Probability Theory

Francisco Moreira Machado

October 27, 2023

1 Chapter 1 - $\sigma-$fields and Measures

1.1 $\sigma-$fields

Definition. Let Ω be a set. A $\sigma-$field $\mathcal{A}$ 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 not a $\sigma-$field.

Remark 1 (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 (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 → **Remark 2.** 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 not be confused with the usual notion of lim sup and lim inf for sequences of real numbers.

Proposition 1.2. Let $(A_i)_{i\in I}$ be a collection of σ−fields on Ω (I not necessarily countable). Then, $\bigcap_{i\in I} A_i$ is itself a σ−field.

Proof. It suffices to check the three properties of σ−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 $A^c \in \bigcap_{i\in I} \mathcal{A}_i$.
3. Similar reasoning

Exercise →

1.1.1 Generated σ−field

Definition 2. 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 σ−field generated by $\mathcal{C}$.

Notice that the generated σ−field by $\mathcal{C}$ is indeed a σ−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** σ−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 3. If $\mathcal{C}$ is a σ−field, then $\sigma(\mathcal{C}) = \mathcal{C}$.

Proposition 1.3. If $\mathcal{C} \subset \mathcal{C}'$ then $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$.

Example 1.4 (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 3 (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 σ−algebra $\mathcal{C}_{cyl}$ is defined to be the $\sigma - \mathtt{field}$ generated by the cylinders.

Example 1.5. $\{(1,1,\ldots)\} \in \mathcal{C}_{cyl}$ because it is the same set as $\bigcap_{n\geq 1} \mathcal{C}_{\underbrace{1,\ldots,1}_{n \text{ times}}}$

Example 1.6. Take $\Omega = \mathbb{R}$ and $\mathcal{A} = \sigma(\{x\}, x \in \mathbb{R})$, one can check that $\mathcal{A} = \{A \subset \mathbb{R} \colon A$ or A^c is countable$\}$.

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

Definition 4. Borel $\sigma-$field If (E, d) is a metric space (take $E = \mathbb{R}$), the **Borel $\sigma-$field** is $\sigma(\{U \colon U \subset E, U \text{ open set}\})$. It is denoted by $\mathcal{B}(E)$ or $\mathcal{B}_E$. It is also the $\sigma-$field generated by closed sets.

Example 1.7. 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 5 (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

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

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

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

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 1.8. 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 $(A_i)_{i\geq 1}$ are measurable and $A_1 \subset A_2 \dots$ then $\mu(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} \mu(A_n)$.
3. If $(A_i)_{i\geq 1}$ are measurable and $A_1 \supset A_2 \dots$ and $\mu(A_1) < \infty$ then $\mu(\bigcap_{n\geq 1} A_n) = \lim_{n\to\infty} \mu(A_n)$.
4. If $(A_i)_{i\geq 1}$ are measurable, $\mu(\bigcup_{n\geq 1} A_n) \leq \sum_{n\geq 1}^{\infty} \mu(A_n)$.

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

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

4. Since $B \setminus A \cap B \subset B$, we have $\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \leq \mu(A) + \mu(B)$. 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 1.9 (The Counting Measure)**.** The cardinality on a set E is defined by $\mathrm{Card}(B)$ and can be used when E is finite or countable

Example 1.10 (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 1.11 (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 4. 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 σ–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 1.12.
- 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 7. 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 σ–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 σ–field

Lemma 1.13. 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 σ–field.

Proof. It suffices to prove the last condition of a σ–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 σ–field.

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

Definition 8. If $\mathcal{C} \subset \mathcal{P}(\Omega)$ is a class of subsets of Ω, we set

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

It is called the <u>Dynkin system generated by $\mathcal{C}$</u>.

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 σ−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 σ−field, but $\sigma(\mathcal{C})$ is the smallest one containing all others, which would finish the proof.

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

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

- <u>$\mathcal{C} \in \lambda_A$:</u> If $B \in \mathcal{C}$, we have $A \cap B \in \lambda(\mathcal{C})$ due to stability under finite intersection.
- <u>Dynkin system</u>
  - $\Omega \in \lambda_A$ 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 → <u>Second:</u> 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 9. 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 $\underline{\pi-\text{system}}$ if it is stable by finite intersections.
We say that $\mathcal{C}$ is a $\underline{\text{generating } \pi-\text{system}}$ if $\sigma(\mathcal{C}) = \mathcal{A}$.

Example 1.14. $\{(-\infty, a) : a \in \mathbb{R}\}$ is generating of $\mathbb{B}(\mathbb{R})$.

Example 1.15. For $\Omega = \{0,1\}^{\mathbb{N}}$ cylinder sets are generating $\pi-$system of the cylinder $\sigma-$field.

Corollary 1.16. Let $(\Omega, \mathcal{A})$ be a measurable space, $\mathcal{C}$ a generating $\pi-$system.

1. Let μ, ν be two $\underline{\text{finite}}$ 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 1.17 (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.
$\underline{\text{Goal:}}$ $\mu(A) = \nu(A) \forall A \in \mathcal{A}$.
To do that, take

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

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

$\underline{\text{Interpretation:}}$ 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 1.18. 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 1.19. Throw one die $\Omega = [6]$ with even probabilities. Then $A = \{1, 2\}$ and $B = \{1, 3, 5\}$ are independent.

Definition 10. 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 5. Independence is relative to $\mathbb{P}$. Moreover in general pairwise independence does not imply independence.

Proposition 1.20. Events $A_1, \ldots, A_n$ are $\perp\!\!\!\perp$ iff $\mathbb{P}(B_1 \cap \ldots, B_n) = \mathbb{P}(B_1) \ldots \mathbb{P}(A_n)$, 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 11. Let $\mathcal{B}_1, \ldots, \mathcal{B}_n \subset \mathcal{A}$ be $\sigma-$fields. They are independent ($\perp\!\!\!\perp$) if $\forall B_1 \in \mathcal{B}_1, \ldots, \forall B_n \in \mathcal{B}_n$,

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

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 1.21. Let $\mathcal{B}_1, \ldots, \mathcal{B}_n \subset \mathcal{A}$ 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

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

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

Application 1.22 (Coalition Principle)**.** Let $\mathcal{B}_1, \ldots, \mathcal{B}_n \subset \mathcal{A}$ independent $\sigma-$fields. Fix $1 \leq n_1 < n_2 \ldots \leq n_p = n$, then $\mathcal{D}_1 = \sigma(\mathcal{B}_1, \ldots, \mathcal{B}_{n_1})$, $\mathcal{D}_{i+1} = \sigma(\mathcal{B}_{n_i+1}, \ldots, \mathcal{B}_{n_{i+1}})$ for $i < p$ are all $\perp\!\!\!\perp$.

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

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

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

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

$\sigma(\mathcal{B}_1, \ldots, \mathcal{B}_{n_1}) \subset \sigma(\mathcal{C})$

$\mathcal{C}_1$ 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 $\forall C_1 \in \mathcal{C}_1, \ldots, \forall C_p \in \mathcal{C}_p$,

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

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

Definition 12 (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 1.23 (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 $\sum_{n=1}^{\infty} A_n = \infty$ 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 Chapter 2: Random Variables

2.1 Measurable Function

Definition 13. 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 $\rightarrow$

Proposition 2.1. 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 $\rightarrow$

Definition 14 (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 6. 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 15 (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 2.2 (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 2.3 (Uniform Distribution). Assume that the law of X is the Lebesgue measure on $[0,1]$

Proposition 2.4. 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 $F\colon\mathbb{R}\longrightarrow[0,1]$ 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. $\longrightarrow$ 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. $\longrightarrow$ **Remark 7.** 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 $f\colon E\longrightarrow(F,\mathcal{F})$ is a function we set $\sigma(f)=\{f^{-1}(B)\colon B\in\mathcal{F}\}$. It is a $\sigma-$field

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

Similarly if $(f_i)_{i\in I}$ is a collection of functions $f_i\colon E\longrightarrow(F_i,\mathcal{F}_i)$ 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 2.5. $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Then $\sigma(f) = \{A \in \mathcal{B}(\mathbb{R}) : A = -A\}$.

Proposition 2.6.

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

Proof. We check that $f: (E, \sigma(f)) \to (F, \mathcal{F})$ is measurable. This is indeed true by definition of $\sigma(f)$. Assume now that $f: (E, \mathcal{E}) \to (F, \mathcal{F})$ 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 2.7. Let E, F be metric spaces. Let $f: E \to F$ be continuous, then $f: (E, \mathcal{B}(E)) \to (F, \mathcal{B}(F))$ 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 $\mathcal{C} = \{O : O \subset F, \text{ open}\}$, 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 σ–fields and families of functions

Product σ–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 16 (Product σ–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 $\Pi_i: E \to E_i$ is the projection onto the i–th coordinate called the canonical projections.

Example 2.8. $E = \{0, 1\}^{\mathbb{N}}$, then $\Pi_j: E \to \{0, 1\}, \Pi_j((x_i)_{i \in I}) = x_j$.

Example 2.9. $E = \prod_{i \in [0,1]} \mathbb{R} = \{f : [0, 1] \to \mathbb{R}\}$ is the space of functions from $[0, 1]$ to $\mathbb{R}$.

Definition 17 (Product σ–field or Cylinder σ–field)**.** We define $\otimes\mathcal{E}_i = \sigma(\Pi_i : i \in I)$ to be the smallest σ–field on $\prod_{i\in I} E_i$ for which the canonical projections are measurable.

Definition 18 (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 π–system of $\otimes_{i\in I}\mathcal{E}_i$

Proposition 2.10. 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 σ–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 σ–field.

Definition 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 σ–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 2.11. The Lebesgue measure on $\mathbb{R}^n$.

Remark 8. If $\mathcal{C}_i$ is a generating π–system of $\mathcal{E}_i$, then $\{A_1\times\ldots\times\times A_n : A_i \in \mathcal{C}_i\}$ is a generating π–system of $\otimes\mathcal{E}_i$.

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

Example 2.12. 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 2.13.

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

Remark 9. 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 10. 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 2.14.

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 21. 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 <u>but</u> when $X_1, \dots, X_n$ $\perp\!\!\!\perp$.

Lemma 2.15 (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 22 (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 2.16 (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. $\to$ Take $\mathcal{C}_1 = \sigma(X_1, \dots, X_p)$ and $\mathcal{C}_2 = \bigcup_{k=p+1}^{\infty} \sigma(X_{p+1}, \dots, X_k)$. Check that this works.

Application 2.17. 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 2.18. The two random variables $(X_i)_{i \in I}$ and $(Y_i)_{i \in I}$ with values in $\prod_{i \in I} E_i$ and $\prod_{i \in I} F_i$ are $\perp\!\!\!\perp$ iff

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

Definition 23. If $(X_i)_{i \geq 1}$ are random variables we set $B_n = \sigma(X_K \colon k \geq n)$ 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 2.19. If $(X_i)_{i \geq 1}$ are $\mathbb{R}-$valued rv. Set $S_n = X_1 + \dots + X_n$ then $\{\sup_{n \geq 1} S_n = +\infty\} \in B_\infty$

Theorem 2.20 (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 $\mathcal{D}_n = \sigma(X_1, \dots, X_n)$, 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 $\bigcup_{n \geq 1} \mathcal{D}_n = \bigcup_{n \geq 1} \sigma(X_1, \dots, X_n)$ is a generating $\pi-$system of $\sigma(X_i \colon i \geq 1)$. Thus

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

Finally, observe that $B_\infty \subset \sigma(X_n \colon n \geq 1)$, 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 2.21. Let $f_n \colon (E, \mathcal{E}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})$ 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 $\mathcal{B}(\overline{\mathbb{R}})$, this shows that f is measurable.

Definition 24 (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 2.22

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 2.23 (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 11. 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 25 (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 26 (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 27 (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 2.24. If $0 \leq f \leq g \leq \infty$

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

2.5.2 Monotone Convergence

Theorem 2.25
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 2.26 (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 2.27.

1. If $f, g \geq 0$, $a, b \geq 0$, $\int(af + bf)d\mu = a\int fd\mu + b\int gd\mu$
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 2.28.

- If we use δ_a, the diract 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 2.29 (Fatou Lemma)
Let $f_n \geq 0$ be measurable functions then

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

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 2.30. 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 2.31 (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 28. 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 2.32 (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 2.33 (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 29 (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 30 (Bernoulli). $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$.
Interpretation Rigged coing giving heads with probability p.

Definition 31 (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 32 (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 33 (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 12 (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 34. 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 35 (Uniform law). $a < b$, $p(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$.

Definition 36 (Exponential law of parameter $\lambda > 0$). $p(x) = \lambda e^{-\lambda x}\mathbb{1}_{x \geq 0}$.

Definition 37 (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 2.34. 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 38 (Density in $\mathbb{R}^n$). Take $p\colon \mathbb{R}^n \to \mathbb{R}^+$ with $\mathbb{P}_{\mathbb{R}^n} p(x)dx = 1$. $X = (X_1, \dots, X_n)$ with values in $\mathbb{R}^n$ has density p if

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

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

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

2.7 Independence and Integration

Theorem 2.35 (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.

Application 2.36. 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 2.37. 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 2.38. 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 2.39

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

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

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

$$\begin{aligned}\mathbb{E}[f(X_1,\ldots,X_n)] &= \int_{\mathbb{R}^N} f(x_1,\ldots,x_n)\mathbb{P}_{(X_1,\ldots,X_n)}(dx_1dx_2\ldots dx_n)\\ &= \int_{\mathbb{R}^N} f(x_1,\ldots,x_n)\mathbb{P}_{X_1}(dx_1)\otimes\ldots\otimes\mathbb{P}_{X_n}(dx_n) \text{ by } \perp\!\!\!\perp\\ &= \int_{\mathbb{R}^n} f(x_1,\ldots,x_n)p_1(x_1)\ldots p_n(x_n) \text{ by Fubini-Tonelli}\end{aligned}$$

Theorem 2.40

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

Application 2.41. 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 2.42. 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 2.43

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

1. $X_1, \dots, X_n$ are $\perp\!\!\!\perp$
2. $\forall f_i : E_i \rightarrow \mathbb{R}_+$ 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 2.44. 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 $f_i : E_i \rightarrow \mathbb{R}$ 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 2.45.

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 $Var(X_1 + \dots X_n) = Var(X_1) + \dots + Var(X_n)$.

3 Sequences and Series of Independent Random Variables

Goal Study limits of $X_1 + \dots + X_n$ as $n \to \infty$ where $X_1 \dots, X_n$ 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 $\frac{X_1 + \dots X_n}{n}$ 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. $\rightarrow$ Item 2. goes simularly, 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

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

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

Application 3.5. Let $(A_i)_{i\geq 1}$ be independent events with same probability p, then

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

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 $(Z_k)_{1\leq k\leq n}$ be $\perp\!\!\!\perp$ real-valued r.v. in L^2. Set $S_K = Z_1 + \dots Z_k$ for $1 \leq k \leq n$. Assume tat $\mathbb{E}[Z_K] = 0$ for every $1 \leq k \leq n$. Then $\forall \lambda > 0$

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

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

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

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

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

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

observe that $2S_k\mathbb{1}_{A_k}$ is $\sigma(Z_1,\dots Z_k)$-measurable and $(S_n - S_k)$ is $\sigma(Z_{k+1},\dots,Z_n)$-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 2 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. $\sum_{n\geq 1}\mathrm{Var}(Z_n) < \infty$.

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

Remark 14. We do not assume that (Z_k) have the same law. In fact, if this was the case, for any $\mathrm{Var}(Z_1) > 0$, 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 $\mathrm{Var}(Z_n - \mathbb{E}[Z_n]) = \mathrm{Var}(Z_n)$, 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:

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

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

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

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 $A_m = \{\forall n \geq m, |S_n - S_m| \leq 1/k\}$. 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

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

which is enough to conclude!

3.4 Three Series Theorem

Theorem 3.8 (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. $\sum_{k=1}^{\infty} (X_k \mathbb{1}_{|X_k| < a}) < \infty$

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

Remark 15. $X_k \mathbb{1}_{|X_k| > a}$ is bounded random variable so it is in L^2.

Remark 16. 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 17. 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 3.9
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 3.10 (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

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

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

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

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

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

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