X' = \lim_{n\to\infty} X'_n$. We check that this X' satisfies the two properties.

First take $Z \geq 0$ real valued $\mathcal{A}-$measurable and bonuded. We have $\mathbb{E}[ZX_n] = \mathbb{E}[ZX'_n]$, but by monotone converge theorem twice, this gives us $\mathbb{E}[ZX] = \mathbb{E}[ZX']$. Moreover, since X' is an almost sure limit of $\mathcal{A}-$measurable random variables, it is $\mathcal{A}-$measurable, so $X' \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Now it sufficies to extend the characteristic property for real valued $\mathcal{A}-$measurable random variables, but it is straight from linearity by writing $Z = Z^+ - Z^-$.

Finally, we use the same argument as above to extend it from $X \geq 0$ to any real-valued.

4.3 Conditional Expectations for $[0, \infty]$-valued Random Variables

Theorem 4.8

Fix X a $[0, \infty]-$valued random variable. Let $\mathcal{A} \subset \mathcal{F}$ be a sub $\sigma-$field. Then there exists a random variable X' such that

1. X' $[0, \infty]-$valued and $\mathcal{A}-$measurable
2. For every $Z \geq 0$ $\mathcal{A}-$measurable and bounded $\mathbb{E}[ZX] = \mathbb{E}[ZX']$.

Moreover, if X'' is another such random variable, $X' = X''$ almost surely. We denote this random variable by $\mathbb{E}[X|\mathcal{A}]$ and call it the conditional expectation of X given $\mathcal{A}$.

Proof. Uniquenes follow by a similar argument as in the L^1 case.

Existence by a truncation argument: as above, we set $X_n = min(X, n)$, $X'_n = \mathbb{E}[X_n|\mathcal{A}]$ and take $X' = \lim_{n\to\infty} X'_n$.

As in L^1 we have the following properties for conditional expectations in the $[0, \infty]-$valued case:

Properties: Let X be $[0, \infty]-$valued and let $\mathcal{A} \subset \mathcal{F}$ be a $\sigma-$field.

1. $\mathbb{E}[X|\{\emptyset, \Omega\} = \mathbb{E}[X]$.
2. If X is $\mathcal{A}-$measurable $\mathbb{E}[X|\mathcal{A}] = X$.
3. If X, Y are $[0, \infty]$ valued, $a, b \geq 0$, $\mathbb{E}[aX + bY|\mathcal{A}] = a\mathbb{E}[X|\mathcal{A}] + b\mathbb{E}[Y|\mathcal{B}]$.
4. $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]] = \mathbb{E}[X]$.
5. $X \geq Y \geq 0$, then $\mathbb{E}[X|\mathcal{A}] \geq \mathbb{E}[Y|\mathcal{A}]$.
6. If $Y : \Omega \to E$ is a random variable with $X \perp\!\!\!\perp Y$ then $\mathbb{E}[X|Y] = \mathbb{E}[X]$

(almost surely is implicit in every statement).

4.4 Convergence Theorems

Theorem 4.9

Let $\mathcal{A} \subset \mathcal{F}$ be a $\sigma-$field.

1. **(Conditional Monotone Convergence)** Let $(X_n)_{n\geq 0}$ be an increaseing sequence of $[0,\infty]-$valued random variables with $X = \lim_{n\to\infty} X_n$ then $\mathbb{E}[X_n|\mathcal{A}]$ converges increasingly to $\mathbb{E}[X|\mathcal{A}]$ as $n \to \infty$ almost surely.

2. **(Conditional Fatou)** Let $(X_n)_{n\geq 1}$ be $[0,\infty]-$valued rv then $\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{A}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{A}]$.

3. **(Conditional Dominated Convergence)** Let (X_n) be a sequence of integrable random variables with
   - $X_n \xrightarrow{a.s.} X$
   - $\exists Y \geq 0$ in L^1 such that $|X_n| \leq Y$ for every $n \geq 1$.

   then $\mathbb{E}[X_n|\mathcal{A}] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[X|\mathcal{A}]$ almost surely and in L^1.

4. **(Conditional Jensen)** Let $f\colon \mathbb{R} \to \mathbb{R}_+$ be a convex function. Assume $X \in L^1$ then $f(\mathbb{E}[X|\mathcal{A}]) \leq \mathbb{E}[f(X)|\mathcal{A}]$ almost surely.

Proof. Exercise: Solve 1,2,3.

For 4, set $E_f = \{(a,b) \in \mathbb{R}^2 \colon \forall x \in \mathbb{R}, f(x) \geq a \geq ax + b\}$. Then $\forall x \in \mathbb{R}$, $f(x) = \sup_{(a,b)\in E_f}(ax+b) = \sup_{(a,b)\in E_f\cap\mathbb{Q}^2}(ax+b)$.
Then

$$\mathbb{E}[f(x)|\mathcal{A}] = \mathbb{E}[sup_{(a,b)\in E_f\cap\mathbb{Q}^2} ax + b|\mathcal{A}] \geq \mathbb{E}[ax+b|\mathcal{A}]\forall(a,b) \in E_f \cap \mathbb{Q}^2$$

Therefore

$$\mathbb{E}[f(x)|\mathcal{A}] \geq \sup_{(a,b)\in E_f\cap\mathbb{Q}^2} a\mathbb{E}[X|\mathcal{A}] + b = f(\mathbb{E}[X|\mathcal{A}])$$

where we used countability because conditional expectations are defined almost surely.

Warning! $\mathbb{E}[X]$ is defined as $\int \mathbb{P}(d\omega)$ but $\mathbb{E}[X|\mathcal{A}]$ is not defined using an integral, it is defined using the characteristic property.

4.5 Some other useful properties

There are other useful properties when we have several random variables or $\sigma-$fields.

Proposition 4.10. Let $\mathcal{A}$ be a $\sigma-$field, X, Y are random variables with X, Y $[0,\infty]$-valued or X and XY integrable. Assume that Y is $\mathcal{A}$ measurable. then

$$\mathbb{E}[XY|\mathcal{A}] = Y\mathbb{E}[X|\mathcal{A}].$$

Proof. Based on the fact that if Z is $\mathcal{A}-$measurable then YZ is also $\mathcal{A}-$measurable, which allows to show that $X' = Y\mathbb{E}[X|\mathcal{A}]$ satisfies $\mathbb{E}[X'Z] = \mathbb{E}[XYZ]$ for every Z $\mathcal{A}-$measurable, positive and bonuded.

Proposition 4.11 (Tower Property). Let $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{F}$ be $\sigma-$field. Take X a random variable with $X \in [0, \infty]$ or $X \in L^1$. Then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{A}_2]|\mathcal{A}_1] = \mathbb{E}[X|\mathcal{A}_1].$$

Proof. Let $Z \geq 0$ be $\mathcal{A}_1$ measurable and bounded. We check that $\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[\mathbb{E}[X|\mathcal{A}_2]|\mathcal{A}_1]]$.

To see this, we write $\mathbb{E}[Z\mathbb{E}[\mathbb{E}[X|\mathcal{A}_2]|\mathcal{A}_1]] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{A}_2]]]$ because Z is $\mathcal{A}_1-$measurable. But notice that Z is also $\mathcal{A}_2$ measurable, bounded and positive, so we can use the characteristic property again to conclude $\mathbb{E}[Z\mathbb{E}[X|\mathcal{A}_2]] = \mathbb{E}[ZX]$.

Hence, $\mathbb{E}[\mathbb{E}[X|\mathcal{A}_2]|\mathcal{A}_1]$ satisfies the characteristic property, finishing the proof.

Lemma 4.12. Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{F}$ be $\sigma-$fields and X random variable in $[0, \infty]$ or integrable. Assume $\mathcal{A}_2 \perp\!\!\!\perp \sigma(\sigma(X), \mathcal{A}_1)$. Then

$$\mathbb{E}[X|\sigma(\mathcal{A}_1, \mathcal{A}_2)] = \mathbb{E}[X|\mathcal{A}_1].$$

Proof. We show that $\mathbb{E}[\mathbb{1}_C X] = \mathbb{E}[\mathbb{1}_C \mathbb{E}[X|\mathcal{A}_1]]$ for every C is a generating π-system of $\sigma(\mathcal{A}_1, \mathcal{A}_2)$. Indeed, by an exercise of PSet 8, this implies the result.

We use $\{A_1 \cap A_2 \colon A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ as the generating $\pi-$system of $\sigma(\mathcal{A}_1, \mathcal{A}_2)$. Indeed, for $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$ we have $\mathbb{E}[\mathbb{1}_{A_1 \cap A_2} X] = \mathbb{E}[\mathbb{1}_{A_1}\mathbb{1}_{A_2} X]$, but $\mathcal{A}_2 \perp\!\!\!\perp \sigma(\sigma(X), \mathcal{A}_1)$, so we continue by

$$\mathbb{E}[\mathbb{1}_{A_2}\mathbb{1}_{A_1} X] = \mathbb{E}[\mathbb{1}_{A_2}]\mathbb{E}[\mathbb{1}_{A_1} X] = \mathbb{E}[\mathbb{1}_{A_2}]\mathbb{E}[\mathbb{1}_{A_1}\mathbb{E}[X|\mathcal{A}_1]] = \mathbb{E}[\mathbb{1}_{A_1 \cap A_2}\mathbb{E}[X|\mathcal{A}_1]$$

Therefore we conclude by the Dynkin Lemma.

Remark 4.13 (Approximation Toolbox).

- $Z \in \mathbb{R}, Z = Z^+ - Z^-$ with $Z^+, Z^- \geq 0$
- If $Z \geq 0$, $\exists 0 \leq Z_n \to Z$ increasingly with Z_n simple.
- $Z \in \mathbb{R}$, $Z\mathbb{1}_{|Z| \leq n} \underset{nto\infty}{\longrightarrow} Z$
- $Z \geq 0$, $Z\mathbb{1}_{Z \leq n} \to Z$ increasingly.

4 Martingales and their a.s. convergence

4.1 Definintions and first properties

We work on $\Omega, \mathcal{F}, \mathbb{P})$.

Definition 4.1. A filtration $(\mathcal{F}_n)_{n\geq 0}$ is an weakly increasing sequence of $\sigma-$fields in $\mathcal{F}$.

Interpretation: n is the time and $\mathcal{F}_n$ represents the information accesible at time n.

Definition 4.2. Let $(M_n)_{n\geq 0}$ be a sequence of real-valued random vrabiels such that $M_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ $\forall n \geq 0$ (we say that "(M_n) is adapted and integrable"). It is called a

- A $(\mathcal{F}_n)$ **martingale** if $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ $\forall n \geq 0$.
- A $(\mathcal{F}_n)$ **submartingale** if $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n$ $\forall n \geq 0$.
- A $(\mathcal{F}_n)$ **supermartingale** if $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$ $\forall n \geq 0$.

Interpretation: Imagine a player betting at a casino. M_n corresponds to her wealth at time n and $\mathcal{F}_n$ is the information the player has at time n to lace a bet and "win" an amount of $M_{n+1} - M_n$.

- (M_n) martingale: fair game
- (M_n) supermartingale: defavorable game
- (M_n) submartingale: favorable game

Remark 4.3. The definitions are always with respect to some filtration, however if (M_n) is a $(\mathcal{F}_n)$ martingale, set $\mathcal{A}_n = \sigma(M_0, \ldots, M_n)$ called canonical filtration. Then (M_n) is a $(\mathcal{A}_n)$ martingale. Indeed, this holds by the tower property.

Remark 4.4. If (M_n) is a $(\mathcal{F}_n)$ martingale, then $\mathbb{E}[M_n|\mathcal{F}_m] = M_m$ for $0 \leq m \leq n$. Indeed this holds by induction on n. For $n = m$ it clearly holds. Now for the induction step, assume $\mathbb{E}[M_n|\mathcal{F}_m] = M_n$, then

$$\mathbb{E}[M_{n+1}|\mathcal{F}_m] = \mathbb{E}[\mathbb{E}[M_{n+1}|\mathcal{F}_n]|\mathcal{F}_m] = \mathbb{E}[M_n|\mathcal{F}_m] = M_m.$$

Moreover, this implies that $\mathbb{E}[M_n] = \mathbb{E}[M_m]$ for any n, m, hence the expectation of the martingales are constant.

Very similarly:
For a submartingale $\mathbb{E}[M_n|\mathcal{F}_m] \geq M_m$ for $0 \leq m \leq n$ and $(\mathbb{E}[M_n])$ is weakly increasing
For a supermartingale $\mathbb{E}[M_n|\mathcal{F}_m] \leq M_m$ for $0 \leq m \leq n$ and $(\mathbb{E}[M_n])$ is weakly decreasing

Remark 4.5. (M_n) is a $(\mathcal{F}_n)$ supermartingale iff $(-M_n)$ is a $(\mathcal{F}_n)$ submartingale. For this reason, results are often written using either submartingales or supermartingales.

Example 4.6.

1. Random walk in $\mathbb{R}$: Fix $x \in \mathbb{R}$, and let $(X_i)_{i\geq 1}$ be iid integrable rv. Set $M_0 = x$, $M_n = x + X_1 + \ldots + X_n$ for $n \geq 1$. Let $(\mathcal{F}_n)$ be the canonical filtration. Then $\mathbb{E}[M_{m+1}|\mathcal{F}_n] = x + X_1 + \ldots + X_n + \mathbb{E}[X_{n+1}] = M_n + \mathbb{E}[X_1]$.
2. If $M \in L^1(\omega, \mathcal{F}, \mathbb{P})$, set $M_n = \mathbb{E}[M|\mathcal{F}_n]$. Then (M_n) is a $(\mathcal{F}_n)$ martingale, called a closed martingale.
3. If $M_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $M_{n+1} \leq M_n$ for every $n \geq 0$, then (M_n) is a $(\mathcal{F}_n))$ supermartingale.

Proposition 4.7. Assume that $M_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ and let $\phi \colon \mathbb{R} \to \mathbb{R}_+$ be a convex function with $\mathbb{E}[|\phi(M_n)|] < \infty$ for $n \geq 0$ then

- If (M_n) is a $(\mathcal{F}_n)$ martingale, then $(\phi(M_n))$ is a $(\mathcal{F}_n)$ submartingale.
- If (M_n) is a $(\mathcal{F}_n)$ submartingale and ϕ is weakly increasing, then $(\phi(M_n))$ is a $(\mathcal{F}_n)$ submartingale.

Sketch. Apply Jensen's inequality for conditional expectation in both cases!

Corollary 4.8. If (M_n) is a $(\mathcal{F}_n)$ martingale, then

- $(|M_n|)$ is a submartingale
- (M_n^+) is a submartingale
- If $\mathbb{E}[M_n^2] < \infty$, then (M_n^2) is a submartingale
- If (M_n) is a submartingale, then (M_n^+) is a submartingale

Proposition 4.9 (Discrete Stochastic Calculus (You can't trick the game))**.** A sequence $(H_n)_{n\geq 1}$ of real-valued random variables is called predictable if $\forall n \geq 1$, H_n is bounded and $\mathcal{F}_{n-1}$ measurable.

For a sequence $(M_n)_{n\geq 0}$ we define $(H \cdot M)_m = \sum_{k=1}^{n} H_k(M_k - M_{k-1})$.

- If $(M_n)_{n\geq 0}$ is a martingale, then $(H \cdot M)_n)$ is a martingale. In particular $\mathbb{E}[(H \cdot M)_n)] = 0$
- If (M_n) is a sub/supermartingale and $H_n \geq 0 \forall n \geq 1$ then $(H \cdot M)_n)$ is a sub/supermartingale.

Interpretation: If M_n represents the wealth of a player at time n, $M_{n+1} - M_n$ represents the amount "won" at time n, and $H_{n+1}(M_{n+1} - M_n$ represents the amount won if the player had multiplied by H_{n+1} the bet at time n.

Proof.

- $(H \cdot M)_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ by definition. We check that $\mathbb{E}[(H \cdot M)_{n+1} - (H \cdot M)_n | \mathcal{F}_n] = 0 \ \forall n \geq 0$.

  Indeed, $\mathbb{E}[(H \cdot M)_n | \mathcal{F}_n] = (H \cdot M)_n$ so this implies $\mathbb{E}[(H \cdot M)_{n+1} | \mathcal{F}_n] = (H \cdot M)_n$.

  Thus it suffices to check the identity above

  $$\mathbb{E}[(H \cdot M)_{n+1} - (H \cdot M)_n | \mathcal{F}_n] = \mathbb{E}[H_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n] = H_{n+1}\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n] = 0.$$

Ex. $\rightarrow$
- It goes similarly.

4.2 The (sub/super)martingale a.s. convergence theorem

Recall that a family $(X_i)_{i \in I}$ of random variables is bounded in L^1 if $\sup_{i \in I} \mathbb{E}[|X_i|] < \infty$.

Theorem 4.10 (Doob)
Let (M_n) be a (sub/super)martingale **bounded in** L^1. Then (M_n) converges a.s. to some real valued random variable M_∞ with $\mathbb{E}[|M_\infty|] < \infty$.

Corollary 4.11. A non-negative supermartingale or martingale converges almost surely.

We need some ingredients before starting to tackle the proof.
First we may assume that (M_n) is a supermartingale.
The main idea is to introduce the notation of upcrossing. Fix $a < b$ and set $S_1 = \inf\{n \geq 0 \colon M_n \leq a\}$, $T_1 = \inf\{n \geq S_1 \colon M_n \geq b\}$ and by induction $S_{k+1} = \inf\{n \geq T_k \colon M_n \leq a\}, T_{k+1} = \inf\{n \geq S_{k+1} \colon M_n \geq b\}$ with the convention $\inf \emptyset = \infty$.
Then for $n \geq 1$, define $N_n([a,b]) = \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k \leq n\}}$ which are the number of upcrossings of $[a,b]$ by $(M_n)_{n \geq 0}$ up to time n.

Lemma 4.12. $(M_n)_{n \geq 1}$ converges in $[-\infty, \infty]$ iff $\forall a < b, a, b \in \mathbb{Q}$, $N_\infty[a,b] < \infty$.

Lemma 4.13 (Doob Upcrossing Lemma)**.** Let (M_n) be a supermartingale. Then $\forall a < b, \forall n \geq 1$ $\mathbb{E}[N_n([a,b])] \leq \frac{1}{b-a}\mathbb{E}[(a - M_n)^+]$

Proof.
Step 1 Observe that $\forall k, n \geq 1$, $\{T_k \leq n\}, \{S_k \leq n\} \in \mathcal{F}_n$. The idea is to define $H_n = \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k < n \leq T_k\}}$ which is one iff M is in the process of doing an upcrossing at time n. Notice that this is predictable, as for each k, $\{S_k < n \leq T_k\} = \{S_k \leq n-1\} \setminus \{T_k \leq n-1\} \in \mathcal{F}_{n-1}$.

We now consider $(H \cdot M)_n$ which is a supermartingale. Write

$$\begin{aligned}
(H \cdot M)_l &= \sum_{n=1}^{l} H_n(M_n - M_{n-1}) \\
&= \sum_{n=1}^{l} \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k < n \le T_k\}}(M_n - M_{n-1}) \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{l} \mathbb{1}_{\{S_k < n \le T_k\}}(M_n - M_{n-1}) \\
&= \sum_{k=1}^{\infty} \sum_{n=S_k+1}^{\min(T_k,l)} (M_n - M_{n-1}) \\
&= \sum_{k=1}^{N_l([a,b])} (M_{T_k} - M_{S_k}) \\
&+ \mathbb{1}_{S_{N_l([a,b])+1} \le l}(M_l - M_{S_{N_l([a,b])+1}}) \\
&\ge (b-a)N_l([a,b]) - (a - M_l)^+
\end{aligned}$$

But now notice that $(H \cdot M)_l$ is a supermartingale and $\mathbb{E}[(H \cdot M)_0] = 0$, thus we get by taking expectation

$$0 \ge \mathbb{E}[(H \cdot M)_l] \ge (b-a)\mathbb{E}[N_l([a,b])] - \mathbb{E}[(a - M_l)^+]$$

from which we get the result.

Proof (Proof of the Theorem using the lemma)**.** Take $(M_n)_{n \ge 0}$ a supermartingale, bounded in L^1. Set $K = \sup_{n \ge 1} \mathbb{E}[|M_n|] < \infty$. By the "deterministic" upcrossing result, it is enough to show that $\forall a < b, a, b \in \mathbb{Q}$ almost surely $N_\infty([a,b]) < \infty$. Indeed, we then have $a.s.\forall a < b, a, b \ni \mathbb{Q} N_\infty([a,b]) < \infty$ thus almost surely (M_n) converges.

First, by the Doob upcrossing lemma, $\mathbb{E}[N_n([a,b])] \le \frac{a+K}{b-a}$ but $N_n([a,b]) \to N_\infty([a,b])$ increasingly, thus by monotone convergence,

$$\mathbb{E}[N_\infty([a,b])] = \lim_{n \to \infty} \mathbb{E}[N_n([a,b])] \le \frac{a+K}{b-a} < \infty.$$

Thus $N_\infty([a,b]) < \infty$ almost surely. This shows that $M_n \xrightarrow{a.s.} M_\infty$.

Next we show that $\mathbb{E}[|M_\infty|] < \infty$. By Fatou's Lemma:

$$\mathbb{E}[|M_\infty|] = \mathbb{E}[\liminf_{n \to \infty} |M_n|] \le \liminf_{n \to \infty} \mathbb{E}[|M_n|] \le K < \infty.$$

Remark 4.14. A (sub/super)martingale bounded in L^p with $p > 1$ is also bounded in L^1 because $\mathbb{E}[|X|] \le \mathbb{E}[|X|^p]^{1/p}$. If this is the case, it converges almost surely by Doob's theorem.

But we also seen that bounded in L^p implies uniform integrability thus $M_n \xrightarrow{\mathbb{P}} M_\infty$ and (M_n) UI, so $M_n \xrightarrow{L^1} M_\infty$.

Warning for this to hold p must be strictly greater than 1.

4.3 Example: The Bienaymé Galton-Watson branching processes

Goal: introducea simple model for the evolution of a population.

Let μ be a probability distribution on $\mathbb{N} = \{0, 1, \dots\}$. Interpretation $\mu(k)$ is the probability of having k children.

Let $(K_{n,j})_{n\geq 0, j\geq 1}$ be an iid family of $\mu-$ distributed random variables. Define by induction $X_0 = 1$ and for $n \geq 0$ $X_{n+1} = \sum_{j=1}^{X_n} K_{n,j}(w)$. Interpretation: X_n is the size of the population at generation n.

Question What is the behavior of X_n as $n \to \infty$?

To void degenerate cases, assume $\mu(0) \neq 1$, $\mu(1) \neq 1$. Our main assumption is $R = \sum_{i=0}^{\infty} i\mu(i) < \infty$. Now to define a Martingale, set $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_n = \sigma(K_{i,j} : i < n, j \geq 1)$.

Claim. $M_n = \dfrac{X_n}{R^n}$ is a $\mathcal{F}_n$ martingale

Proof. First, M_n is $\mathcal{F}_n$ measurable because the definition of X_n only involves $X_{i,j}$ for $i < n, j \geq 1$. Also, $M_n \geq 0$, so it suffices to prove it is integrable to guarantee it is $L^1(\Omega, \mathcal{F}_n, \mu)$.

This can be proved by computing $\mathbb{E}[X_{n+1}|\mathcal{F}_n]$.

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[\sum_{j=1}^{X_n} K_{n,j}|\mathcal{F}_n] = \mathbb{E}[\sum_{j=1}^{\infty} \mathbb{1}_{j\leq X_n} K_{n,j}|\mathcal{F}_n] \\ &= \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{1}_{j\leq X_n} K_{n,j}|\mathcal{F}_n] = \sum_{j=1}^{\infty} \mathbb{1}_{j\leq X_n} \mathbb{E}[K_{n,j}|\mathcal{F}_n]\end{aligned}$$

where the last inequality holds by monotone convergence and because X_n is $\mathcal{F}_n$ measurable.

Moreover, $\mathbb{E}[K_{n,j}|\mathcal{F}_n] = \mathbb{E}[K_{n,j}] = R$ because $K_{n,j} \perp\!\!\!\perp \mathcal{F}_n$ by the coalition principle, from which we conclude that

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \sum_{j=1}^{\infty} \mathbb{1}_{j\leq X_n} R = RX_n$$

from which it follows that (M_n) is a martingale and in particular, $\mathbb{E}[M_n] = 1$ for all n, hence it is also integrable, finishing the proof of the claim.

Recall that we just proved that (M_n) is bounded in L^1, hence it converges almost surely to some r.v. $M_\infty \geq 0$.

Thus

$$\frac{X_n}{R^n} \xrightarrow{a.s.} M_\infty.$$

Questions. Does this convergence also hold in L^1? Is $M_\infty > 0$.

To answer these questions, we distinguish 3 cases.

Case 1. $R < 1$ (subcritical).
In this case it is clear from the equation above that X_n will converge to 0 a.s.

Case 2. $R = 1$ (critical).
Then $X_n \xrightarrow{a.s.} M_\infty$. But because $X_n \in \mathbb{N}$, we have $M_\infty \in \mathbb{N}$, which allows us to show $\forall k \geq 1$, $\mathbb{P}(M_\infty = k) = 0$.

If $M_\infty = k \geq 1$ then for every n sufficiently large, $X_n = X_{n+1} \ldots = k$. This is very unlikely as the events $\{\sum_{j=1}^k K_{n,j} \neq k\}_{n\geq 1}$ are $\perp\!\!\!\perp$.

Let us prove there is positive probability of each of them happening. Indeed, $R = 1$, $\mu(1) \neq 1$ implies $\mu(0) > 0$, so

$$\mathbb{P}(\sum_{j=1}^n K_{n,j} \neq k) \geq \mathbb{P}(K_{n,j} = 0\colon 1 \leq j \leq k) = \mu(0)^k > 0$$

Hence, by Borel-Cantelli, we get that almost surely for infinitely many n, if $X_n = k$, then $X_{n+1} \neq k$, which contradicts our previous assumption.

So we conclude with $X_n \xrightarrow{a.s.} 0$, so almost surely $X_n = 0$ for n sufficiently large.

Moreover, If $X_n = M_n \xrightarrow{a.s.} 0$ and in particular, M_n does not converge in L^1, because $\mathbb{E}[M_n] = 1$ does not converge to $\mathbb{E}[0] = 0$.

Case 3. $R > 1$ (supercritical)
In this case, if $M_\infty > 0$, $X_n \sim M_\infty R^n$. This raises the question of whether $M_\infty > 0$.

One can show that $\mathbb{P}(\forall n \geq 0, X_n \neq 0) > 0$, but it could still be the case that $\mathbb{P}(M_\infty) = 0$.

However, if we can have $M_\infty > 0$ with positive probability, which is the case when $\sum_{k=0}^\infty k^2 \mu(k) < \infty$.

Indeed, one can then show by computing $\mathbb{E}[X_{n+1}^2 | \mathcal{F}_n]$ that $(\mathbb{E}[M_n^2])_{n\geq 1}$ is bounded.

So (M_n) is a L^2 bounded martingale, so M_n converges to M_∞ almost surely and in L^1. In particular, $\mathbb{E}[M_\infty] = 1$ which gives us $\mathbb{P}(M_\infty) > 0$.

5 Uniformly Integrable Martingales

5.1 Reminder on uniform integrability

Definition 5.1. $(X_i)_{i\in I}$ family of $\mathbb{R}-$valued is uniformly integrable (UI) if $\sup_{i\in I}\mathbb{E}[|X_i|\mathbb{1}_{|X_i|\geq k}] \underset{k\to\infty}{\longrightarrow} 0$.

We saw that this is equivalent to $\sup_{i\in I}\mathbb{E}[|X_i|]<\infty$ and $\forall\varepsilon,\exists\delta>0$ such that $\mathbb{P}(A)\leq\delta \implies \mathbb{E}[|X_i|\mathbb{1}_A]\leq\varepsilon$ for all $i\in I$ ($\varepsilon-\delta$ condition).

We saw $X_n \xrightarrow{L^1} X$ iff (X_n) UI and $X_n \xrightarrow{\mathbb{P}} X$ which is called Superdominated Convergence Theorem.

Theorem 5.2 (Strong Law of large numbers: a.s. and L^1)
Let $(X_n)_{n\geq 1}$ be iid $\mathbb{R}-$valued integrable r.v. then

$$\frac{X_1+\ldots+X_n}{n} \underset{n\to\infty}{\longrightarrow} \mathbb{E}[X_1]$$

almost surely and in L^1.

Proof. We already proved it for a.s.

For L^1 convergence, we use super dominated convergence. Indeed, set $Z_n=(X_1+\ldots+X_n)/n$. We know $Z_n \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$. It thus remains to check that Z_n is UI. We use the $\varepsilon-\delta$ condition.

First, $\mathbb{E}[|Z_n|]\leq\mathbb{E}[|X_1|]$.

Second, take $\varepsilon>0$. Since $X_1\in L^1$, the family $(X_i)_{i\geq 1}$ is UI. So we can find $\delta>0$ such that $\mathbb{P}(A)\leq\delta$ implies $\mathbb{E}[|X_i|\mathbb{1}_A]\leq\sigma$ for $i\in I$.

Now write

$$\mathbb{E}[|Z_n|\mathbb{1}_A]\leq\sum_{k=1}^{n}\frac{\mathbb{E}[|X_k|\mathbb{1}_A]}{n}\leq\sum_{k=1}^{n}\frac{\varepsilon}{n}=\varepsilon.$$

Proposition 5.3. Take $X\in L^1(\Omega,\mathcal{F},\mathbb{P})$ and $(\mathcal{A}_i)_{i\in I}$ a collection of $\sigma-$fields contained in $\mathcal{F}$. Then $(\mathbb{E}[X|\mathcal{A}_i])_{i\in I}$ is UI.

Proof. Step 1: By writting $X=X^+-X^-$ and using the fact that if $(Y_i)_{i\in I}$ and $(Z_i)_{i\in I}$ are UI, then $(Y_i-Z_i)_{i\geq I}$ is UI, we may assume that $X\geq 0$.

Step 2: Fix $\varepsilon>0$. Since $X\in L^1$ we can find $\delta>0$ such that $\mathbb{P}(A)\leq\delta \implies \mathbb{E}[X\mathbb{1}_A]\leq\varepsilon$. Now choose $k\geq\mathbb{E}[X]/\delta$ and write $\mathbb{E}[\mathbb{E}[X|\mathcal{A}_i]\mathbb{1}_{\mathbb{E}[X|\mathcal{A}_i]\geq k}]=\mathbb{E}[X\mathbb{1}_{\mathbb{E}[X|\mathcal{A}_i]\geq k}]$ by the characteristic property of conditional expectation. Now take $A=\{\mathbb{E}[X|\mathcal{A}_i]\geq k\}$, and by Markov's Inequality

$$\mathbb{P}(A)\leq\frac{1}{k}\mathbb{E}[\mathbb{E}[X|\mathcal{A}_i]]=\frac{1}{k}\mathbb{E}[X]\leq\delta.$$

Hence $\mathbb{E}[X\mathbb{1}_A] \leq \varepsilon$ which shows that $(\mathbb{E}[X|\mathcal{A}_i])_{i\in I}$ satisfies the definition of UI.

5.2 UI Martingales

Take $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}$ to be a filtration.

Theorem 5.4

Let (M_n) be a $(\mathcal{F}_n)$ martingale. The following are equivalent

1. $(M_n)_{n\geq 0}$ converges almost surely and in L^1 to a random variable denoted by M_∞.
2. $\exists X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $\forall n \geq 0$, $M_n = \mathbb{E}[X|\mathcal{F}_n]$.
3. $(M_n)_n$ is UI.

If these conditions holds, we may take $X = M_\infty$ in 2.

Proof.

2. implies 3. as we have just seen $(\mathbb{E}[X|\mathcal{F}_n])_{n\geq 0}$ is UI.

3. implies 1. If (M_n) is UI martingale, then it is bounded in L^1, so it converges a.s. to some random variable M_∞ and thus also in probability. Since it is UI, we get L^1 convergence.

1. implies 2. Fix $n \geq 1$. We know that for $p \geq n$, $\mathbb{E}[M_p|\mathcal{F}_n] = M_n$. Then write $|\mathbb{E}[M_\infty|\mathcal{F}_n] - \mathbb{E}[M_p|\mathcal{F}_n]| \leq \mathbb{E}[|M_\infty - M_p||\mathcal{F}_n]$. So $\mathbb{E}[|\mathbb{E}[M_\infty|\mathcal{F}_n] - M_n|] \leq \mathbb{E}[|M_\infty - M_p|] \underset{p\to\infty}{\longrightarrow} 0$ because $M_p \xrightarrow{L^1} M_\infty$.

We conclude $\mathbb{E}[M_\infty|\mathcal{F}_n] = M_n$.

Corollary 5.5. Take $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. The martingale $M_n = \mathbb{E}[Z|\mathcal{F}_n]$ converges a.s. and in L^1 to $M_\infty = \mathbb{E}[Z|\mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \sigma(\bigcup_{n\geq 0} \mathcal{F}_n)$

Proof. By the theorem, M_n converges a.s. and in L^1 to some r.v. M_∞, now our goal is to prove $M_\infty = \mathbb{E}[Z|\mathcal{F}_\infty]$. Let us use the defining properties of conditional expectations.

First, M_∞ is F_∞ measurable because it is a a.s. limit of M_n, hence it is $\mathcal{F}_n$ measurable for all n and thus F_∞ measurable.

Now we check it is L^1. Indeed, it is a L^1 limit of random variables, so it is in L^1.

Finally, we check the characteristic property: $\mathbb{E}[M_\infty Y] = \mathbb{E}[ZY]$ for all bounded and F_∞ measurable r.v. Y. To do this, we show that for any $\mathbb{E}[M_\infty \mathbb{1}_A] = \mathbb{E}[Z\mathbb{1}_A]$ for all $A \in \bigcup_{n\geq 0} \mathcal{F}_n$, as this is a generating $\pi-$system of $\mathcal{F}_\infty$ containing Ω (which we showed on Exercise 2 of PSet 8 is equivalent to the characteristic property).

To do this, take $A \in \mathcal{F}_n$ for fixed $n \geq 0$. Take $p \geq n$ and knowing that $\mathbb{1}_A$ is $\mathcal{F}_p$ measurable we write

$$\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_p]\mathbb{1}_A] = \mathbb{E}[M_p\mathbb{1}_A] \underset{p\to\infty}{\longrightarrow} \mathbb{E}[M_\infty\mathbb{1}_A].$$

indeed, $M_p \mathbb{1}_A \xrightarrow{L^1} M_\infty \mathbb{1}_A$ as M_p converges in L^1 to M_∞.

5.3 Optional Stopping

Motivation: If (M_n) is a martingale, $\forall n \geq 0, \mathbb{E}[M_n] = \mathbb{E}[M_0]$. But what if we stop at random times?

Definition 5.6 (Stopping Time). A r.v. $T : (\Omega, \mathcal{F}) \to \mathbb{N} \cup \{+\infty\}$ (here $\mathbb{N}$ also contains 0) is called a $(\mathcal{F}_n)$ stopping time if $\forall n \geq 0$, $\{T = n\} \in \mathcal{F}_n$.

If $T < \infty$ a.s., then we say that T is a finite stopping time.

Interpretation: In the game interpretation, stopping times are the random times at which we can decide to stop to play "without looking at the future".

Remark 5.7. T is a stopping time iff $\forall n \geq 0$ $\{T \leq n\} \in \mathcal{F}_n$ iff $\forall n \geq 0$ $\{T > n\} \in \mathcal{F}_n$.

$\{T = \infty\} = \Omega \setminus \bigcup_{n \geq 0} \{T = n\} \in \mathcal{F}_\infty$.

Example 5.8.

1. If $k \geq 0$ is a fixed constant, $T = k$ is a stopping time.
2. If X_n is $\mathcal{F}_n$ measurable, $A \in \mathcal{B}(\mathbb{R})$, then $T_A = \inf\{n \geq 0 \colon X_n \in A\}$ with the convention $\inf \emptyset = \infty$ is a stopping time, called hitting time of A.

Lemma 5.9. Let (M_n) be a $(\mathcal{F}_n)$ martingale, T be a $\mathcal{F}_n$ stopping time, then the so-called stopped process $(M_{n\wedge T})_{n\geq 0}$ with $n \wedge T = \min(n,t)$ is a $(\mathcal{F}_n)$ martingale. As a consequence, for every $n \geq 0$, $\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_0]$.

Proof. First, to be more formal on the definition of our new martingale, set $M_{n\wedge T}(\omega) = M_{n\wedge T(\omega)}(\omega)$.

For $n \geq 0$, write $M_{n\wedge T} = \sum_{j=0}^{n} \mathbb{1}_{T=j} M_j + \mathbb{1}_{T>n} M_n$. In particular, all the elements in this expression are $\mathcal{F}_n$ measurable, hence so is $M_{n\wedge T}$ and in L^1 as a finite sum of L^1 random variables.

Now we check that $\mathbb{E}[M_{(n+1)\wedge T} | \mathcal{F}_n] = M_{n\wedge T}$. Indeed, observe that if $T \leq n$, then $M_{(n+1)\wedge T} = M_{n\wedge T}$, in particular

$$\begin{aligned}
\mathbb{E}[M_{(n+1)\wedge T} - M_{n\wedge T} | \mathcal{F}_n] &= \mathbb{E}[(M_{(n+1)\wedge T} - M_{n\wedge T})\mathbb{1}_{T>n} | \mathcal{F}_n] \\
&= \mathbb{E}[(M_{n+1} - M_n)\mathbb{1}_{T>n} | \mathcal{F}_n] \\
&= \mathbb{1}_{T>n}\mathbb{E}[(M_{n+1} - M_n) | \mathcal{F}_n] \\
&= 0.
\end{aligned}$$

because M_n is a martingale. Hence we conclude $(M_{n\wedge T})$ is a $(\mathcal{F}_n)$ martingale.

Goal: Get rid of n in $\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_0]$ and hope that $\mathbb{E}[M_t] = \mathbb{E}[M_0]$.
Unfortunately, this is **false** in general.

Example 5.10. Take $(X_n)_{n\geq 1}$ iid $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ and $S_0 = 0$, $S_n = X_1 + \ldots + X_n$ for $n \geq 1$. Consider the canonical filtration, then (S_n) is a martingale.

If we set $T = \inf\{n \geq 1 \colon S_n = -1\}$ (we will later see $T < \infty$ a.s.), then $S_T = -1$, thus our goal does not hold.

The optional stopping theorem gives a condition for $\mathbb{E}[M_T] = \mathbb{E}[M_0]$ to hold.

We need the following:

Definition 5.11. Let T be a stopping time. Set $\mathcal{F}_T = \{A \in \mathcal{F} \colon \forall n \geq 0, A \cap \{T = n\} \in \mathcal{F}_n\}$

Remark 5.12. Although T is a random variable, $\mathcal{F}_T$ is not. In particular, $\mathcal{F}_T$ is a $\sigma-$field. Moreover, if $T = n$ is a constant r.v., then $\mathcal{F}_T = \mathcal{F}_n$.

Interpretation: $\mathcal{F}_T$ is the information concerning what happened until time T.

Lemma 5.13. Assume that $\forall n \geq 0$, M_n is $\mathcal{F}_n$ measurable and let T be a $(\mathcal{F}_n)$ stopping time.

1. Assume that $T < \infty$ a.s. then $M_T = \sum_{n=0}^{\infty} \mathbb{1}_{\{T=n\}} M_n$ ($M_T = 0$ if $T = \infty$) is $\mathcal{F}_T$ measurable.

2. Assume now that $M_n \xrightarrow{a.s.} M_\infty$. Then $M_T = \sum_{n=0}^{\infty} \mathbb{1}_{\{T=n\}} M_n + \mathbb{1}_{\{T=\infty\}M_\infty}$, then M_T is $\mathcal{F}_T$ measurable.

Proof. For 1., we check that $\forall n \geq 0$, $\mathbb{1}_{\{T=n}\} M_n$ is $\mathcal{F}_T$ measurable and that $\{T = \infty\}$ is $\mathcal{F}_T$ measurable.

Take $n \geq 0$, $\{T = \infty\} \cap \{T = n\} = \emptyset \in \mathcal{F}_n$.

Now take $B \in \mathcal{B}(\mathbb{R})$ and show that for $n \geq 0$, $\{\mathbb{1}_{T=n} M_n \in B\} \in \mathcal{F}_n$. Take $p \geq 0$, $\{\mathbb{1}_{T=n} M_n \in B\} \cap \{T = p\} \in \mathcal{F}_p$, but this intersection is $\emptyset$ if $p \neq n$ and $\{T = n\} \cap \{M_n \in B\}$ when $n = p$ (If $0 \notin B$), both of which are $\mathcal{F}_n$ measurable, finishing this step for B such that $0 \notin B$. For the other case just use the same result but with B^C.

Ex. $\longrightarrow$ For 2. is similar.

Theorem 5.14 (Optimal Stopping Theorem)

Let $(M_n)_{n \geq 1}$ be a UI martingale, converging a.s. and in L^1 to M_∞. Let T be a stopping time. Then $M_T = \mathbb{E}[M_\infty | \mathcal{F}_T]$.

In particular, $\mathbb{E}[M_T] = \mathbb{E}[M_\infty] = \mathbb{E}[M_0]$.

Corollary 5.15. If (M_n) is a martingale, T a finite stopping time such that $(M_{n \wedge T})_{n \geq 0}$ is UI, then $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

Tip: In practice, we use often the fact that a bounded sequence of r.v. is UI.

Proof. Recall that $M_T = \sum_{n=0}^{\infty} \mathbb{1}_{\{T=n\}} M_n + \mathbb{1}_{\{T=\infty\}} M_\infty$ we saw that M_T is $\mathcal{F}_T$ measurable.

Let us check that M_T is in L^1.

$$
\begin{aligned}
\mathbb{E}[|M_T|] &= \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}[|M_T| \mathbb{1}_{T=n}] = \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}[|M_n| \mathbb{1}_{T=n}] \\
&= \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}[|\mathbb{E}[M_\infty | \mathcal{F}_n]| \mathbb{1}_{T=n}] \leq \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}[\mathbb{E}[|M_\infty| | \mathcal{F}_n]] \mathbb{1}_{T=n}] \\
&= \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}[|M_\infty| \mathbb{1}_{T=n}] = \mathbb{E}[|M_\infty|] < \infty.
\end{aligned}
$$

Now we show that $M_T = \mathbb{E}[M_\infty | \mathcal{F}_T]$. To do this, we show that $\forall A \in \mathcal{F}_T$ $\mathbb{E}[M_T \mathbb{1}_A] = \mathbb{E}[M_\infty \mathbb{1}_A]$, from which the results follow by standard approximation approaches.

We use the same method. For $A \in \mathcal{F}_T$

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A M_T] &= \sum_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{E}[\mathbb{1}_{A\cap\{T=n\}} M_n] \\ &= \sum_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{E}[\mathbb{1}_{A\cap\{T=n\}} \mathbb{E}[M_\infty|\mathcal{F}_n]] \\ &= \sum_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{E}[\mathbb{1}_{A\cap\{T=n\}} M_\infty] = \mathbb{E}[\mathbb{1}_A M_\infty]\end{aligned}$$

Goal: Use optional stopping to study properties of simple random walks.

Assume $(X_i)_{i\geq 1}$ iid $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and define $S_0 = 0$, $S_n = \sum_{i=0}^n X_i$ for $n \geq 0$. Take $(\mathcal{F}_n)_{n\geq 0}$ to be the canonical filtration for S_n. Since $\mathbb{E}[X_1] = 0$, we know that (S_n) is a $(\mathcal{F}_n)$ martingale.

For $x \in \mathbb{Z}$, set $T_x = \inf\{n \geq 1 \colon S_n = x\}$, with the convetion that $\inf \emptyset = \infty$, which is a $(\mathcal{F}_n)$ stopping time.

Finally, for $a < 0 < b$, set $T_{a,b} = T_a \wedge T_b$.

Proposition 5.16. For $x \in \mathbb{Z}$, $a < 0 < b$ with $a, b \in \mathbb{Z}$

1. $\mathbb{P}(T_a < T_b) = \frac{b}{b-a}$.
2. a.s $T_x < \infty$.
3. $\mathbb{E}[T_{a,b}] = |a|b$
4. For $u \geq 0$, $\mathbb{E}[e^{-uT_b}] = \exp(-\cosh^{-1}(\exp(u))b)$.

Proof.

$\boxed{1.}$ $T_{a,b}$ is a stopping time, so let us check that $T_{a,b} < \infty$ a.s. We check that a.s. there exists $|a| + b$ consecutive " $+1$" in the outcomes of X_i. Indeed, this event is included in $\{T_{a,b} < \infty\}$.

The idea is to use "block" type arguments. More formally, let $(A_i)_{i\geq 1}$ be events defined by $A_1 = \{X_1 = \ldots = X_k = +1\}, \ldots, A_i = \{X_{(i-1)k+1} = \ldots X_{ik} = 1\}$. Then, by the coalition principle, $(A_i)_{i\geq 1}$ are $\perp\!\!\!\perp$ and $\mathbb{P}(A_i) = 1/2^k$, hence in particular, $\sum_{i=0}^\infty \mathbb{P}(A_i) = \infty$, thus by Borel Cantelli, a.s. (A_i) occours infinitely often. This is enough to conclude that the stopping time is finite.

Now we can resort to the Optional Stopping Theorem. $(S_{n\wedge T_{a,b}})_{n\geq 0}$ is a martingale which converges a.s. to $S_{T_{a,b}}$. But $|S_{n\wedge T_{a,b}}| < |a| + b$, thus this is a UI martingale, which then we conclude by the Optional Stopping that $\mathbb{E}[S_{T_{a,b}}] = 0$.

Finally, we can alternatively write $\mathbb{E}[S_{T_{a,b}}] = b\mathbb{P}(T_b < T_a) + a\mathbb{P}(T_a < T_b) = 0$ (they cannot be equal), which directly implies the result.

$\boxed{2.}$ Idea is to take $b \to \infty$ in 1. Indeed, since $T_b \geq b$, $T_b \underset{b\to\infty}{\longrightarrow} \infty$, and $(T_b)_{b\geq 1}$ is increasing, so $\mathbb{P}(T_a < T_b) \underset{b\to\infty}{\longrightarrow} \mathbb{P}(T_a < \infty)$. But $\frac{b}{b-a} \underset{b\to\infty}{\longrightarrow} 1$, so $\mathbb{P}(T_a < \infty) = 1$. By symmetry of taking $(-S_n)_{n\geq 1}$ we get $\mathbb{P}(T_b < \infty) = 1$.

$\boxed{3.}$ The idea is to consider the quadratic martingale $Q_n = S_n^2 - n$. $Q_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ because $|Q_n| \leq n^2 + n$. Moreover, $\mathbb{E}[Q_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_{n+1}^2 - (n+1)|\mathcal{F}_n] = S_n^2 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] - n - 1 = S_n^2 - n = Q_n$.

Hence, $(Q_{n\wedge T_{a,b}})_{n\geq 0}$ is also a martingale. However, it is not immediate to check that it is UI, so we argue directly: $\mathbb{E}[Q_{n\wedge T_{a,b}}] = 0$, thus $\mathbb{E}[S_{n\wedge T_{a,b}}^2] = \mathbb{E}[n \wedge T_{a,b}]$. The idea is to make $n \to \infty$ in this equality, which by Monotone Convergence gives us $\mathbb{E}[n \wedge T_{a,b}] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[T_{a,b}]$

Moreover, $S_{n\wedge T_{a,b}}^2 \xrightarrow{a.s.} S_{T_{a,b}}^2$ because $T_{a,b} < \infty$ a.s. In addition, $S_{n\wedge T_{a,b}}^2 < (|a| + b)^2$, hence by dominated convergence, $S_{n\wedge T_{a,b}}^2 \xrightarrow{L^1} S_{T_{a,b}}^2$

Now we conclude with $\mathbb{E}[T_{a,b}] = \mathbb{E}[S_{T_{a,b}}^2] = a^2\mathbb{P}(T_a < T_b) + b^2\mathbb{P}(T_b < T_a) = |a|b$.

$\boxed{4.}$ We show that $\mathbb{E}[(\cosh\lambda)^{-T_b}] = e^{-\lambda b}$ for $\lambda \geq 0$.

For this, the idea is to consider the so-called exponential martingale: $M_n = \frac{e^{\lambda S_n}}{(\cosh\lambda)^n}$. We know that M_n is $\mathcal{F}_n$ measurable and bounded, so it is in $L^1(\Omega, \mathcal{F}_n, \mathbb{P})$. We can easily check that it satisfies the other martingale defining property.

Now, $(M_{n\wedge T_b})_{n\geq 0}$ is a martingale which is UI because it is bounded by $e^{\lambda b}$, so we can apply optional stopping: $1 = \mathbb{E}[M_0] = \mathbb{E}[M_{T_b}] = \mathbb{E}[e^{\lambda S_{T_b}}/(\cosh\lambda)^{T_b}]$, hence $\mathbb{E}[(\cosh\lambda)^{-T_b} = e^{-\lambda b}$

(Observe that here, $(M_{n\wedge T_b})_{n\geq 0}$ is UI but $(S_{n\wedge T_b})_{n\geq 0}$ is not)

6 Martingales bounded in L^p, $p > 1$

We saw that if (M_n) is a martingale bounded in L^1, then (M_n) converges a.s., but not necessarily in L^1. In $L^p, p > 1$ the situation is different.

$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}$ a filtration.

6.1 Doob maximal inequalities

Theorem 6.1 (Doob Maximal Inequalities)

1. Let (M_n) be a submartingale. Then $\forall a > 0, n \geq 0$,
$$a\mathbb{P}(\max_{0\leq k\leq n} M_n \geq a) \leq \mathbb{E}[M_n \mathbb{1}_{\{\max_{0\leq k\leq n} M_k \geq a\}}] \leq \mathbb{E}[M_n^+]$$
2. Let (M_n) be a martingal. Set $M_n^* = \max_{0\leq k\leq n} |M_k|$. Then for $a > 0, n \geq 0$ we have
$$a\mathbb{P}(M_n^* \geq a) \leq \mathbb{E}[|M_n| \mathbb{1}_{M_n^* \geq a}] \leq \mathbb{E}[|M_n|].$$

Remark 6.2. $\boxed{2.}$ immediately follows from $\boxed{1.}$ since (M_n) martingale implies $(|M_n|)$ is a submartingale.

Proof ($\boxed{1.}$). THe idea is to introduce the stopping time $T = \inf\{k \geq 0 \colon M_k \geq a\}$, with $\inf \emptyset = \infty$ and observe that $\mathbb{P}(\max_{0\leq k\leq k} M_k \geq a) = \mathbb{P}(T \leq n)$.

Now, let us expand $a\mathbb{P}(T \leq n) = a\sum_{k=0}^n \mathbb{P}(T = k) = \sum_{k=0}^n \mathbb{E}[a\mathbb{1}_{a=k}]$, but $a\mathbb{1}_{a=k} \leq M_k \mathbb{1}_{a=k}$, thus $\sum_{k=0}^n \mathbb{E}[a\mathbb{1}_{T=k}] \leq \sum_{k=0}^n \mathbb{E}[M_k \mathbb{1}_{T=k}]$.

Moreover, (M_n) is a submartingale, so $M_k \leq \mathbb{E}[M_n|\mathcal{F}_k]$ and thus $\sum_{k=0}^n \mathbb{E}[M_k \mathbb{1}_{T=k}] \leq \sum_{k=0}^n \mathbb{E}[\mathbb{E}[M_n|\mathcal{F}_k]\mathbb{1}_{T=k}] = \sum_{k=0}^n \mathbb{E}[M_n \mathbb{1}_{T=k}] = \mathbb{E}[M_n \mathbb{1}_{T\leq n}]$ which is what we wanted.

Theorem 6.3 (Doob L^p inequalities, $p > 1$)

Fix $p > 1$.

1. Let (M_n) be a positive submartingale, then $\forall n \geq 0$
$$\mathbb{E}[(\max_{0\leq k\leq n} M_k)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[M_n^p]).$$
2. Let (M_n) be a martingale. Write $M_n^* = \max_{0\leq k\leq n} |M_k|$ then
$$\mathbb{E}[(M_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|M_n|^p]).$$

Remark 6.4. Again, $\boxed{2.}$ follows immediately from $\boxed{1.}$ as if (M_n) is martingale, then $(|M_n|)$ is a submartingale.

Before proving the theorem, we prove some useful results:

Lemma 6.5 (Holder's inequality)**.** Let $q > 1$ be such that $1/p + 1/q = 1$. Let X, Y be $\mathbb{R}-$valued r.v. with $X \in L^p$ and $Y \in L^q$, then

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Y|^q]^{1/q}.$$

Proof. First step: (Young's Inequality) for $a, b \geq 0$, it holds that $ab \leq a^p/p + b^q/q$.

Second step: We may assume that the expectations on the RHS are positive, as otherwise either X or Y would be almost surely 0, and thus the inquality is trivially true. Moreover, we may divide X by $\mathbb{E}[|X|^p]^{1/p}$ and similarly for Y and obtain $\mathbb{E}[|X|^p] = 1$, $\mathbb{E}[|Y|^q] = 1$.

Now we apply Young's inequality and conclude $|XY| \leq |X|^p/p + |Y|^q/q$, so taking expectations $\mathbb{E}[|XY|] \leq 1/p + 1/q = 1$.

Lemma 6.6 (Moment-tail)**.** Let $X \geq 0$ be a random variable, then $\forall p > 0$

$$\mathbb{E}[X^p] = p\int_0^\infty x^{p-1}\mathbb{P}(X \geq x)dx.$$

Proof. Using Fubini-Tonelli we get

$$\begin{aligned} p\int_0^\infty x^{p-1}\mathbb{P}(X \geq x)dx &= p\int_0^\infty x^{p-1}\mathbb{E}[\mathbb{1}_{X\geq x}]dx \\ &= \mathbb{E}[\int_0^\infty px^{p-1}\mathbb{1}_{X\geq x}dx] \\ &= \mathbb{E}[\int_0^\infty px^{p-1}dx] = \mathbb{E}[X^p] \end{aligned}$$

Proof (Doob's L^p Inequality)**.** If $\mathbb{E}[M_n^p] = \infty$, then there is nothing to prove. Assume $M_n \in L^p$. We further check that for $0 \leq k \leq n$ $M_k \in L^p$ as well: write $\mathbb{E}[M_k^p] \leq \mathbb{E}[(\mathbb{E}[M_n|\mathcal{F}_k])^p] \leq \mathbb{E}[\mathbb{E}[M_n^p|\mathcal{F}_k]] = \mathbb{E}[M_n^p] < \infty$ because (M_n) is a submartingale and applying conditional Jensen.

Now write $M_n^* = \max_{0\leq k\leq n} M_k$. We check $M_n^* \in L^p$ as $\mathbb{E}[(M_n^*)^p] \leq \sum_{k=0}^n \mathbb{E}[M_k^p] < \infty$.

Using the tail-moment lemma and Doob's maximal inequality

$$\begin{aligned}
\mathbb{E}[(M_n^*)^p] &= p\int_0^\infty a^{p-2}a\mathbb{P}(M_n^* \geq a)da && \text{Tail-Moment}\\
&\leq p\int_0^\infty a^{p-2}\mathbb{E}[M_n \mathbb{1}_{M_n^*\geq a}]da && \text{Doob's Max Ineq}\\
&= \mathbb{E}[\int_0^\infty pa^{p-2}M_n\mathbb{1}_{M_n^*\geq a}da] && \text{Fubini}\\
&= \mathbb{E}[pM_n\int_0^{M_n^*} a^{p-2}da]\\
&= p\mathbb{E}[M_n\frac{(M_n^*)^{p-1}}{p-1}]\\
&\leq \frac{p}{p-1}\mathbb{E}[M_n^p]^{1/p}\mathbb{E}[(M_n^*)^p]^{p-1/p} && \text{By Holder}
\end{aligned}$$

from which the conclusion holds directly.

6.2 Martingales bounded in L^p

Recall that for $p > 1$, X real-valued r.v.

$$\mathbb{E}[|X|] \leq \mathbb{E}[|X|^p]^{1/p}$$

So (X_n) bounded in L^p implies it is bounded in L^1 and $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{L^1} X$

Theorem 6.7 (L^p Martingales)
Let (M_n) be a martingale bounded in L^p, $p > 1(\sup_n \mathbb{E}[|M_n|^p] < \infty$. Then

1. M_n converges a.s. and in L^p to a random variable M_∞ with $\mathbb{E}[|M_\infty|^p] = \sum_{n\geq 0}\mathbb{E}[|M_n|^p]$
2. Setting $M_\infty^* = \sum_{n\geq 0}|M_n|$, we have

$$\mathbb{E}[(M_\infty^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|M_\infty|^p].$$

Proof. For $\boxed{1.}$ (M_n) is bounded in L^p, so it is bounded in L^1 and in particular, it converges a.s to some r.v. M_∞.

To show that this convergence also holds in L^p, we use Doob's L^p inequality: $\mathbb{E}[(M_\infty^*)^p] \leq (p/(p-1))^p \sup_{k\geq 0}\mathbb{E}[|M_k|^p]$, but M_n^* converges increasingly to M_∞^*, so by monotone convergence $\mathbb{E}[(M_n^*)^p] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[(M_\infty^*)^p]$ and we conclude that $\mathbb{E}[(M_\infty^*)^p] \leq (p/(p-1))^p \sup_{k\geq 0}\mathbb{E}[|M_k|^p]$ thus $M_\infty^* \in L^p$ and so $M_\infty \in L^p$.

Now, notice $|M_n - M_\infty|^p \xrightarrow{a.s.} 0$ and $|M_n - M_\infty|^p \leq (|M_n| + |M_\infty|)^p \leq 2^p(|M_n|^p + |M_\infty|^p) \leq 2^p(|M_\infty^*|^p + |M_\infty|^p)$, so it is bounded and we may apply dominated convergence to conclude convergence in L^p. Now, since $M_n \xrightarrow{L^p} M_\infty$ we have $\mathbb{E}[|M_n|^p] \to$

> $\mathbb{E}[|M_\infty|^p]$. Since $(|M_n|^p)$ is a submartingale, the sequence $(\mathbb{E}[|M_n|^p|])$ is nondecreasing, so the limit and the supremum coincide

Remark 6.8. If (M_n) converges in L^p, then it is bounded in L^p (true for any sequence of random variables).

7 Convergence in distribution of random variables

In a.s., $\mathbb{P}$, L^p convergence of random variables $X_n \to X$, the quantity "$X_n(\omega) - X(\omega)$" was involved, it says something about the joint realization of X_n and the limit X.

Here we define a notion of convergence for the laws of random variables.

7.1 Definition and first properties

We work with $\mathbb{R}^d-$valued random variables (but most that follows can be extended to general metric spaces).

Notation (Set of bounded continuous functions). $\mathcal{C}_b(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} \text{ cotinuous and bounded}\}$. For $f \in \mathcal{C}_b(\mathbb{R}^d)$, we write $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. Here $|\cdot|$ is any norm in $\mathbb{R}^d$.

Definition 7.1. A sequence (μ_n) of probability measures on $\mathbb{R}^d$ is said to converge weakly to a probability measure μ on $\mathbb{R}^d$ if

$$\forall f \in \mathcal{C}_b(\mathbb{R}^d), \int_{\mathbb{R}^d} f(x)\mu_n(dx) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} f(x)\mu(dx)$$

(f is called a test function).

Moreover, a sequence (X_n) of $\mathbb{R}^d-$valued r.v. is said to converge in distribution or converge in law to a $\mathbb{R}^d-$valued r.v. if $\mathbb{P}_{X_n} \to \mathbb{P}_X$ weakly, that is

$$\forall f \in \mathcal{C}_b(\mathbb{R}^d), \mathbb{E}[f(X_n)] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(X)].$$

Remark 7.2. When we say X_n converges in distribution to X, there is an abuse of notation. The limiting random variable is not uniquely defined, only its LAW is.

For this, we sometimes say that "X_n converges in distribution to μ" a probability measure.

Finally, the random variables $(X_n), X$ are not necessarily defined on the same space.

Example 7.3.

- If X_n is uniform on $\{1, 2, \ldots, n\}$, then X_n/n converges in distribution to the Uniform Law in $[0, 1]$
- Let $X_n \sim N(0, \sigma_n^2)$ with $\sigma_n \to 0$, then X_n converges in distribution to 0, i.e., to the random variable whose law is δ_0.
- If $\mu_n = \delta_{1/n}$ then $\mu_n \overset{weakly}{\longrightarrow} \delta_0$. In particular, $\mu_n(\{0\}) = 0$ and $\mu(\{0\}) = 1$.

Lemma 7.4. If $X_n \xrightarrow{(d)} X$, $X_n \xrightarrow{(d)} Y$ then $X \overset{(d)}{=} Y$

Proof. Notice that this implies that for all bounded continue function $f : \mathbb{R}^d \to \mathbb{R}$, we have $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$. To prove the desired equality, we need to establish that $\forall A \in \mathcal{B}(\mathbb{R}), \mathbb{P}_X(A) = \mathbb{P}_Y(A)$.

Let us first restrict ourselves to $F \subset \mathbb{R}^d$ closed. Indeed, we can do this by approximating $\mathbb{1}_F$ by bounded continuous functions.

Define $f_n(x) = \max(1 - nd(x, F), 0)$ then $\mathbb{E}[f_n(X)] = \mathbb{E}[f_n(Y)]$ as $f_n \in \mathcal{C}_b(\mathbb{R}^d)$. It is clear also that $f_n \overset{pointwise}{\longrightarrow} \mathbb{1}_F$ and $|f_n| \leq 1$, thus by dominated convergence twice

$$\mathbb{E}[\mathbb{1}_F(X)] \longleftarrow \mathbb{E}[f_n(X)] = \mathbb{E}[f_n(Y)] \longrightarrow \mathbb{E}[\mathbb{1}_F(Y)]$$

Thus $\mathbb{P}_X(F) = \mathbb{P}_Y(F)$. Therefore, we have two probability measures equal on a generating $\pi-$system, thus they are equal by the Dynkin Lemma.

Proposition 7.5 (Continuous Mapping)**.** Take X_n, X $\mathbb{R}^d$ valued random variables such that $X_n \overset{(d)}{\longrightarrow} X$. Take $F : \mathbb{R}^d \to \mathbb{R}^n$ continuous then $F(X_n) \overset{(d)}{\longrightarrow} F(X)$ in $\mathbb{R}^n$.

Proposition 7.6. Let X_n, X be $\mathbb{R}^d$ valued r.v. such that $X_n \overset{a.s.}{\longrightarrow} X$, $X_n \overset{L^p}{\longrightarrow} X$ $X_n \overset{\mathbb{P}}{\longrightarrow} X$, then $X_n \overset{(d)}{\longrightarrow} X$

7.2 Portemanteau Thereom

Theorem 7.7 (Portemanteau Theorem)

Let μ_n, μ be probability measures on $\mathbb{R}^d$. The following are equivalent:

1. $\mu_n \to \mu$ weakly.
2. $\forall f : \mathbb{R}^d \to \mathbb{R}$ bounded and Lipschitz, $\int f(x)\mu_n(dx) \to \int f(x)\mu(dx)$.
3. $\forall F \subset \mathbb{R}^d$ closed, $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.
4. $\forall O \subset \mathbb{R}^d$ open, $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$.
5. $\forall A \in \mathbb{R}^d$ such that $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$
6. $\forall f : \mathbb{R}^d \to \mathbb{R}$ measurable and bounded, continuous at $\mu-$almost every point (i.e. $\mu(\{x \in \mathbb{R}^d : f \text{ continuous at } x\}) = 1)$ $\int f(x)\mu_n(dx) \to \int f(x)\mu(dx)$.

Theorem (Probabilistic Formulation)
Let X_n, X be r.v. in $\mathbb{R}^d$. The following are equivalent

1. $X_n \xrightarrow{(d)} X$
2. $\forall f : \mathbb{R}^d \to \mathbb{R}$ lipschitz bounded
3. $\forall F \subset \mathbb{R}^d$ closed, $\limsup_{n\to\infty} \mathbb{P}(X_n \in f) \leq \mathbb{P}(X \in f)$
4. $\forall O \subset \mathbb{R}^d$ open, $\liminf_{n\to\infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$
5. $\forall A \subset \mathbb{R}^d$ with $\mathbb{P}(X \in \partial A) = 0$, $\mathbb{P}(X_n \in A) \to \mathbb{P}(X_A)$.
6. $\forall f : \mathbb{R}^d \to \mathbb{R}$ measurable bounded, a.s. continuous ate X, $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

Corollary 7.8 (Extended Continuous Mapping). If $X_n \xrightarrow{(d)} X$, $F : \mathbb{R}^d \to \mathbb{R}^n$ is almost surely continuous at X, then $F(X_n) \xrightarrow{(d)} F(X)$.

Proof. This comes from the fact that if $f : \mathbb{R}^n \to \mathbb{R}$ is continuous bounded, then $f \circ F \colon \mathbb{R}^d \to \mathbb{R}$ is bounded, almost surely continous at X and the result follows from $\boxed{6.}$

Example 7.9. If X_n is $\mathbb{R}$ valued and $X_n \xrightarrow{(d)} X$ with $X \neq 0$ a.s. then $1/X_n \xrightarrow{(d)} 1/X$

<u>Connection with CDF's in $\mathbb{R}$</u>

If X is a $\mathbb{R}-$valued r.v., $F_X(t) = \mathbb{P}(X \leq t)$ for $t \in \mathbb{R}$ is its CDF.

- F_X is continuous at x iff $\mathbb{P}(X = x) = 0$
- F_X has at most a countable number of discontinuity points

Theorem 7.10

Let X_n, X be a $\mathbb{R}-$valued r.v. then $X_n \xrightarrow{(d)} X$ iff $\mathbb{P}(X_n \leq t) \to \mathbb{P}(X \leq t)$ for every $t \in \mathbb{R}$ that is a continuity point of F_x.

Example 7.11. $X_n = 1/n$, $X_n \xrightarrow{(d)} 0$.

Proof. $\boxed{\Longrightarrow}$ Let $t \in \mathbb{R}$ be a continuity point of F_X, so $\mathbb{P}(X = t) = 0$. Take $A = (-\infty, t]$ in $\boxed{5.}$ of Portemanteau, $\partial A = \{t\}$ so $\mathbb{P}(X \in \partial A)\mathbb{P}(X = t) = 0$. Thus $F_{X_n}(t) = \mathbb{P}(X_n \in A) \to \mathbb{P}(X \in A) = F_X(t)$.

$\boxed{\Longleftarrow}$ We show $\boxed{4.}$ in Portemanteau, i.e. $\forall O \subset \mathbb{R}$ open, $\limsup_{n\to\infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)(\star)$

We show first that $\forall a, b \in \mathbb{R}$, $\limsup_{n\to\infty} \mathbb{P}(X_n \le b) \le \mathbb{P}(X_b)$ and $\liminf_{n\to\infty} \mathbb{P}(X_n > a) \ge \mathbb{P}(X > a)$, puttings this together $(\star)$ will hold for all open intervals.

Since F_X has at most countable number of discontinuity points, its continuity points are dense in $\mathbb{R}$, so we can choose $t > b$ with F_X continuous at t.

Then $\limsup_{n\to\infty} \mathbb{P}(X_n \le b) = \limsup_{n\to\infty} \mathbb{P}(X_n \le t) = \mathbb{P}(X \le t)$ by assumption. Now take t converge decreasingly to b and this together with the right continuity of the

Ex. $\rightarrow$ cdf implies $\mathbb{P}(X \le t) \to \mathbb{P}(X \le b)$. Finally, to prove the limsup result is very similar

Now we go back to taking $O \subset \mathbb{R}$ open. We know that we can write $O = \bigcup_{i\in I}(a_i, b_i)$ with I being at most countable and $(a_i, b)i)$ being pairwise disjoint open intervals. In particular

$$\begin{aligned}
\mathbb{P}(X \in O) = \mathbb{P}(X \in \bigcup_{i\in I}(a_i, b_i)) &= \sum_{i\in I} \mathbb{P}(X \in (a_i, b_i)) \\
&\le \sum_{i\in I} \liminf_{n\to\infty} \mathbb{P}(X_n \in (a_i, b_i)) \\
&\le \liminf_{n\to\infty} \sum_{i\in I} \mathbb{P}(X_n \in (a_i, b_i)) && \text{by Fatou} \\
&= \liminf_{n\to\infty} \mathbb{P}(X_n \in O)
\end{aligned}$$

Corollary 7.12. $X_n \xrightarrow{(d)} X$ with density p iff $\forall t \in \mathbb{R}, \mathbb{P}(X_n \le t) \to \mathbb{P}(X \le t)$ iff $\forall t \in \mathbb{R}, \mathbb{P}(X_n < t) \to \mathbb{P}(X < t) = \mathbb{P}(X \le t)$ iff $\forall a < b$ $\mathbb{P}(a \le X_n \le b) \to \int_a^b p(t)dt$.

Application 7.13. Fix $\lambda > 0$, and take $X_n \sim Geo(\frac{\lambda}{n})$, then $X_n/n \xrightarrow{(d)} Exp(\lambda)$.

Proposition 7.14. Let X_n be $\mathbb{R}^d$ valued and $a \in \mathbb{R}^d$ a constant, then $X_n \xrightarrow{(d)} a$ iff $X_n \xrightarrow{\mathbb{P}} a$

Proof. $\boxed{\Longleftarrow}$ We have already seen that convergence in probability implies convergence in distribution

$\boxed{\Longrightarrow}$ We show $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - a| \ge \varepsilon) \to 0$. Take $B(x, \varepsilon)$ to be the open ball of radius ε around x, in particular, $\mathbb{P}(|X_n - a| \ge \varepsilon) = \mathbb{P}(X_n \in B(a, \varepsilon)^c)$. Then by Portemanteau for closed sets

$$\limsup_{n\to\infty} \mathbb{P}(|X_n - a| \ge \varepsilon) \le \mathbb{P}(a \in B(a, \varepsilon)^c) = 0.$$

Theorem 7.15 (Slutsky's Theorem)

Let X_n, X, Y_n be $\mathbb{R}^d$−valued random variable, $a \in \mathbb{R}^d$ constant. Assume $X_n \xrightarrow{(d)} X$, $Y_n \xrightarrow{\mathbb{P}} a$, then $(X_n, Y_n) \xrightarrow{(d)} (X, a)$.

Application 7.16. If $a = 0$, then $X_n + Y_n \xrightarrow{(d)} X$. Indeed, $(X_n, Y_n) \xrightarrow{(d)} (X, 0)$, thus by continuous mapping $f(X_n, Y_n) \xrightarrow{(d)} f(X, 0)$ with $f(x, y) = x + y$

Moreover, if $a \neq 0$, then $X_n/Y_n \xrightarrow{(d)} X/a$, which we can prove by extended continuous mapping with $f(x, y) = x/y$ if $y \neq 0$ and 0 otherwise. One can check that f is almost surely continuus at (X, a).

Take home message: in a cv in (d) one can replace a random varirable by its limiting values when it converges in probability without changing the limit.

Warning! In general, $X_n \xrightarrow{(d)} X, Y_n \xrightarrow{(d)} Y$ does not imply $(X_n, Y_n) \xrightarrow{(d)} (X, Y)$. Indeed take X with $\mathbb{P}(X = 1) = \mathbb{P}(X = -1/2) = 1/2$ and $X_n = X$, $Y_n = -X$, then it will not hold.

We will prove later that the implication works under assumption of $\perp\!\!\!\perp$.

Lemma 7.17. Let X_n, X, Y_n be $\mathbb{R}^d-$valued. Assume $X_n \xrightarrow{(d)} X$ and $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$, then $Y_n \xrightarrow{(d)} X$

Proof. We show that $\forall F$ closed, $\limsup \mathbb{P}(Y_n \in F) \leq \mathbb{P}(X \in F)$. Define for $p \geq 1$ $F^{(1/p)} = \{x \in \mathbb{R}^d \colon d(x, F) \leq \frac{1}{p}\}$ called the $1/p-$closed enlargement of F.

$$\begin{aligned}
\mathbb{P}(Y_n \in F) &= \mathbb{P}(Y_n \in F, |X_n - Y_n| \leq \frac{1}{p}) + \mathbb{P}(Y_n \in F, |X_n - Y_n| > \frac{1}{p}) \\
&= \mathbb{P}(X_n \in F^{(1/p)}) + \mathbb{P}(|X_n - Y_n| > \frac{1}{p}).
\end{aligned}$$

So $\limsup \mathbb{P}(Y_n \in F) \leq \mathbb{P}(X \in F^{(1/p)} + 0$.

Now take $p \to \infty$ since $F^{(1/p)}$ is decreasing and $\bigcap_{p \geq 1} F^{(1/p)} = F$ as F is closed, we get $\mathbb{P}(X \in F^{(1/p)}) \underset{p \to \infty}{\longrightarrow} \mathbb{P}(X \in F)$.

Proof (Slutsky's Theorem)**.** By continuous mapping, we have $(X_n, a) \xrightarrow{(d)} (X, a)$.

Now equp $\mathbb{R}^2$ with the L^1 norm and observe that $|(X_n, a) - (X_n, Y_n)| = |Y_n - a| \xrightarrow{\mathbb{P}} 0$ by assumption, thus by the lemma $(X_n, Y_n) \xrightarrow{(d)} (X, a)$.

7.3 Restricting Test Functions

Let $\mathcal{C}_c(\mathbb{R}^d) = \{f \colon \mathbb{R}^d \to \mathbb{R}$ continuous with compact support$\}$.

Theorem 7.18

Take, μ_n, μ prob measures on $\mathbb{R}^d$. Then $\mu_n \to \mu$ weakly iff $\forall f \in \mathcal{C}_c(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} f(x) \mu_n(dx) \to \int_{\mathbb{R}^d} f(x) \mu(dx)$$

Warning! this result is specific to $\mathbb{R}^d$ and is not true for general metric spaces.

Proof. $\boxed{\Longrightarrow}$ is clear because $\mathcal{C}_c(\mathbb{R}^d) \subset \mathcal{C}_b(\mathbb{R}^d)$.

$\boxed{\Longleftarrow}$ Take $f \in \mathcal{C}_b(\mathbb{R}^d)$, let us use a truncation argument.

Take $R > 1$ and define $g_R(x) = 1$ if $|x| < R$ and $\max(R+1-|x|, 0)$ if $|x| \geq R$. Notice $fg_R \in \mathcal{C}_c(\mathbb{R}^d) \forall r \geq 1$.

For $R > 0$ fixed,

$$\begin{aligned}\left|\int f(x)\mu_n(dx) - \int f(x)\mu(dx)\right| \leq \int |f(x) - f(x)g_R(x)\mu_n(dx)| \\ + \left|\int f(x)g_R(x)\mu_n(dx) - \int f(x)g_R(x)\mu(dx)\right| \\ + \int |f(x) - f(x)g_R(x)|\mu(dx)\end{aligned}$$

Hence, taking limsup on both sides and using tkat $g_R, fg_R \in \mathcal{C}_c(\mathbb{R}^d)$

$$\begin{aligned}\limsup_{n\to\infty}\left|\int f(x)\mu_n(dx) - \int f(x)\mu(dx)\right| &\leq \limsup_{n\to\infty} \|f\|_\infty(1 - \int g_R(x)\mu_n(dx)) + 0 \\ &+ \|f\|_\infty(1 - \int g_R(x)\mu(dx)) \\ &= 2\|f\|_\infty(1 - \int g_R(x)\mu(dx))\end{aligned}$$

But $\int g_R(x)\mu(dx) \underset{R\to\infty}{\longrightarrow} 1$ by dominated convergence, finishing the proof

Corollary 7.19. Let X_n, X be $\mathbb{Z}-$valued r.v. then $X_n \xrightarrow{(d)} X$ iff $\forall k \in \mathbb{Z}$, $\mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X = k)$.

Proof. $\boxed{\Longrightarrow}$ Fix $k \in \mathbb{Z}$ and take $f_k(x) = \max(1 - |x-k|, 0)$ which is continuous and bounded, thus $\mathbb{E}[f_k(X_n)] \to \mathbb{E}[f_k(X)]$ thus $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)$

$\boxed{\Longleftarrow}$ Take $f \in \mathcal{C}_c(\mathbb{R}^d)$ and assume $\forall k \in \mathbb{Z}$, $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)$.

Write $\mathbb{E}[f(X_n)] = \sum_{j\in\mathbb{Z}} \mathbb{P}(X_n = j)f(j)$ and $\mathbb{E}[f(X)] = \sum_{j\in\mathbb{Z}} \mathbb{P}(X = j)f(j)$. To prove convergence, notice that as f has compact support, the two sums can be indexed by a finite set ($\mathbb{Z} \cap support(f)$). Then we can interchange the limit and sum over a finite set and conclude.

Application 7.20. Take $\lambda > 0$, $X_n \sim Bin(n, \frac{\lambda}{n})$, then $X_n \xrightarrow{(d)} Poi(\lambda)$. (This is the reason why the Poisson distribution is used to model rare events)

7.4 Characteristic functions and Lévy's theorem

Characteristic functions are defined as expectations of $\mathbb{C}-$valued rnadom variables. When Z is a $\mathbb{C}-$valued r.v., $\mathbb{E}[|Z|] < \infty$, we say Z is integrable and define $\mathbb{E}[Z] = \mathbb{E}[ReZ] + i\mathbb{E}[ImZ]$.

Definition 7.21. The characteristic functionn of a $\mathbb{R}^d$−valued r.v. X is defined by

$$\begin{aligned}\varphi_X : \mathbb{R}^d &\to \mathbb{C}\\ u &\mapsto \mathbb{E}[e^{i\langle X|u\rangle}].\end{aligned}$$

Remark 7.22. φ_X is well defined as $e^{i\langle X|u\rangle}$ is an integrable r.v. because its absolute value is 1.

Example 7.23. If $X \sim Poi(\lambda)$, for $u \in \mathbb{R}$, $\varphi_X(u) = \mathbb{E}[e^{iXu}] = e^{\lambda(e^{iu}-1)}$

Remark 7.24. By the transfer theorem, $\varphi_X(u) = \int_{\mathbb{R}^d} e^{i\langle x|u\rangle}\mathbb{P}_X(dx)$ for $u \in \mathbb{R}^d$. In measure theoretical terms, φ_X is the Fourier transform of $\mathbb{P}_X$.

Proposition 7.25. φ_X always satisfies the following:

- $\varphi_X(0) = 1$.
- $\varphi_X(-u) = \overline{\varphi_X(u)}$ for $u \in \mathbb{R}^d$.
- $|\varphi_X(u)| \leq \mathbb{E}[|e^{i\langle X|u\rangle}|] = 1$ for $u \in \mathbb{R}^d$
- $|\varphi_X(u+h) - \varphi_X(u)| \leq \mathbb{E}[|e^{i\langle X|h} - 1|]$ for $u, h \in \mathbb{R}^d$. In particular, φ_X is uniformly continuous.

In what follows Guassian r.v.s play a crucial role.

Example 7.26. Take $X \sim N(m, \sigma^2)$, $\varphi_X(u) = e^{imu - \frac{\sigma^2 u^2}{2}}$ for $u \in \mathbb{R}$.

Sketch. Using properties of the Gaussian, it is enough to show the result for $N(0,1)$. Since $\varphi_X(u) = \varphi_X(-u) = \overline{\varphi_X(u)}$ as $N(0,1) = -N(0,1)$ we have that $\varphi_X(u) \in \mathbb{R}$.

Thus

$$\varphi_X(u) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \cos(xu)e^{-x^2/2}dx$$

To compute this, the idea is to see that φ_X solves a differential equation.

Indeed, we have

$$\varphi_X'(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} -x\sin(xu)e^{-x^2/2}dx.$$

(Justification: use the theorem that allows to differentiate an integral depending on a parameter, which is possible because $|-x\sin(xu)e^{-x^2}| \leq xe^{-x^2/2}$ which is integrable)

Now, integration by parts gives

$$\varphi_X' u = -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} u\cos(ux)dx$$

hence $\varphi_X'(u) = -u\phi_X(u)$. But this system has the initial condition $\varphi_X(0) = 1$, which one can solve to obtain $\varphi_X(u) = e^{-u^2/2}$.

Remark 7.27. If $g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$ this example shows that $g_\sigma(z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_{1/\sigma}(u)du$ for every $z \in \mathbb{R}$.

Theorem 7.28

Let X, Y be r.v. in $\mathbb{R}^d$, $\varphi_X = \varphi_Y$ iff X and Y have the same law.

Proof. To simplify, assume $d = 1$ and that X, Y are defined in the same probability space.

$\boxed{\Longleftarrow}$ If $\mathbb{P}_X = \mathbb{P}_y$, this is directly true by the transfer theorem.

$\boxed{\Longrightarrow}$ Idea: use a small gaussian perturbation. More precisely let Z_n be a $N(0, 1/n)$ r.v. $\perp\!\!\!\perp X, Y$.

The idea now is to show $\varphi_X = \varphi_Y$ implies $X + Z_n \overset{(d)}{=} Y + Z_n$ $(\star)$. Indeed, assume this hows, let us see how we can conclude.

$\mathbb{E}[Z_n^2] = 1/n$, so $Z_n \xrightarrow{L^2} 0$ thus $Z_n \xrightarrow{\mathbb{P}} 0$. Thus $X + Z_n \xrightarrow{\mathbb{P}} X + 0$. Thus $X + Z_n \xrightarrow{(d)} X$, similarly $Y + Z_n \xrightarrow{(d)} Y$, from which the conclusion holds.

Now we must go back to show $(\star)$. For this, we show that for $F \colon \mathbb{R} \to \mathbb{R}_+$ measurable, $\mathbb{E}[F(X + Z_n)] = \mathbb{E}[F(Y + Z_n)]$.

We will prove that $\mathbb{E}[F(X+Z_n)] = \mathbb{E}[F(Y+Z_n)]$ only depends on ϕ_X. Let's compute

$$\begin{aligned}
\mathbb{E}[F(X + Z_n)] &= \int_{\mathbb{R}} \mathbb{P}_X(dx) \left(\int_{\mathbb{R}} F(x+z) \mathbb{P}_{Z_n}(dz) \right) && \text{Transfer+Fubini} \\
&= \mathbb{E}[\int_{\mathbb{R}} F(x+z) g_{1/\sqrt{n}}(z) dz] \\
&= \mathbb{E}[\int_{\mathbb{R}} F(z) g_{1/\sqrt{n}}(z - X) dz] \\
&= \int_{\mathbb{R}} F(z) dz \mathbb{E}[g_{1/\sqrt{n}}(z - X)]. && \text{Fubini-Tonelli}
\end{aligned}$$

Now we look at $\mathbb{E}[g_{1/\sqrt{n}}(z - X)]$. But we know that $g_{1/\sqrt{n}}(z - X) = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iu(z-X)} g_{\sqrt{n}}(u)du$.

Taking the expectation and using Fubini-Lebesgue

$$\mathbb{E}[g_{\sqrt{\frac{1}{n}}}(z - X)] = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathbb{E}[e^{-iux}] e^{-iuz} g_{\sqrt{n}}(u) du$$

but $\mathbb{E}[e^{-iux}] = \varphi_X - u$, thus $\mathbb{E}[F(X + Z_n)]$ only depends on $\varphi_X = \varphi_Y$ and the equality holds.

Important Consequence: For $X_1, \ldots, X_k$ $\mathbb{R}-$valued r.v.s, $X_1, \ldots, X_k$ $\perp\!\!\!\perp$ iff $\forall u_1, \ldots, u_k \in \mathbb{R}$

$$\varphi_{(X_1, \ldots, X_k)}(u_1, \ldots, u_k) = \varphi_{X_1}(u_1) \ldots \varphi_{X_k}(u_k)$$

Proof. $\boxed{\Longrightarrow}$ We have seen that for $X_1, \ldots, X_k$ $\perp\!\!\!\perp$ and $f_1, f_2, \ldots, f_k$ integrable, $\mathbb{E}[f_1(X_1) \ldots f_k(X_k)] = \mathbb{E}[f_1(X_1)] \ldots \mathbb{E}[f_k(X_k)]$, from which it follows.

$\boxed{\Longleftarrow}$ We have seen that $(X_1, \ldots, X_k)$ has the same characteristic function as $\mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_k}$. Thus $\mathbb{P}_{(X_1,\ldots,X_k)}$ and $\mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_k}$ have the same characteristic function and they are equal.j

Application 7.29. Take $X \sim N(m_1, \sigma_1^2)$, $Y \sim N(m_2, \sigma_2^2)$. Assume $X \perp\!\!\!\perp Y$, then $X + Y \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Rule of thumb:

- characteristic functions are often well adapted when we have sums of $\perp\!\!\!\perp$ r.v.
- cdfs are often adapted when we have r.v. defined using min, max.

Theorem 7.30 (Lévy)

Take X_n, X $\mathbb{R}^d$−valued r.v. then $X_n \xrightarrow{(d)} X$ iff $\varphi_{X_n} \to \varphi_X$ pointwise, i.e. $\forall u \in \mathbb{R}^d$, $\varphi_{X_n}(u) \underset{n\to\infty}{\longrightarrow} \varphi_X(u)$

Proof. We assume $d = 1$ to simplify.

$\boxed{\Longrightarrow}$ If $X_n \xrightarrow{(d)} X$, observe that for $u \in \mathbb{R}$, $f(x) = e^{iux}$ is continious bounded, so $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(x)]$ which is what we want.

$\boxed{\Longleftarrow}$ We use the idea of a small gaussian perturbation $Z_k \sim N(0, 1/k^2)$ with $Z_k \perp\!\!\!\perp X_n$, $Z_k \perp\!\!\!\perp X$. Assuming φ_{X_n} converges pointwise to φ_X, we have two steps:

Step 1 Show that for $k \geq 1$ fixed, $X_n + Z_k \xrightarrow{(d)} X + Z_k$.

Step 2 Conclude that $X_n \xrightarrow{(d)} X$.

Let us deal with step 2 assuming step 1 first. By Portemanteau, it is enough to show that $\forall f\colon \mathbb{R} \to \mathbb{R}$ L−Lipschitz, we have $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$. To do this, write

$$\begin{aligned}
|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq \mathbb{E}[|f(X_n) - f(X_n + Z_k)|] \\
&\quad + |\mathbb{E}[f(X_n + Z_k)] - \mathbb{E}[f(X + Z_k)]| + \mathbb{E}[|f(X) - f(X + Z_k)|] \\
&\leq 2L\mathbb{E}[|Z_k|] + |\mathbb{E}[f(X_n + Z_k)] - \mathbb{E}[f(X + Z_k)]|
\end{aligned}$$

Thus

$$\limsup_{n\to\infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2L\mathbb{E}[|Z_k|] = \frac{2L}{k}\mathbb{E}[|Z_1|] \to 0.$$

Now we shall prove step 1.

Recall that for $g_\sigma(x) = \frac{1}{\sigma\sqrt{n}} e^{-x^2/2\sigma^2}$ and $Z_k \sim N(0, 1/k^2) \perp\!\!\!\perp X$, for $F \geq 0$

$$\mathbb{E}[F(X + Z_k)] = \int_{\mathbb{R}} dz F(z) \left(\frac{k}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iuz} g_k(u) \varphi_X(-u) du \right). \tag{$\star$}$$

We take $f\colon \mathbb{R} \to \mathbb{R}$ continuous with compact support and show

$$\mathbb{E}[f(X_n + Z_k)] \to \mathbb{E}[f(X + Z_k)].$$

We know that φ_{X_n} converges pointwise to φ_X and will use this with dominated convergence twice.

First, $e^{iuz}g_k(u)\varphi_{X_n}(-u) \to e^{iuz}g_k(u)\varphi_X(u)$ and $|e^{iuz}g_k(y)\varphi_{X_n}(-u)| \leq g_k(u) \in L^1(du)$. Hence by dominated convergence

$$f(z)\int_{\mathbb{R}} e^{iuz}g_k(u)\varphi_{X_n}(-u)du \to f(z)\int_{\mathbb{R}} e^{iuz}g_k(u)\varphi_X(-u)du.$$

Second, let us prove that the expression above is bounded. Indeed

$$\left|f(z)\frac{k}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{iuz}g_k(u)\varphi_{X_n}(-u)du\right| \leq |f(Z)|\frac{k}{\sqrt{2\pi}}\int_{\mathbb{R}} g_k(u)du \in L^1(dz)$$

because f is continuous with compact support. We conclude by $\star$.

Application 7.31. Take X_n, Y_n, X, Y $\mathbb{R}-$valued r.v. Assume $X_n \xrightarrow{(d)} X$, $Y_n \xrightarrow{(d)} Y$ and $X_n \perp\!\!\!\perp Y_n \forall n \geq 1$. Then $(X_n, Y_n) \xrightarrow{(d)} (X, Y)$ where X, Y $\perp\!\!\!\perp$.

Proof. We show $\varphi_{(X_n,Y_n)} \to \varphi_{(X,Y)}$ pointwise in $\mathbb{R}^2$ with X, Y $\perp\!\!\!\perp$. Take $(u_1, u_2) \in \mathbb{R}^2$, then

$$\begin{aligned}\varphi_{(X_n,Y_n)}(u_1,u_2) &= \mathbb{E}[e^{i(u_1X_n+u_2Y_n)}] = \varphi_{X_n}(u_1)\varphi_{Y_n}(u_2)\\ &\underset{n\to\infty}{\longrightarrow} \varphi_X(u_1)\varphi_Y(u_2) = \varphi_{(X,Y)}(u_1,u_2).\end{aligned}$$

from which we conclude by Lévy's theorem.

Remark 7.32. If μ, ν are two probability measures on $\mathbb{R}$, a **coupling** of μ, ν is a r.v. (X, Y) with $Law(X) = \mu, Law(Y) = \nu$.

Application 7.33. Take $0 \leq p \leq q \leq 1$. Then for every $0 \leq k \leq n$,

$$\mathbb{P}(Bin(n,q) \geq k) \geq \mathbb{P}(Bin(n,p) \geq k).$$

Proof. Take $U_1, \ldots, U_n$ iid $Uni([0,1])$ r.v. Define $Y_n = \sum_{k=1}^n \mathbb{1}_{U_k \leq q}$ $X_n = \sum_{k=1}^n \mathbb{1}_{U_k \leq p}$, then $X_n \sim Bin(n,p), Y_n \sim Bin(n,q)$ but $Y_n \geq X_n$, which yields the result.

7.5 Central Limit Theorem

Theorem 7.34

Let $(X_i)_{i\geq 1}$ be a sequence of iid $\mathbb{R}-$valued r.v.s with $\mathbb{E}[X_1^2] < \infty$. Set $\sigma = Var(X_1)$ and assume $\sigma > 0$. Then

$$\frac{X_1 + \ldots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \xrightarrow{(d)} N(0,1).$$

Remark 7.35.

- $\sigma > 0$ rules out the case of constant r.v
- Since
$$\frac{X_1 + \ldots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X_1]\right))$$
this tells that when $\mathbb{E}[X_1^2] < \infty$ the "speed" of convergence in the SLN is of order $1/\sqrt{n}$

Lemma 7.36. Assume that X is a $\mathbb{R}-$valued with $\mathbb{E}[X^2] < \infty$ then

$$\varphi_X(t) = 1 + i\mathbb{E}[X_1]t - \frac{\mathbb{E}[X^2]}{2}t^2 + o(t^2).$$

Proof. $\varphi_X(t) = \mathbb{E}[e^{itX}]$. This comes from Taylor's formular as φ_X is twice differentiable at 0.

Indeed, we use the following result from measure theory (essentially consequence of dominated convergence): if $\forall t \in \mathbb{R}$, $F(t, X) \in L^1$, a.s. $t \mapsto F(t, X)$ is differentiable, $\exists Y \in L^1$ s.t. a.s $\forall t \in \mathbb{R}$ $\left|\frac{\partial}{\partial t}F(t, X)\right| \leq Y$, then $t \mapsto \mathbb{E}[F(t, X)]$ is differentiable and

$$\frac{d}{dt}\mathbb{E}[F(t, X)] = \mathbb{E}[\frac{d}{dt}F(t, X)].$$

We use this result with $F(t, x) = e^{itx}$

Proof (Central Limit Theorem). Up to replacing X_i with $X_i - \mathbb{E}[X_1]$, we can assume $\mathbb{E}[X_1] = 0$, so $\sigma^2 = \mathbb{E}[X_i^2]$.

We use Lévy's theorem and the lemma

$$\begin{aligned}
\varphi_{\frac{X_1+\ldots+X_n}{\sigma\sqrt{n}}}(t) &= \mathbb{E}\left[e^{i\frac{(X_1+\ldots+X_n)}{\sigma\sqrt{n}}t}\right] = \prod_{i=1}^{n}\mathbb{E}[e^{i\frac{t}{\sigma\sqrt{n}}X_i}] \qquad \text{by } \perp\!\!\!\perp \\
&= \varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)^n \\
&= \left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^2\varepsilon\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \\
&= \left(1 - \frac{t^2}{2n} + \frac{t^2}{\sigma n}\varepsilon\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n
\end{aligned}$$

Now we use a trick to avoid using ln of complex numbers: for $u, v \in \mathbb{C}$, $|u^n - v^n| \leq n|u - v|\max(|u|^{n-1}, |v|^{n-1})$. We get

$$\left|\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)^n - \left(1 - \frac{t^2}{2n}\right)^n\right| \leq n\frac{t^2}{\sigma n}\varepsilon\left(\frac{t}{\sigma\sqrt{n}}\right) \underset{n\to\infty}{\longrightarrow} 0.$$

And

$$\left(1 - \frac{t^2}{2n}\right)^n = \exp(n \ln(1 - t^2/2n)) \to \exp(-t^2).$$

So we conclude

$$\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)^n \xrightarrow[n\to\infty]{} \exp(-t^2/2) = \varphi_{N(0,1)}(t)$$

and thus the theorem holds by Lévy's theorem.

Obs. If $\forall t \in \mathbb{R}$ $\mathbb{P}(X_n \leq t)$ has a limit as $n \to \infty$, this does not imply X_n converge in distribution to X.

Take for example $X_n = n$, $\mathbb{P}(X_n \leq t) \to 0$, but 0 is not a cdf of a random variable.

Obs. X_n $\mathbb{R}-$valued, $\forall t \in \mathbb{R}$, $\phi_{X_n}(t) = \mathbb{E}[e^{itX_n}]$ has a limit as $n \to \infty$ does not imply X_n converge in distribution

Take for example $X_n \sim N(0, n^2)$. It is clear that its characteristic function converges, however if we take $a < b$, then $\mathbb{P}(a < X_n < b) = \mathbb{P}(a/n < N(0,1) < b/n) \to 0$. Indeed, if we argue by contradiction assuming it converges in distribution to X, we could pick $a < b$ such that $\mathbb{P}(a < X < b) \geq 1/2$.

Remark 7.37 (Improved Lévy Theorem)**.** Assume $\varphi_{X_n}(t) \xrightarrow[n\to\infty]{} f(t)$ $\forall t \in \mathbb{R}$, then X_n converges in distribution iff f is continuous at 0.

7.6 Gaussian vectorrs and the multidimensional CLT

Definition 7.38. A r.v. $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ is a **gaussian vector** if any linear combination of its cordinates is a gaussian r.v with the convention $N(m, 0) = m$ constant.

Recall that if $X \sim N(0, \sigma^2)$, $\mathbb{E}[e^{itx}] = e^{itm - \sigma^2 \frac{t^2}{2}}$.

Example 7.39.

- If $X_1, \ldots, X_n$ $\perp\!\!\!\perp$ Gaussian r.v. then $(X_1, \ldots, X_d)$ is a gaussian vector.
- If X, Y are $\perp\!\!\!\perp$ Gaussian, $(X, X + Y)$ is a gaussian vector

Warning! If $(X_1, \ldots, X_d)$ is a gaussian vector, then $X_1, \ldots, X_d$ are gaussian, but the converse is false.

Indeed take $X \sim N(0,1)$ and $\varepsilon \sim \pm 1$ with probability $1/2$, $\perp\!\!\!\perp$ X. We can check that $(X, \varepsilon X)$ is not a gaussian vector since $\mathbb{P}(X + \varepsilon X) = 1/2$, so $X + \varepsilon X)$ is not gaussian.

Definition 7.40. Let $X = (X_1, \ldots, X_d)$ be a gaussian vector.

- $m_X = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$ is called the mean of X.
- $K_X = (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])_{1 \leq i,j \leq d} \in \mathcal{M}_{d\times d}(\mathbb{R})$ is the covariance matrix of X.

- X is centered if $m_X = (0, 0, \dots, 0)$.

Proposition 7.41. Let X be a gaussian vector in $\mathbb{R}^d$. Take $\lambda \in \mathbb{R}^d$, then $\langle \lambda, X \rangle = N(m_\lambda, \sigma_\lambda^2)$ with $m_\lambda = \langle m_X, \lambda \rangle$, $\sigma_\lambda^2 = \langle \lambda, K_X \lambda \rangle$.

Corollary 7.42. $\forall \lambda \in \mathbb{R}^d$, $\langle \lambda, K_X \lambda \rangle \geq 0$. Thus K_X is a positive semi-definite matrix.

Corollary 7.43. The characteristic function of a gaussian vector X is given by $\Phi_X(\lambda) = \mathbb{E}[e^{i\langle \lambda, X \rangle}] = \exp(i\langle \lambda, m_X \rangle - \frac{1}{2}\langle \lambda, K_X, \lambda \rangle)$ for $\lambda \in \mathbb{R}^d$. Indeed this is a straight consequence of $\langle \lambda, X \rangle$ being gaussian.

Since characteristic functions characterize laws, the law of a gaussian vector X is characterized by m_X, K_X.

Application 7.44. If $X, Y \perp\!\!\!\perp$ gaussian vectors, $X + Y$ is a gaussian vector with $m_{X+Y} = m_X + m_Y$ and $K_{X+Y} = K_X + K_Y$.

Remark 7.45. One can show that if K is $d \times d$ positive semi-definite and $m \in \mathbb{R}^d$, then there exists a gaussian vector X with mean m and covariance matrix K.

Theorem 7.46

1. Let $X = (X_1, \dots, X_d)$ be a gaussian vector in $\mathbb{R}^d$. Then $(X_1, \dots, X_d)$ are $\perp\!\!\!\perp$ iff K_X is diagonal (i.e. $\forall i \neq j \mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$).
2. Let $Z = (X_1, \dots, X_p, Y_1, \dots, Y_q)$ be a gaussian vector in $\mathbb{R}^{p+q}$ then $(X_1, \dots, X_p) \perp \perp (Y_1, \dots, Y_q)$ iff $\forall 1 \leq i \leq p, 1 \leq j \leq q$, $\mathbb{E}[X_i Y_j] = \mathbb{E}[X_i]\mathbb{E}[Y_j]$.

Take home message: for gaussian vectors independence is equivalent to 0 covariance.

Theorem 7.47 (Multidimensional CLT)
Let $(X^i)_{i \geq 1}$ be iid r.v. in $\mathbb{R}^d$. Assume $\mathbb{E}[|X^1|] < \infty$. Then

$$\frac{X^1 + \ldots + X^n - n\mathbb{E}[X^1]}{\sqrt{n}} \xrightarrow{(d)} N(0, K_{X^1}).$$

Sketch. Similar to $d = 1$, based on characteristic function and taylor expansion of φ_{X^1} at 0.

8 A glimpse of statistical theory

Outline:

1. Estimators
2. Confidence interval

So far, we use sequences $(X_i)_{i\geq 1}$ of r.v. with known laws. In statistical theory, it is different: we observe a sequence of values (which we often assume to be the realization of an iid sequence of r.v.) called **sample** but with unknown law.

Goal: Use the sample to estimate the unknown law or decide to accept or reject some hypothesis on it.

8.1 Estimators

In practice, it often happens that the unkown law belongs to a certain family of probability measures depending on a parameter θ.

For example: a company would like to sell a product and the goal is to estimate the proportion $\theta \in [0,1]$ of people susceptible of buying the product.

Definition 8.1. A statistical model is a space Ω with a $\sigma-$field $\mathcal{F}$ and a family $(P_\theta)_{\theta\in\Theta}$ of probability measures on it, where Θ is the space of parameters.

Example 8.2.

- $\Theta = [0,1]$ and P_θ is the law of $Ber(\theta)$.
- $\Theta = (0,\infty)$ and P_θ is the law of $Exp(\theta)$.
- $\Theta = \mathbb{R}\times\mathbb{R}_+$ and $P_{(m,\sigma^2)}$ is the law of $N(m,\sigma^2)$.

Definition 8.3. A sample of size n of a probability measure P is a sequence $X_1,\dots,X_n$ of r.v. $\perp\!\!\!\perp$ with law P.

An **estimator** is a function d with values in Θ which depends on the sample, i.e. of the form $d(X_1,\dots,X_n)$. It is **unbiased** if $\forall\theta\in\Theta$, $\mathbb{E}_\theta[d(X_1,\dots,X_n)] = \theta$. (when $\Theta\subset\mathbb{R}^+$, $\mathbb{E}_\theta$ denotes the expectation with respect to P_θ).

It is strongly consistent if for $\theta\in\Theta$, under P_θ, $d(X_1,\dots,X_n) \xrightarrow{a.s.} \theta$.

In practice, we often view data as the realization of r.v that are independent under P_θ with θ unknown.

Example 8.4. In the model $\Theta = [0,1]$, P_θ the law of $Ber(\theta)$, then $d(X_1,\dots,X_n) = \frac{X_1+\dots+X_n}{n}$ is an unbiased, strongly consistent estimator of θ.

8.2 Confidence intervals

In practice, we do not just give a numerical estimation of a parameters, but also a "small" interval in which the parameter lies with given probability.

Definition 8.5 (Confidence interval)**.** Fix a confidence level $1-\alpha$ with $\alpha \in (0,1)$ representing the "error" allowed. A **confidence interval** of level $1-\alpha$ is an interval $I(X_1,\dots,X_n) = [a(X_1,\dots,X_n), b(X_1,\dots,X_n)]$ such that

$$P_\theta(\theta \in I(X_1,\dots,X_n)) \geq 1-\alpha \ \ \forall \theta \in \Theta.$$

We hope to have large $1-\alpha$ with a small confidence interval, but generally these are antagonistic.

Example 8.6. In the model $\Theta = [0,1]$, P_θ the law of $Ber(\theta)$, take again $d(X_1,\dots,X_n) = \frac{X_1+\ldots+X_n}{n}$. For $\theta \in \Theta$

$$P_\theta(|d(X_1,\dots,X_n)-\theta| \geq \varepsilon) \leq \frac{Var(\frac{X_1+\ldots+X_n}{n})}{\varepsilon^2} = \frac{\theta(1-\theta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Therefore, for α fixed, we can pick $\varepsilon = \frac{1}{\sqrt{4n\alpha}}$ to obtain a confidence interval of level $1-\alpha$.

An **asymptotic** confidence interval $I(X_1,\dots,X_n)$ satisfies $\forall \theta \in \Theta$

$$\liminf_{n\to\infty} P_\theta(\theta \in I(X_1,\dots,X_n)) \geq 1-\alpha$$

The central limit theorem often gives such intervals. Indeed, assume $Z_n \xrightarrow{(d)} N(0,1)$, then $\forall a < b, \mathbb{P}(a < Z_n < b) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(a < N(0,1) < b)$.

Hence choosing q_α with $\mathbb{P}(|N(0,1)| > q_\alpha) = \alpha$ we get $\mathbb{P}(|Z_n| > q_\alpha) \underset{n\to\infty}{\longrightarrow} \alpha$.

Indeed, if we apply this to the Bernoulli example, we get $I(X_1,\dots,X_n) = [\overline{X_n} - q_\alpha \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}}, \overline{X_n} + q_\alpha \frac{\sqrt{\theta(1-\theta)}}{\sqrt{n}}]$. The problem here is that the interval cannot depend on θ. We could fix ths by bounding $\theta(1-\theta)$.

However, another interesting solution would be to replace θ by a strongly consistent estimator $(\overline{X_n})$. Indeed, by Slutsky's theorem, we also have

$$\frac{\sqrt{n}}{\sqrt{\overline{X_n} - \overline{X_n}^2}}(\overline{X_n} - \theta) \xrightarrow{(d)} N(0,1)$$

so we can repeat similar steps we get $I(X_1,\dots,X_n) = [\overline{X_n} - q_\alpha \frac{\sqrt{\overline{X_n} - \overline{X_n}^2}}{\sqrt{n}}, \overline{X_n} + q_\alpha \frac{\sqrt{\overline{X_n} - \overline{X_n}^2}}{\sqrt{n}}]$