Part II — Probability and Measure

Based on lectures by Dr Sarkar

Michaelmas 2023

Contents

§0 Holes in classical theory

Analysis

1. What is the "volume" of a subset of $\mathbb{R}^d$.
2. Integration (Riemann Integration has holes)

- $\{f_n\}$ a sequence of continuous functions on $[0,1]$ s.t.
  - $0 \leq f_n(x) \leq 1 \; \forall \; x \in [0,1]$.
  - $f_n(x)$ is monotonically decreasing on $n \to \infty$, i.e. $f_n(x) \geq f_{n+1}(x) \; \forall \; x/$

  So, $\lim_{n\to\infty} f_n(x)$ exists. But f is not Riemann integrable . We want a theory of integration s.t. f is integrable and $\lim_{n\to\infty} \int_0^1 f_n(x)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x$.

3. $L^1 = ()$ If $f \in L^1$ is f Riemann integrable? Will have to change the definition of integral. L^2 a hilbert space

Probability

1. Discrete probability has its limitations,
   - Toss a unbiased coin 5 times. What is the probability if getting 3 heads?
   - Take an infinite sequence of coin tosses ($E = \{0,1\}^{\mathbb{N}}$ which is uncountable) and an event A that depends on that infinite sequence. How do you define $\mathbb{P}(A)$? E.g. $X_i \sim$ Ber $\left(\frac{1}{2}\right)$ and $A = \frac{\sum_{i=1}^n X_i}{n}$, the average number of heads. By strong law of large numbers $\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{n} \to \frac{1}{2}\right) = 1$.
   - How to draw a point uniformly at random from $[0,1]$? $U \sim U[0,1]$. Probability needs axioms to be made rigorous.
2. Define Expectation for a r.v.. Also would want the following if $0 \leq X_n \leq 1$ and $X_n \downarrow X$ then $\mathbb{E}X_n \to \mathbb{E}X$.

§1 Introduction

Notation. $A_n \uparrow A$ means that the sequence A_n is increasing ($A_1 \subseteq A_2 \subseteq \dots$) and $\bigcup_n A_n = A$.

§1.1 Definitions

Definition 1.1 (σ-algebra)

Let E be a (nonempty) set. A collection $\mathcal{E}$ of subsets of E is called a **σ-algebra** if the following properties hold:

- $\varnothing \in \mathcal{E}$;
- $A \in \mathcal{E} \implies A^c = E \setminus A \in \mathcal{E}$;
- if $(A_n)_{n\in\mathbb{N}}$ is a countable collection of sets in $\mathcal{E}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}$.

Example 1.1

Let $\mathcal{E} = \{\varnothing, E\}$. This is a σ-algebra. Also, $\mathcal{P}(E) = \{A \subseteq E\}$ is a σ-algebra.

Remark 1. Since $\bigcap_n A_n = \left(\bigcup_n A_n^c\right)^c$, any σ-algebra $\mathcal{E}$ is closed under countable intersections as well as under countable unions. Note that $B \setminus A = B \cap A^c \in \mathcal{E}$, so σ-algebras are closed under set difference.

Definition 1.2 (Measurable Space and Set)

A set E with a σ-algebra $\mathcal{E}$ is called a **measurable space**. The elements of $\mathcal{E}$ are called **measurable sets**.

Definition 1.3 (Measure)

A **measure** μ is a set function $\mu : \mathcal{E} \to [0, \infty]$, such that $\mu(\varnothing) = 0$, and for a sequence $(A_n)_{n\in\mathbb{N}}$ such that the A_n are disjoint, we have

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

This is the **countable additivity** property of the measure.

Remark 2. $(E, \mathcal{E}, \mu)$ is a measure space.

Remark 3. If E is countable, then for any $A \in \mathcal{P}(E)$ and measure μ, we have

$$\mu(A) = \mu\left(\bigcup_{x \in A} \{x\}\right) = \sum_{x \in A} \mu(\{x\})$$

Hence, measures are uniquely defined by the measure of each singleton.

Define $m : E \to [0, \infty]$ s.t. $m(x) = \mu(\{x\})$, such an m is called a "mass function", and measures μ are in 1-1 correspondence with the mass function m. This corresponds to the notion of a probability mass function.

Here $\mathcal{E} = \mathcal{P}(E)$ and this is the theory in elementary discrete prob. (when $\mu(\{x\}) = 1\ \forall\, x \in E$, μ is called the counting measure. Here $\mu(A) = |A|\ \forall\ A \subset E$).

For uncountable E however, the story is not so simple and $\mathcal{E} = \mathcal{P}(E)$ is generally not feasible. Indeed measures are defined on σ-algebra "generated" by a smaller class $\mathcal{A}$ of simple subsets of E.

Definition 1.4 (Generated $\sigma-$algebra)

For a collection $\mathcal{A}$ of subsets of E, we define the σ-algebra $\sigma(A)$ **generated by $\mathcal{A}$** by

$$\sigma(\mathcal{A}) = \{A \subseteq E \colon A \in \mathcal{E} \text{ for all } \sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A}\}$$

So it is the smallest σ-algebra containing $\mathcal{A}$. Equivalently,

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{E} \supseteq \mathcal{A},\ \mathcal{E} \text{ a } \sigma\text{-algebra}} \mathcal{E}$$

Question

Why is $\sigma(A)$ a σ-algebra? See Sheet 1, Q1.

§1.2 Rings and algebras

The class $\mathcal{A}$ will usually satisfy some properties too, let E be a set and $\mathcal{A}$ a collection of subsets of E. To construct good generators, we define the following.

Definition 1.5 (Ring)

$\mathcal{A} \subseteq \mathcal{P}(E)$ is called a **ring** over E if $\varnothing \in \mathcal{A}$ and $A, B \in \mathcal{A}$ implies $B \setminus A \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.

Rings are easier to manage than σ-algebras because there are only finitary operators.

Definition 1.6 (Algebra)

$\mathcal{A}$ is called an **algebra** over E if $\varnothing \in \mathcal{A}$ and $A, B \in \mathcal{A}$ implies $A^c \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.

Remark 4. Rings are closed under symmetric difference $A \triangle B = (B \setminus A) \cup (A \setminus B)$, and are closed under intersections $A \cap B = A \cup B \setminus A \triangle B$. Algebras are rings, because $B \setminus A = B \cap A^c = (B^c \cup A)^c$. Not all rings are algebras, because rings do not need to include the entire space.

The idea:

- Define a set function on a suitable collection $\mathcal{A}$.
- Extend the set function to a measure on $\sigma(\mathcal{A})$. (Carathéodory's Extension theorem)
- Such an extension is unique. (Dynkin's Lemma)

Goal: Start with a "measure" on $\mathcal{A}$ that has some nice properties and then extend it to $\sigma(A)$.

Definition 1.7 (Set Function)

A **set function** on a collection $\mathcal{A}$ of subsets of E, where $\varnothing \in \mathcal{A}$, is a map $\mu \colon \mathcal{A} \to [0, \infty]$ such that $\mu(\varnothing) = 0$.

- We say μ is **increasing** if $\mu(A) \leq \mu(B)$ for all $A \subseteq B$ in $\mathcal{A}$.
- We say μ is **additive** if $\mu(A \cup B) = \mu(A) + \mu(B)$ for disjoint $A, B \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.
- We say μ is **countably additive** if $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ for disjoint sequences A_n where $\bigcup_n A_n$ and each A_n lie in $\mathcal{A}$.
- We say μ is **countably subadditive** if $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$ for arbitrary sequences A_n under the above conditions.

Remark 5. If μ is countably additive set function on $\mathcal{A}$ and $\mathcal{A}$ is a ring then μ satisfies all the previous listed properties.

Proposition 1.1 (Disjointification of countable unions)

Consider $\bigcup_n A_n$ for $A_n \in \mathcal{E}$, where $\mathcal{E}$ is a σ-algebra (or a ring, if the union is finite). Then there exist $B_n \in \mathcal{E}$ that are disjoint such that $\bigcup_n A_n = \bigcup_n B_n$.

Proof. Define $\widetilde{A}_n = \bigcup_{j \leq n} A_j$, then $B_n = \widetilde{A}_n \setminus \widetilde{A}_{n-1}$. □

Remark 6. A measure satisfies all four of the above conditions. Countable additivity implies the other conditions. Proof on Sheet 1.

Theorem 1.1 (Carathéodory's theorem)

Let μ be a countably additive set function on a ring $\mathcal{A}$ of subsets of E. Then there exists a measure $\mu^\star$ on $\sigma(\mathcal{A})$ such that $\mu^\star|_\mathcal{A} = \mu$.

We will later prove that this extended measure is unique.

Proof. For $B \subseteq E$, we define the *outer measure* $\mu^\star$ as

$$\mu^\star(B) = \inf\left\{\sum_{n\in\mathbb{N}} \mu(A_n) : A_n \in \mathcal{A}, B \subseteq \bigcup_{n\in\mathbb{N}} A_n\right\}$$

If there is no sequence A_n such that $B \subseteq \bigcup_{n\in\mathbb{N}} A_n$, we declare the outer measure $\mu^\star(B)$ to be ∞. Clearly, $\mu^\star(\varnothing)$ and $\mu^\star$ is increasing, so $\mu^\star$ is an increasing set fcn on $\mathcal{P}(E)$.

Definition 1.8 ($\mu^\star$ measurable)

A set $A \subseteq E$ $\mu^\star$ **measurable** if $\forall\, B \subseteq E$ $\mu^\star(B) = \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$.

We define the class

$$\mathcal{M} = \{A \subseteq E : A \text{ is } \mu^\star \text{ measurable}\}$$

We shall show that M is a σ-algebra that contains $\mathcal{A}$, $\mu^\star\,|_M$ is a measure on M that extends μ (i.e. $\mu^\star|_\mathcal{A} = \mu$).

Step 1. $\mu^\star$ is countably sub-additive on $\mathcal{P}(E)$: It suffices to prove that for $B \subseteq E$ and $B_n \subseteq E$ such that $B \subseteq \bigcup_n B_n$ we have

$$\mu^\star(B) \leq \sum_n \mu^\star(B_n) \tag{\dagger}$$

We can assume without loss of generality that $\mu^\star(B_n) < \infty$ for all n, otherwise there is nothing to prove. For all $\varepsilon > 0$ there exists a collection $A_{n,m} \in \mathcal{A}$ such that $B_n \subseteq \bigcup_m A_{n,m}$ and

$$\mu^\star(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu(A_{n,m})$$

as we took an infimum. Now, since $\mu^\star$ is increasing, and $B \subseteq \bigcup_n B_n \subseteq \bigcup_n \bigcup_m A_{n,m}$, we have

$$\mu^\star(B) \leq \mu^\star\left(\bigcup_{n,m} A_{n,m}\right) \leq \sum_{n,m} \mu(A_{n,m}) \leq \sum_n \mu^\star(B_n) + \sum_n \frac{\varepsilon}{2^n} = \sum_n \mu^\star(B_n) + \varepsilon$$

Since ε was arbitrary in the construction, (†) follows by construction.

Step 2. $\mu^\star$ extends μ: Let $A \in \mathcal{A}$, and we want to show $\mu^\star(A) = \mu(A)$.

We can write $A = A \cup \varnothing \cup \dots$, hence $\mu^\star(A) \leq \mu(A) + 0 + \dots = \mu(A)$ by definition of $\mu^\star$.

If $\mu^\star$ is infinite, there is nothing to prove.

We need to prove the converse, that $\mu(A) \leq \mu^\star(A)$. For the finite case, suppose there is a sequence A_n where $\mu(A_n) < \infty$ and $A \subseteq \bigcup_n A_n$. Then, $A = \bigcup_n (A \cap A_n)$, which is a union of elements of the ring $\mathcal{A}$. As μ is countably additive on $\mathcal{A}$ and $\mathcal{A}$ is a ring, μ is countably subadditive on $\mathcal{A}$ and increasing by remark 6. Hence $\mu(A) \leq \sum_n \mu(A \cap A_n) \leq \sum_n \mu(A_n)$. Since the A_n were arbitrary taking the infimum over A_n, we have $\mu(A) \leq \mu^\star(A)$ as required.

Step 3. $\mathcal{M} \supseteq \mathcal{A}$: Let $A \in \mathcal{A}$. We must show that for all $B \subseteq E$, $\mu^\star(B) = \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$.

We have $B \subseteq (B \cap A) \cup (B \cap A^c) \cup \varnothing \cup \dots$, hence by countable subadditivity (†), $\mu^\star(B) \leq \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$.

It now suffices to prove the converse, that $\mu^\star(B) \geq \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$.
We can assume $\mu^\star(B)$ is finite, and so $\forall\ \varepsilon > 0\ \exists\ A_n \in \mathcal{A}$ s.t. $B \subseteq \bigcup_n A_n$ and $\mu^\star(B) + \varepsilon \geq \sum_n \mu(A_n)$. Now, $B \cap A \subseteq \bigcup_n (A_n \cap A)$, and $B \cap A^c \subseteq \bigcup_n (A_n \cap A^c)$. All of the members of these two unions are elements of $\mathcal{A}$, since $A_n \cap A^c = A_n \setminus A$. Therefore,

$$\begin{aligned}\mu^\star(B \cap A) + \mu^\star(B \cap A^c) &\leq \sum_n \mu(A_n \cap A) + \sum_n \mu(A_n \cap A^c) \\ &\leq \sum_n \left[\mu(A_n \cap A) + \mu(A_n \cap A^c)\right] \\ &\leq \sum_n \mu(A_n) \leq \mu^\star(B) + \varepsilon\end{aligned}$$

Since ε was arbitrary, $\mu^\star(B) = \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$ as required.

Step 4. $\mathcal{M}$ is an algebra: Clearly $\varnothing$ lies in $\mathcal{M}$, and by the symmetry in the definition of $\mathcal{M}$, complements lie in $\mathcal{M}$. We need to check $\mathcal{M}$ is stable under finite intersections. Let $A_1, A_2 \in \mathcal{M}$ and let $B \subseteq E$. We have

$$\begin{aligned}\mu^\star(B) &= \mu^\star(B \cap A_1) + \mu^\star(B \cap A_1^c) \text{ as } A_1 \in M \\ &= \mu^\star(B \cap A_1 \cap A_2) + \mu^\star(B \cap A_1 \cap A_2^c) + \mu^\star(B \cap A_1^c) \text{ taking } \tilde{B} = B \cap A_1\end{aligned}$$

We can write $A_1 \cap A_2^c = (A_1 \cap A_2)^c \cap A_1$, and $A_1^c = (A_1 \cap A_2)^c \cap A_1^c$. Hence

$$\mu^\star(B) = \mu^\star(B \cap A_1 \cap A_2) + \underbrace{\mu^\star(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^\star(B \cap (A_1 \cap A_2)^c \cap A_1^c)}_{\mu^\star(B \cap (A_1 \cap A_2)^c) \text{ as } A_1 \in M}$$

$$= \mu^\star(B \cap A_1 \cap A_2) + \mu^\star(B \cap (A_1 \cap A_2)^c)$$

which is the requirement for $A_1 \cap A_2$ to lie in $\mathcal{M}$.

Step 5. $\mathcal{M}$ is a σ-algebra and $\mu^\star$ is a measure on $\mathcal{M}$:
It suffices now to show that $\mathcal{M}$ has countable unions and the measure respects these countable unions. Let $A = \bigcup_n A_n$ for $A_n \in \mathcal{M}$. Without loss of generality, let the A_n be disjoint. We want to show $A \in \mathcal{M}$, and that $\mu^\star(A) = \sum_n \mu^\star(A_n)$.

By (†), we have for any $B \subseteq E$ $\mu^\star(B) \leq \mu^\star(B \cap A) + \mu^\star(B \cap A^c) + 0 + \dots$ so we need to check only the converse of this inequality. Also, $\mu^\star(A) \leq \sum_n \mu^\star(A_n)$, so we need only check the converse of this inequality as well. Similarly to before,

$$\begin{aligned}
\mu^\star(B) &= \mu^\star(B \cap A_1) + \mu^\star(B \cap A_1^c) \\
&= \mu^\star(B \cap A_1) + \mu^\star(B \cap \underbrace{A_1^c \cap A_2}_{A_2 \text{ as } A_1, A_2 \text{ disjoint}}) + \mu^\star(B \cap A_1^c \cap A_2^c) \\
&= \mu^\star(B \cap A_1) + \mu^\star(B \cap A_2) + \mu^\star(B \cap A_1^c \cap A_2^c) \\
&= \mu^\star(B \cap A_1) + \mu^\star(B \cap A_2) + \mu^\star(B \cap A_1^c \cap A_2^c \cap A_3) + \mu^\star(B \cap A_1^c \cap A_2^c \cap A_3^c) \\
&= \mu^\star(B \cap A_1) + \mu^\star(B \cap A_2) + \mu^\star(B \cap A_3) + \mu^\star(B \cap A_1^c \cap A_2^c \cap A_3^c) \\
&= \cdots \\
&= \sum_{n \leq N} \mu^\star(B \cap A_n) + \mu^\star(B \cap A_1^c \cap \dots \cap A_N^c)
\end{aligned}$$

Since $\bigcup_{n \leq N} A_n \subseteq A$, we have $\bigcap_{n \leq N} A_n^c \supseteq A^c$. $\mu^\star$ is increasing, hence, taking limits,

$$\mu^\star(B) \geq \sum_{n=1}^\infty \mu^\star(B \cap A_n) + \mu^\star(B \cap A^c)$$

By (†),

$$\mu^\star(B) \geq \mu^\star(B \cap A) + \mu^\star(B \cap A^c)$$

as required. Hence $\mathcal{M}$ is a σ-algebra. For the other inequality, we take the above result for $B = A$.

$$\mu^\star(A) \geq \sum_{n=1}^\infty \mu^\star(A \cap A_n) + \mu^\star(A \cap A^c) = \sum_{n=1}^\infty \mu^\star(A_n)$$

So $\mu^\star$ is countably additive on $\mathcal{M}$ and is hence a measure on $\mathcal{M}$. □

§1.3 Uniqueness of extension

To address uniqueness of extension, we introduce further subclasses of $\mathcal{P}(E)$. Let $\mathcal{A}$ be a collection of subsets of E.

Definition 1.9 (π-system)

A collection $\mathcal{A}$ of subsets of E is called a **π-system** if $\varnothing \in \mathcal{A}$ and $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$.

Definition 1.10 (d-system)

A collection $\mathcal{A}$ of subsets of E is called a **d-system** if

- $E \in \mathcal{A}$;
- $A, B \in \mathcal{A}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{A}$;
- $A_n \in \mathcal{A}$ is an increasing sequence of sets then $\bigcup_n A_n \in \mathcal{A}$.

Remark 7. Equivalently, $\mathcal{A}$ is a d-system if

- $\varnothing \in \mathcal{A}$;
- $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- $A_n \in \mathcal{A}$ is a sequence of disjoint sets then $\bigcup_n A_n \in \mathcal{A}$.

The difference between this and a σ-algebra is the requirement for disjoint sets.

Proof on Sheet 1.

Proposition 1.2

A d-system which is also a π-system is a σ-algebra.

Proof. Sheet 1. □

Lemma 1.1 (Dynkin's Lemma/π-λ/π-d theorem)

Let $\mathcal{A}$ be a π-system. Then any d-system that contains $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.

Proof. We define

$$\mathcal{D} = \bigcap_{\mathcal{D}' \text{ is a } d\text{-system; } \mathcal{D}' \supseteq \mathcal{A}} \mathcal{D}'$$

We can show this is a d-system (proof same as in $\sigma(\mathcal{A})$ on Sheet 1). It suffices to prove that $\mathcal{D}$ is a π-system, because then it is a σ-algebra[a].

We now define

$$\mathcal{D}' = \{B \in \mathcal{D} : \forall A \in \mathcal{A}, B \cap A \in \mathcal{D}\}$$

We can see that $\mathcal{A} \subseteq \mathcal{D}'$, as $\mathcal{A}$ is a π-system.

We now show that $\mathcal{D}'$ is a d-system, fix $A \in \mathcal{A}$.

- Clearly $E \cap A = A \in \mathcal{A} \subseteq \mathcal{D}'$ hence $E \in \mathcal{D}'$.
- Let $B_1, B_2 \in \mathcal{D}'$ such that $B_1 \subseteq B_2$. Then $(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A)$, and since $B_i \cap A \in \mathcal{D}$ this difference also lies in $\mathcal{D}$, so $B_2 \setminus B_1 \in \mathcal{D}'$.
- Now, suppose B_n is an increasing sequence converging to B, and $B_n \in \mathcal{D}'$. Then $B_n \cap A \in \mathcal{D}$, and $\mathcal{D}$ is a d-system, we have $B \cap A \in \mathcal{D}$, so $B \in \mathcal{D}'$.

Hence $\mathcal{D}'$ is a d-system. Also, $\mathcal{D}' \subseteq \mathcal{D}$ by construction of $\mathcal{D}'$. But also $\mathcal{A} \subseteq \mathcal{D}'$ and $\mathcal{D}'$ is a d-system so $\mathcal{D} \subset \mathcal{D}'$ as $\mathcal{D}$ is the smallest d-system containing $\mathcal{A}$. Thus $\mathcal{D} = \mathcal{D}'$, i.e $\forall\ B \in \mathcal{D}$ and $A \in \mathcal{A}, B \cap A \in \mathcal{D}$ $(*)$.

We then define

$$\mathcal{D}'' = \{B \in \mathcal{D} : \forall A \in \mathcal{D}, B \cap A \in \mathcal{D}\}$$

Note that $\mathcal{A} \subseteq \mathcal{D}''$ by $(*)$. Running the same argument as before, we can show that $\mathcal{D}''$ is a d-system. So $\mathcal{D}'' = \mathbb{D}$. But then (by the definition of $\mathcal{D}''$), $\forall\ B \in \mathcal{D}, A \in \mathcal{D} \implies B \cap A \in \mathcal{D}$, i.e. $\mathcal{D}$ is a π-system (check that $\varnothing \in \mathcal{D}$).

So $\mathcal{D}$ is a σ-algebra containing $\mathcal{A}$, hence $\mathcal{D} \supseteq \sigma(\mathcal{A})$. □

[a]As $\mathcal{D} \supseteq \mathcal{A}$ and $\sigma(\mathcal{A})$ the intersection of all σ-algebras containing $\mathcal{A}$, $\mathcal{D} \supseteq \sigma(\mathcal{A})$.

Theorem 1.2 (Uniqueness of extension)

Let μ_1, μ_2 be measures on a measurable space $(E, \mathcal{E})$, such that $\mu_1(E) = \mu_2(E) < \infty$. Suppose that μ_1 and μ_2 coincide on a π-system $\mathcal{A}$, such that $\mathcal{E} \subseteq \sigma(\mathcal{A})$. Then $\mu_1 = \mu_2$ on $\sigma(\mathcal{A})$, and hence on $\mathcal{E}$.

Proof. We define

$$\mathcal{D} = \{A \in \mathcal{E} : \mu_1(A) = \mu_2(A)\}$$

This collection contains $\mathcal{A}$ by assumption. By Dynkin's lemma, it suffices to prove $\mathcal{D}$ is a d-system, because then $\mathcal{D} \supseteq \sigma(\mathcal{A}) \supseteq \mathcal{E}$ giving $\mathcal{D} = \mathcal{E}$ as $\mathcal{D} \subseteq \mathcal{E}$.

- $\varnothing \in \mathcal{D}$, since $\mu_1(\varnothing) = \mu_2(\varnothing) = 0$;
- $A \in \mathcal{D} \implies \mu_1(A) = \mu_2(A)$, thus $\mu_1(A^c) = \mu_1(E) - \mu_1(A) = \mu_2(E) - \mu_2(A) = \mu_2(A^c)$, so $A^c \in \mathcal{D}$ (μ_1, μ_2 finite so this works);
- Let $A_n \in \mathcal{D}$ be a disjoint sequence then, $\mu_1(\bigcup_n A_n) = \sum \mu_1(A_n) = \sum \mu_2(A_n) = \mu_2(\bigcup_n A_n)$ by countable additivity. So $\bigcup_n A_n \in \mathcal{D}$.

So $\mathcal{D}$ is a d-system. □

Remark 8. If $A_n \in \mathcal{A}$ an increasing sequence, then $\mu(\mathcal{A}) = \lim_{n\to\infty} \mu(A_n)$. Use this to show that $\mathcal{D}$ is a d-system satisfying conditions in d-system.

The above theorem applies to finite measures (μ such that $\mu(E) < \infty$) only. However, the theorem can be extended to measures that are σ-finite, for which $E = \bigcup_{n\in\mathbb{N}} E_n$ where $\mu(E_n) < \infty$.

Question

How to show all sets of a σ-algebra $\mathcal{E}$ generated by $\mathcal{A}$ has a certain property $\mathcal{P}$?

Answer

Consider set $\mathcal{G} = \{A \subseteq E : A \text{ has the property } \mathcal{P}\}$ and have that all elements of $\mathcal{A}$ have the property $\mathcal{A}$.

Method 1: Show that $\mathcal{G}$ is a σ-algebra, as it then must contain $\sigma(\mathcal{A}) = \mathcal{E}$.

Method 2: Show that $\mathcal{G}$ is a d-system and pick $\mathcal{A}$ s.t. it is a π-system and use Dynkin's Lemma/π-λ/π-d theorem.

Method 3: Monotone Convergence Theorem, we will see it shortly.

§1.4 Borel measures

Definition 1.11 (Borel Sets)

Let (E, τ) be a Hausdorff topological space. The σ-algebra generated by the open sets of E, i.e. $\sigma(\mathcal{A})$ where $\mathcal{A} = \{A \subseteq E : A \text{ open}\}$, is called the **Borel σ-algebra** on E, denoted $\mathcal{B}(E)$.
A measure μ on $(E, \mathcal{B}(E))$ is called a **Borel measure on E**.

Members of $\mathcal{B}(E)$ are called **Borel sets**.

Notation. We write $\mathcal{B} = \mathcal{B}(\mathbb{R})$.

Definition 1.12 (Radon Measure)

A **Radon measure** is a Borel measure μ on E such that $\mu(K) < \infty$ for all $K \subseteq E$ compact.

Note that in a Hausdorff space, compact sets are closed and hence measurable.

Definition 1.13 (Probability Measure)

If $\mu(E) = 1$, μ is called a **probability measure** on E, and $(E, \mathcal{E}, \mu)$ is called a probability space, typically denoted instead by $(\Omega, \mathcal{F}, \mathcal{P})$.

Definition 1.14 (Finite Measure)
If $\mu(E) < \infty$, μ is a **finite measure** on E.

Definition 1.15 (σ-finite Measure)
If $\exists$ sequence $E_n \in \mathcal{E}$ s.t. $\mu(E_n) < \infty \ \forall\, n$ and $E = \bigcup_n E_n$, then μ is called a **σ-finite measure**.

Remark 9. Arguments that hold for finite measures can usually be extended to σ-finite measures.

§1.5 Lebesgue measure

One of the main goals for this course is to define a notion of volume for arbitrary sets, we can do this by constructing a Borel measure μ on $\mathcal{B}(\mathbb{R}^d)$ s.t $\mu\left(\prod_{i=1}^d (a_i, b_i)\right) = \prod_{i=1}^d (b_i - a_i)$ where $a_i < b_i$ corresponding to the usual notion of volume of rectangles.

Initially, we will perform this construction for $d = 1$, and later we will consider product measures to extend this to higher dimensions.

Theorem 1.3 (Construction of the Lebesgue measure)
There exists a unique Borel measure μ on $\mathbb{R}$ such that

$$a < b \implies \mu\left((a, b]\right) = b - a. \tag{$\dagger$}$$

μ is called the Lebesgue measure on $\mathbb{R}$.

Proof. First we shall prove the existence of the measure and then uniqueness.

Consider the ring $\mathcal{A}$ of finite unions of disjoint intervals[a] of the form

$$\mathcal{A} = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$$

where $a_1 \le b_1 \le a_2 \le \cdots \le a_n \le b_n$. Note that $\sigma(\mathcal{A}) = \mathcal{B}$ (see Example Sheets[b]).

Define for each $A \in \mathcal{A}$

$$\mu(A) = \sum_{i=1}^n (b_i - a_i).$$

This agrees with ($\dagger$) for $(a, b]$. This is additive and well-defined (check).

So, the existence of μ on $\sigma(\mathcal{A}) = \mathcal{B}$ follows from Carathéodory's theorem if we can show that μ is *countable additive* on $\mathcal{A}$.

Remark 10. Suppose μ a finitely additive set function on a ring $\mathcal{A}$. Then μ is countable additive iff

- $A_n \uparrow$ [c]$A; A_n, A \in \mathcal{A} \implies \mu(A_n) \uparrow \mu(A)$.
- In addition, if μ is finite and $A_n \downarrow A$ s.t. $A_n, A \in \mathcal{A}$ then $\mu(A_n) \downarrow \mu(A)$[d].

See Example Sheet for proof.

So showing μ is countably additive on $\mathcal{A}$ is equivalent to showing the following
If $A_n \in \mathcal{A}, A_n \downarrow \varnothing$ then $\mu(A_n) \downarrow 0$. We require that μ is finite, as A_n decreasing we require A_1 to have finite measure. ?????

We shall prove this by contradiction.

Suppose this is not the case, so there exist $\varepsilon > 0$ and $B_n \in \mathcal{A}$ such that $B_n \downarrow \varnothing$ but $\mu(B_n) \geq 2\varepsilon$ for infinitely many n (and so wlog for all n).
We can approximate B_n from within by a sequence $\overline{C}_n$[e] $\in \mathcal{A}$ s.t. $C_n \subseteq B_n$ and $\mu(B_n \setminus C_n) \leq \varepsilon/2^n$. Suppose $B_n = \bigcup_{i=1}^{N_n}(a_{ni}, b_{ni}]$, then define $C_n = \bigcup_{i=1}^{N_n}(a_{ni} + \frac{2^{-n}\varepsilon}{N_n}, b_{ni}]$. Note that the C_n lie in $\mathcal{A}$, and $\mu(B_n \setminus C_n) \leq 2^{-n}\varepsilon$. Since B_n is decreasing, we have $B_N = \bigcap_{n\leq N} B_n$, and

$$B_N \setminus (C_1 \cap \cdots \cap C_N) = B_n \cap \left(\bigcup_{n\leq N} C_n^c\right) = \bigcup_{n\leq N} B_N \setminus C_n \subseteq \bigcup_{n\leq N} B_n \setminus C_n$$

Since μ is increasing and finitely additive and thus subadditive on $\mathcal{A}$,

$$\mu(B_N \setminus (C_1 \cap \cdots \cap C_N)) \leq \mu\left(\bigcup_{n\leq N} B_n \setminus C_n\right) \leq \sum_{n\leq N} \mu(B_n \setminus C_n) \leq \sum_{n\leq N} 2^{-N}\varepsilon \leq \varepsilon$$

Since $\mu(B_N) \geq 2\varepsilon$, additivity implies that $\mu(C_1 \cap \cdots \cap C_N) \geq \varepsilon$. This means that $C_1 \cap \cdots \cap C_N$ cannot be empty. We can add the left endpoints of the intervals, giving $K_N = \overline{C}_1 \cap \cdots \cap \overline{C}_N \neq \varnothing$. By Analysis I, K_N is a nested sequence of bounded nonempty closed intervals and therefore there is a point $x \in \mathbb{R}$ such that $x \in K_N$ for all N[f]. But $K_N \subseteq \overline{C}_N \subseteq B_N$, so $x \in \bigcap_N B_n$, which is a contradiction since $\bigcap_N B_N$ is empty. Therefore, a measure μ on $\mathcal{B}$ exists.

Now we prove uniqueness. Suppose μ, λ are measures such that the measure of an interval $(a, b]$ is $b - a$. We define truncated measures for $A \in \mathcal{B}$

$$\mu_n(A) = \mu\left(A \cap (n, n+1)\right)$$
$$\lambda_n(A) = \lambda\left(A \cap (n, n+1]\right)$$

Then μ_n, λ_n are *probability measures* on $\mathcal{B}$ and $\mu_n = \lambda_n$ on the π-system of intervals of the form $(a, b]$ with $a < b$[g]. This π-system generates $\mathcal{B}$, so by the uniqueness theorem for finite measures (theorem 1.2) $\mu_n = \lambda_n$ on $\mathcal{B}$. Hence $\forall\, A \in \mathcal{B}$

$$\begin{aligned}\mu(A) &= \mu\left(\bigcup_n A \cap (n, n+1]\right) \\ &= \sum_{n\in\mathbb{Z}} \mu(A \cap (n, n+1]) \\ &= \sum_{n\in\mathbb{Z}} \mu_n(A) \\ &= \sum_{n\in\mathbb{Z}} \lambda_n(A) = \cdots = \lambda(A)\end{aligned}$$

$\square$

[a] We take semi intervals as for $\mathcal{A}$ to be a ring, we require the set difference to be in $\mathcal{A}$.
[b] as all open intervals are in $\sigma(\mathcal{A})$ and open intervals generate open sets
[c] increasing sequence tending to A
[d] E.g. let $A_n = [n, \infty)$ with the Lebesgue measure then $A_n \downarrow \varnothing$. But $\mu(A_n) = \infty$ whilst $\mu(\varnothing) = 0$
[e] $\overline{C}_n$ means the closure of C_n, i.e. make it a closed set by including the left endpoint
[f] As completeness of $\mathbb{R}$ implies $\bigcap_n K_n$ is closed and non empty.
[g] As $(a, b] \cap (c, d] = \varnothing$ or $(e, f]$.

Definition 1.16 (Lebesgue null set)

A Borel set $B \in \mathcal{B}$ is called a **Lebesgue null set** if $\lambda(B) = 0$ where λ is the Lebesgue measure.

Remark 11. A singleton $\{x\}$ can be written as $\bigcap_n \left(x - \frac{1}{n}, x\right]$, hence $\lambda(x) = \lim_n \frac{1}{n} = 0$. Hence singletons are null sets. In particular, $\lambda((a, b)) = \lambda((a, b]) = \lambda([a, b)) = \lambda([a, b])$. Any countable set $Q = \bigcup_q \{q\}$ is a null set. Not all null sets are countable; the Cantor set is an example.

The Lebesgue measure is *translation-invariant*. Let $x \in \mathbb{R}$, then the set $B + x = \{b + x : b \in B\}$ lies in $\mathcal{B}$ iff $B \in \mathcal{B}$, and in this case, it satisfies $\lambda(B + x) = \lambda(B)$. We can define the translated Lebesgue measure $\lambda_x(B) = \lambda(B + x)$ for all $B \in \mathcal{B}$, then $\lambda_x((a, b]) = \lambda((a, b] + x) = \lambda((a + x, b + x]) = b - a = \lambda((a, b])$. So $\lambda_x = \lambda$ on the π-system of intervals and so $\lambda_x = \lambda$ on the sigma algebra $\mathcal{B}$ (i.e. $\forall\, B \in \mathcal{B}, \lambda(B + x) = \lambda(B)$).

Question

Is the Lebesgue measure the only such translation invariant measure on $\mathcal{B}$?

Carathéodory's theorem extends λ from $\mathcal{A}$ to not just $\sigma(\mathcal{A}) = \mathcal{B}$, but actually to $\mathcal{M}$, the set of outer-measurable sets $M \supseteq \mathcal{B}$, but how large is $\mathcal{M}$?

The class of outer measurable sets $\mathcal{M}$ used in Carathéodory's extension theorem is here called the class of Lebesgue measurable sets. This class, the Lebesgue σ-algebra, can be shown to be

$$\mathcal{M} = \{A \cup N : A \in \mathcal{B}, N \subseteq B, B \in \mathcal{B}, \lambda(B) = 0\} \supsetneq \mathcal{B}$$

§1.6 Existence of non-measurable sets

We now show that $\mathcal{B} \subsetneq \mathcal{P}(\mathbb{R})$ (in fact $\mathcal{M}_{leb} \subsetneq \mathcal{P}(\mathbb{R})$).

Consider $E = [0, 1)$ with addition defined modulo one. By the same argument as before, the Lebesgue measure is translation-invariant modulo one. Consider the subgroup $Q = E \cap \mathbb{Q}$ of $(E, +)$. We define $x \sim y$ for $x, y \in E$ if $x - y \in Q$. Assuming the axiom of choice (uncountable version), we can select a representative from each equivalence class, and denote by S the set of such representatives. We shall show that $S \notin \mathcal{B}$.

We can partition E into the union of its cosets, so $E = \bigcup_{q \in Q}(S+q)$ is a disjoint[1] union.

Suppose S is a Borel set. Then $S + q$ is also a Borel set[2]. Therefore by translation invariance of λ and by countably additivity,

$$\lambda([0, 1)) = 1 = \lambda\left(\bigcup_{q \in Q}(S + q)\right) = \sum_{q \in Q} \lambda(S + q) = \sum_{q \in Q} \lambda(S)$$

But no value for $\lambda(S) \in [0, \infty]$ can be assigned to make this equation hold. Therefore S is not a Borel set.

Remark 12. We can extend this proof to show that $S \notin \mathcal{M}_{leb}$.

One can further show that λ cannot be extended to all subsets $\mathcal{P}(E)$.

Theorem 1.4 (Banach - Kuratowski)

Assuming the continuum hypothesis, there exists no measure μ on the set $\mathcal{P}([0, 1))$ such that $\mu([0, 1)) = 1$ and $\mu(\{x\}) = 0$ for $x \in [0, 1)$.

Henceforth, whenever we are on a metric space E, we will work with $\mathcal{B}(E)$, which will be perfectly satisfactory.

§1.7 Probability spaces

Definition 1.17

If a measure space $(E, \mathcal{E}, \mu)$ has $\mu(E) = 1$, we call it a **probability space**, and

[1] Suppose $s_1 + q_1 = s_2 + q_2$ then $s_1 - s_2 = q_1 - q_2 \in \mathbb{Q}$ but then $s_1, s_2 \in S$ by definition $\lightning$.

[2] Consider $\mathcal{G} = \{B \in \mathcal{B} : B + x \in \mathcal{B}\}$ we can show this is a σ-algebra, see page 11.

instead write $(\Omega, \mathcal{F}, \mathbb{P})$. We call Ω the outcome space or sample space, $\mathcal{F}$ the set of events, and $\mathbb{P}$ the probability measure.

The axioms of probability theory (Kolmogorov, 1933), are

1. $\mathbb{P}(\Omega) = 1, \mathbb{P}(\varnothing) = 0$;
2. $0 \leq \mathbb{P}(E) \leq 1$ for all $E \in \mathcal{F}$;
3. if A_n are a disjoint sequence of events in $\mathcal{F}$, then $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$.

This is exactly what is required by our definition: $\mathbb{P}$ is a measure on a σ-algebra.

Remark 13.

- $\mathbb{P}(\bigcup_n A_n) \leq \sum_n \mathbb{P}(A_n)$ for all sequences $A_n \in \mathcal{F}$;
- $A_n \uparrow A \implies \mathbb{P}(A_n) \uparrow \mathbb{P}(A)$;
- $A_n \downarrow A \implies \mathbb{P}(A_n) \downarrow \mathbb{P}(A)$ as $\mathbb{P}$ a finite measure.

This definition is what separates probability from analysis.

Definition 1.18 (Independent)
Events $(A_i, i \in I), A_i \in \mathcal{F}$ are **independent** if for all finite $J \subseteq I$, we have

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

σ-algebras $(\mathcal{A}_i, i \in I), \mathcal{A}_i \subseteq \mathcal{F}$ are **independent** if for any $A_j \in \mathcal{A}_j$, where $J \subseteq I$ is finite, the A_j are independent.

Kolmogorov showed that these definitions are sufficient to derive the law of large numbers.

Proposition 1.3
Let $\mathcal{A}_1, \mathcal{A}_2$ be π-systems of sets in $\mathcal{F}$. Suppose $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$ for all $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$. Then the σ-algebras $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2)$ are independent.

Proof. Fix $A_1 \in \mathcal{A}_1$, and define for all $A \in \sigma(\mathcal{A}_2)$.

$$\mu(A) = \mathbb{P}(A_1 \cap A), \nu(A) = \mathbb{P}(A_1)(A).$$

Then μ, ν are finite measures and they agree on the π-system $\mathcal{A}_2$. Hence by Uniqueness of extension, $\mu(A) = \nu(A)\ \forall\ A \in \sigma(\mathcal{A}_2)$, i.e. $\mathbb{P}(A_1 \cap A) = \mathbb{P}(A_1)\mathbb{P}(A)\ \forall\ A_1 \in$

$\mathcal{A}_1, A_2 \in \sigma(\mathcal{A}_2)$.

Now repeat same argument, but now by fixing $A_2 \in \sigma(\mathcal{A}_2)$ define for all $A \in \sigma(\mathcal{A}_1)$

$$\mu'(A) = \mathbb{P}(A \cap A_2), \nu'(A) = \mathbb{P}(A)(A_2).$$

Then μ', ν' are finite measures and they agree on the π-system $\mathcal{A}_1$. Hence by Uniqueness of extension, $\mu'(A) = \nu'(A)\ \forall\ A \in \sigma(\mathcal{A}_1)$, i.e. $\mathbb{P}(A_1 \cap A) = \mathbb{P}(A_1)\mathbb{P}(A)\ \forall\ A_1 \in \sigma(\mathcal{A}_1), A_2 \in \sigma(\mathcal{A}_2)$.

□

This follows by uniqueness.

§1.8 Borel–Cantelli lemmas

Definition 1.19

Let $A_n \in \mathcal{F}$ be a sequence of events. Then the **limit superior** of A_n is

$$\limsup_n A_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_n \text{ infinitely often}\}^a$$

The **limit inferior** of A_n is

$$\liminf_n A_n = \bigcup_n \bigcap_{m \geq n} A_m = \{A_n \text{ eventually}\}^b$$

[a] Consider ω, if $\omega \in \limsup_n A_n$ then $\forall\ n, \omega \in \bigcup_{m \geq n} A_m$ thus ω must be in an infinite number of A_ns.

[b] ω is in all but finitely many A_n.

Lemma 1.2 (First Borel–Cantelli lemma)

Let $A_n \in \mathcal{F}$ be a sequence of events such that $\sum_n \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}(A_n \text{ infinitely often}) = 0$.

Proof. For all n, we have

$$\mathbb{P}\left(\limsup_n A_n\right) = \mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} A_m\right) \leq \mathbb{P}\left(\bigcup_{m \geq n} A_m\right) \leq^a \sum_{m \geq n} \mathbb{P}(A_m) \to 0$$

□

[a] By countable subadditivity

This proof did not require that $\mathbb{P}$ be a probability measure, just that it is a measure.

Therefore, we can use this for arbitrary measures.

Lemma 1.3 (Second Borel–Cantelli lemma)
Let $A_n \in \mathcal{F}$ be a sequence of independent events with $\sum_n \mathbb{P}(A_n) = \infty$. Then $\mathbb{P}(A_n \text{ infinitely often}) = 1$.

Proof. By independence, for all $N \geq n \in \mathbb{N}$ and using $1 - a \leq e^{-a}$, we find

$$\mathbb{P}\left(\bigcap_{m=n}^{N} A_m^c\right) = \prod_{m=n}^{N} (1 - \mathbb{P}(A_m)) \leq \prod_{m=n}^{N} e^{-\mathbb{P}(A_m)} = e^{-\sum_{m=n}^{N} \mathbb{P}(A_m)}$$

As $N \to \infty$, this approaches zero.
Since $\bigcap_{m=n}^{N} A_m^c$ decreases to $\bigcap_{m=n}^{\infty} A_m^c$, $\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0$ as $\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) \leq \mathbb{P}\left(\bigcap_{m=n}^{N} A_m^c\right) \leq e^{-\sum_{m=n}^{N} \mathbb{P}(A_m)} \to 0$. So by taking complements $\mathbb{P}(\bigcup_{m=n}^{\infty} A_n) = 1 \; \forall \, n (\dagger)$.

Let $B_n = \bigcup_{m=n}^{\infty} A_m$, B_n decreasing and so $B_n \downarrow \bigcap_n B_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_n \text{ i.o}\}$[a]. As $\mathbb{P}(B_n) = 1$ by $(\dagger)$, $\mathbb{P}\{A_n \text{ i.o}\} = \lim_{n\to\infty} \mathbb{P}(B_n) = 1$ as probabilities are a finite measure. $\square$

[a] A_n occurs infinitely often

Remark 14. If A_n independent, then $\{A_n \text{ i.o}\}$ has either probability 0 or 1 and is called a "tail event". Kolmogorov 0-1 law shows this is true for all "tail events".

§2 Measurable Functions

§2.1 Definition

Definition 2.1 (Measurable)

Let $(E, \mathcal{E}), (G, \mathcal{G})$ be measurable spaces. A function $f \colon E \to G$ is called **measurable** if $f^{-1}(A) \in \mathcal{E} \ \forall\ A \in \mathcal{G}$, where $f^{-1}(A)$ is the preimage of A under f i.e. $f^{-1}(A) = \{x \in E : f(x) \in A\}$.

If $G = \mathbb{R}$ and $\mathcal{G} = \mathcal{B}$, we can just say that $f \colon (E, \mathcal{E}) \to G$ is measurable. Moreover, if E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, we say f is Borel measurable.

Note that preimages f^{-1} commute with many set operations such as intersection, union, and complement. This implies that $\{f^{-1}(A) : A \in \mathcal{G}\}$ is a σ-algebra over E, and likewise, $\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra over G. Hence, if $\mathcal{A}$ is a collection of subsets s.t. $G \supset \sigma(\mathcal{A})$ then if $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A}$, the class $\{A : f^{-1} \in \mathcal{E}\}$ is a σ-algebra that contains $\mathcal{A}$ and so $\sigma(\mathcal{A})$. So f is measurable.

If $f \colon (E, \mathcal{E}) \to \mathbb{R}$, the collection $\mathcal{A} = \{(-\infty, y] \colon y \in \mathbb{R}\}$ generates $\mathcal{B}$ (Sheet 1). Hence f is Borel measurable iff $f^{-1}((-\infty, y]) = \{x \in E : f(x) \le y\} \in \mathcal{E}$ for all $y \in \mathbb{R}$.

If E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, then if $f \colon E \to \mathbb{R}$ is continuous, the preimages of open sets B are open, and hence Borel sets. The open sets in $\mathbb{R}$ generate the σ-algebra $\mathcal{B}$. Hence, continuous functions to the real line are measurable.

Example 2.1

Consider the indicator function 1_A of a set $A \subset E$. $1_A^{-1}(1) = A$ and $1_A^{-1}(0) = A^c$ hence measurable iff $A \in \mathcal{E}$.

Example 2.2

The composition of measurable functions is measurable. Note that given a collection of maps $\{f_i \colon E \to (G, \mathcal{G}) : i \in I\}$, we can make them all measurable by taking $\mathcal{E}$ to be a large enough σ-algebra, for instance $\sigma\left(\left\{f_i^{-1}(A) : A \in \mathcal{G}, i \in I\right\}\right)$ called the σ-algebra generated by $\{f_i\}_{i \in I}$.

Proposition 2.1

If $f_1, f_2, \ldots$ are measurable $\mathbb{R}$-valued. Then $f_1 + f_2$, $f_1 f_2$, $\inf_n f_n$, $\sup_n f_n$, $\liminf f_n$, $\limsup f_n$ are all measurable.

Proof. See Sheet 1. □

§2.2 Monotone class theorem

Theorem 2.1 (Monotone class theorem)

Let $(E, \mathcal{E})$ be a measurable space and $\mathcal{A}$ be a π-system that generates the σ-algebra $\mathcal{E}$. Let $\mathcal{V}$ be a vector space of bounded maps from E to $\mathbb{R}$ s.t.

1. $1_E \in \mathcal{V}$;
2. $1_A \in \mathcal{V}$ for all $A \in \mathcal{A}$;
3. if f is bounded and $f_n \in \mathcal{V}$ are nonnegative functions that form an increasing sequence that converge pointwise to f on E, then $f \in \mathcal{V}$.

Then $\mathcal{V}$ contains all bounded measurable functions $f \colon E \to \mathbb{R}$.

Proof. Define $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{V}\}$. Then $\mathcal{D}$ is a d-system as $1_E \in \mathcal{V}$ and for $A \subseteq B$, $1_{B \setminus A} = 1_B - 1_A \in \mathcal{V}$ as $\mathcal{V}$ a vector space so $B \setminus A \in \mathcal{D}$.
If $A_n \in \mathcal{D}$ increases to A, we have 1_{A_n} increases pointwise to 1_A, which lies in $\mathcal{V}$ by the (3.) so $A \in \mathcal{D}$.

$\mathcal{D}$ contains $\mathcal{A}$ by (2.), as well as E itself. So by Dynkin's lemma $\mathcal{D}$ contains $\sigma(\mathcal{A}) = \mathcal{E}$ so $\mathcal{E} = \mathcal{D}$ i.e. $1_A \in V \ \forall\ A \in \mathcal{E}$.

Since V a vector space it contains all finite linear combinations of indicators of measurable sets. Let $f \colon E \to \mathbb{R}$ be a bounded measurable function, which we will assume at first is nonnegative. We define

$$\begin{aligned} f_n(x) &= 2^{-n} \lfloor 2^n f(x) \rfloor \\ &= 2^{-n} \sum_{j=0}^{\infty} 1_{A_{n,j}}(x) \\ A_{n,j} &= \{2^n f(x) \in [j, j+1)\} \\ &= f^{-1}\left(\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right)\right) \in \mathcal{E}. \end{aligned}$$

As f is bounded we do not need an infinite sum but only a finite one. Then $f_n \le f \le f_n + 2^{-n}$. Hence $|f_n - f| \le 2^{-n} \to 0$ and $f_n \uparrow f$.

So $0 \le f_n \uparrow f, f_n \in \mathcal{V}$ and f is bounded non-negative so $f \in \mathcal{V}$ by (3.).

Finally, for any f bounded and measurable, $f = f^{+}$[a] $- f^{-}$[b]. f^+, f^- are bounded, nonnegative and measurable, so in $\mathcal{V}$ and $\mathcal{V}$ a vector space thus $f \in \mathcal{V}$. □

[a] $\max(f, 0)$
[b] $\max(-f, 0)$

§2.3 Image measures

Definition 2.2 (Image Measure)

Let $f\colon (E,\mathcal{E}) \to (G,\mathcal{G})$ be a measurable function and μ a measure on $(E,\mathcal{E})$. Then the **image measure** $\nu = \mu \circ f^{-1}$ is obtained from assigning $\nu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{G}$.

Remark 15. This is well defined as $f^{-1}(A) \in \mathcal{E}$ as f measurable. ν is countably additive because the preimage satisfies set operations and μ countably additive (See Sheet 1).

Starting from the Lebesgue measure, we can get all probability measures (in fact we can get all Radon measures) in this way.

Definition 2.3 (Right-Continuous)

A function f is **right-continuous** if $x_n \downarrow x \implies f(x_n) \to f(x)$.

Lemma 2.1

Let $g\colon \mathbb{R} \to \mathbb{R}$ be a non-constant, increasing, right-continuous function, and set $g(\pm\infty) = \lim_{z\to\pm\infty} g(z)$. On $I = (g(-\infty), g(+\infty))$ we define the **generalised inverse** $f : I \to \mathbb{R}$ by
$$f(x) = \inf\{y \in \mathbb{R} : g(y) \geq x\}.$$
Then f is increasing, left-continuous, and $f(x) \leq y$ iff $x \leq g(y)$ for all $x \in I, y \in \mathbb{R}$.

Remark 16. f and g form a Galois connection.

Proof. Fix $x \in I$.
Let $J_x = \{y \in \mathbb{R} : g(y) \geq x\}$. Since $x > g(-\infty)$, J_x is nonempty and bounded below. Hence $f(x)$ is a well-defined real number.
If $y \in J_x$, then $y' \geq y$ implies $y' \in J_x$ since g is increasing. Since g is right-continuous, if $y_n \downarrow y$, and all $y_n \in J_x$, then $g(y) = \lim_n g(y_n) \geq x$ so $y \in J_x$.
So $J_x = [f(x), \infty)$. Hence $f(x) \leq y \iff x \leq g(y)$ as required.

If $x \leq x'$, we have $J_x \supseteq J_{x'}$ (as $y \in J_x \impliedby y \in J'_x$), i.e. $[f(x), \infty) \supseteq [f(x'), \infty)$ so $f(x) \leq f(x')$.
Similarly, if $x_n \uparrow x$, we have $J_x = \bigcap_n J_{x_n}$[a] so $[f(x), \infty) = \bigcap_n [f(x_n), \infty)$ so $f(x_n) \to f(x)$ as $x_n \to x$. □

[a] As $y \in \bigcap_n J_{x_n} \iff g(y) \geq x_n \ \forall\, n \iff g(y) \geq x \iff y \in J_x$.

Theorem 2.2

Let $g\colon \mathbb{R} \to \mathbb{R}$ as in the previous lemma. Then $\exists$ a unique Radon measure μ_g on $\mathbb{R}$ such that $\mu_g((a,b]) = g(b) - g(a)$ for all $a < b$. Further, all Radon measures on $\mathbb{R}$ can be obtained in this way.

Proof. Define I, f as in the previous lemma and λ the Lebesgue measure on I.

f is Borel measurable since $f^{-1}((-\infty, z]) = \{x \in I\colon f(x) \le z\} = \{x \in I\colon x \le g(z)\} = (-g(\infty), g(z)] \in \mathcal{B}$. As $\{(-\infty, z] : z \in \mathbb{R}\}$ generate $\mathcal{B}$, f measurable.

Therefore, the image measure $\mu_g = \lambda \circ f^{-1}$ exists on $\mathcal{B}$. Then for any $-\infty < a < b < \infty$, we have

$$\begin{aligned}\mu_g((a,b]) &= \lambda\left(f^{-1}((a,b])\right)\\ &= \lambda(\{x\colon a < f(x) \le f(b)\})\\ &= \lambda(\{x\colon g(a) < x \le g(b)\})\\ &= g(b) - g(a)\end{aligned}$$

By the Uniqueness of extension for σ-finite measures, μ_g is uniquely defined.

Conversely, let ν be a Radon measure on $\mathbb{R}$. Define $g : \mathbb{R} \to \mathbb{R}$ as

$$g(y) = \begin{cases} \nu((0,y]) & \text{if } y \ge 0 \\ -\nu((y,0]) & \text{if } y < 0 \end{cases}$$

ν Radon tells us that g is finite. Easy to check g is right-continuous[a]. This is an increasing function in y, since ν is a measure. Finally, $\nu((a,b]) = g(b) - g(a)$ which can be seen by case analysis and additivity of the measure ν. By uniqueness as before, this characterises ν in its entirety. □

[a] For $y_n \downarrow y$ where $y \ge 0$, $(0, y_n] \downarrow (0, y]$ and then $\nu((0, y_n]) \downarrow \nu((0, y])$ by countably additivity. Similarly for $y < 0$.

Remark 17. Such image measures μ_g are called **Lebesgue–Stieltjes measures** associated with g, where g is the **Stieltjes distribution**.

Example 2.3

Fix $x \in \mathbb{R}$ and take $g = 1_{[x,\infty)}$. Then $\mu_g = \delta_x$ the *dirac measure at* x defined for all $A \in \mathcal{B}$ by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

§2.4 Random variables

Definition 2.4 (Random Variable)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E})$ be a measurable space. If $X : \Omega \to E$ a measurable function then X is a **random variable** in E.

When $E = \mathbb{R}$ or $\mathbb{R}^d$ with the Borel σ-algebra, we simply call X a random variable or random vector.

Example 2.4

X models a "random" outcome of an experiment, e.g. when tossing a coin $\Omega = \{H, T\}, X = \#\text{ heads} : \Omega \to \{0, 1\}$.

Definition 2.5 (Distribution)

The **law** or **distribution** μ_X of a random variable X is given by the image measure $\mu_X = \mathbb{P} \circ X^{-1}$. It is a measure on $(E, \mathcal{E})$.

When $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, μ_X is uniquely determined by its values on any π-system, we shall take $\{(-\infty, x] : x \in \mathbb{R}\}$ and

$$F_X(z) = \mu_X((-\infty, z]) = \mathbb{P}(X^{-1}(-\infty, z]) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \leq z\}\right) = \mathbb{P}\left(X \leq z\right)$$

The function F_x is called the **distribution function** of X, because it uniquely determines the distribution of X.

Using the properties of measures, we can show that any distribution function satisfies:

1. F_X is increasing;
2. F_X is right-continuous[3];
3. $F_X(-\infty) = \lim_{z \to -\infty} F_X(z) = \mu_X(\varnothing) = 0$;
4. $F_X(\infty) = \lim_{z \to \infty} F_X(z) = \mu_X(\mathbb{R}) = \mathbb{P}\left(\Omega\right) = 1$.

Proposition 2.2

Given any function F satisfying the previous properties, $\exists$ a random variable X s.t. $F = F_X$.

Proof. Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}(0, 1)$, $\mathbb{P}$ the Lebesgue measure $\lambda|_{(0,1)}$.
Let F be any function satisfying the properties, then F is increasing and right

[3] $x_n \downarrow x \implies (-\infty, x_n] \downarrow (-\infty, x]$ hence by countable additivity of $\mathbb{P} \circ X^{-1}$.

continuous so we can define the generalised inverse

$$X(\omega) = \inf\{x : \omega \le F(x)\} : (0,1) \to \mathbb{R}$$

Hence X is a measurable function and thus a random variable.

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \le x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \le x\}) = \mathbb{P}(\{\omega \in \Omega : \omega \le F(x)\}) \\ &= \mathbb{P}(\{\omega \in (0,1) : \omega \le F(x)\}) \\ &= \mathbb{P}((0, F(x)]) \\ &= F(x) - 0 \end{aligned}$$

□

Remark 18. This is similar to what we saw in IB Probability, if we have F then r.v. $F^{-1}(U)$ where $U \sim U(0,1)$ has the distribution function F, where F^{-1} is the generalised inverse.

Definition 2.6 (Independent)

Consider a countable collection $(X_i \colon (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E}))$ for $i \in I$. This collection of random variables is called **independent** if the σ-algebras $\sigma(X_i)$ are independent, recall $\sigma(X_i)$ is generated by $\left\{X_i^{-1}(A) \colon A \in \mathcal{E}\right\}$, the smallest σ-algebra s.t. X_i measurable.

For $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ we show on an Sheet 1 that this is equivalent to the condition

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \dots \mathbb{P}(X_n \le x_n)$$

for all finite subsets $\{X_1, \dots, X_n\}$ of the X_i.

§2.5 Constructing independent random variables

Question

Given a distribution function F, we know $\exists$ a r.v. X corresponding to it. But given an infinite sequence of distribution functions $F_1, F_2, \dots$ does $\exists$ independent r.v. $(X_1, X_2, \dots)$ corresponding to them?

Let $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0,1), \lambda|_{(0,1)})$. We start with Bernoulli random variables.

Any $\omega \in (0,1)$ has a binary representation given by $(\omega_i) \in \{0,1\}^{\mathbb{N}}$ where $\omega = \sum_{i=1}^{\infty} 2^{-i}\omega_i$, which is unique if we exclude infinitely long tails of zeroes from the binary representation (same reasoning as $1.00000\ldots = 0.99999\ldots$).

Definition 2.7 (nth Rademacher function)

The **nth Rademacher function** $R_n : \Omega \to 0, 1$ is given by $R_n(\omega) = \omega_n$, it extracts the nth bit from the binary expansion.

Observe that $R_1 = 1_{(1/2,1]}$, $R_2 = 1_{(1/4,1/2]} + 1_{(3/4,1]}$ and so on. Since each R_n can be given as the sum of finite (2^{n-1}) indicator functions on measurable sets, they are measurable functions and are hence random variables.

Claim 2.1

R_i are iid $\mathrm{Ber}(\frac{1}{2})$.

Proof. $\mathbb{P}(R_n = 1) = \frac{1}{2} = \mathbb{P}(R_n = 0)$ can be checked by induction.

We now show they are independent. For a finite set $(x_i)_{i=1}^n$, by considering the size of the intervals that ω can lie in,

$$\mathbb{P}(R_1 = x_1, \ldots, R_n = x_n) = 2^{-n} = \mathbb{P}(R_1 = x_1) \ldots \mathbb{P}(R_n = x_n)$$

□

Therefore, the R_n are all independent, so countable sequences of independent random variables indeed exist.
The next step is to construct a sequence of iid r.v.s on $\mathrm{U}(0, 1)$.

Now, take a bijection $m \colon \mathbb{N}^2 \to \mathbb{N}$ and define $Y_{k,n} = R_{m(k,n)}$, the Rademacher functions. We now define $Y_n = \sum_{k=1}^{\infty} 2^{-k} Y_{k,n}$[4].

Claim 2.2

Y_n are iid $\mathrm{U}(0, 1)$, i.e. $\mu_{Y_n} = \lambda|_{(0,1)}$ and Y_n independent.

Lemma 2.2

Any measurable functions of independent random variables are independent.

Proof. They are independent because the Y_i are measurable functions of independent random variables, e.g. Y_1 is a measurable function of $Y_{1,1}, Y_{2,1}, \ldots$; Y_2 of $Y_{1,2}, Y_{2,2}, \ldots$

The π-system of intervals $\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]$ for $i = 0, \ldots, 2^m - 1$ for $m \in \mathbb{N}$ generates $\mathcal{B}(0, 1)$ as $\mathbb{Q}$ dense in $\mathbb{R}$. So by theorem 1.2 the distribution of Y_n is identified on the

[4]This converges for all $\omega \in \Omega$ since $|Y_{k,n}| \leq 1$.

intervals.

$$\begin{aligned}\mathbb{P}\left(Y_n \in \left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]\right) &= \mathbb{P}\left(\frac{i}{2^m} < \sum_{k=1}^{\infty} 2^{-k} Y_{k,n} \le \frac{i+1}{2^n}\right)^{a} \\ &= \mathbb{P}(Y_{1,n} = y_1, \dots, Y_{m,n} = y_m) \text{ where } \frac{i}{2^m} = 0.y_1 y_2 \dots y_m \\ &= \prod_{i=1}^{m} \mathbb{P}(Y_{m,n} = y_m) \text{ by independence.} \\ &= 2^{-m} = \lambda\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]\end{aligned}$$

Hence $\mu_{Y_n} = \lambda|_{(0,1)}$ on the π-system and so on $\mathcal{B}(0,1)$. □

[a] This specifies the first m digits in the binary expansion of Y_n.

As before, set $G_n(x) = F_n^{-1}(x)$ which is the generalised inverse. Then G_n are Borel functions, set $X_n = G_n(Y_n)$ for $n \in \mathbb{N}$, then as before $F_{X_n} = F_n$ and X_n are independent as Y_n are.

§2.6 Convergence of measurable functions

Let $(E, \mathcal{E}, \mu)$ be a measure space. Let $A \in \mathcal{E}$ be defined by some property.

Definition 2.8 (Almost everywhere)

We say that a property defining a set $A \in \mathcal{E}$ holds **μ-almost everywhere** if $\mu(A^c) = 0$.

Definition 2.9 (Almost surely)

If μ is a $\mathbb{P}$- measure, we say a property holds **$\mathbb{P}$-almost surely** or **with probability one**, if $\mathbb{P}(A^c) = 0$, i.e. if $\mathbb{P}(A) = 1$.

Definition 2.10 (Convergence almost everywhere)

If f_n and f are measurable functions on $(E, \mathcal{E}, \mu) \to (\mathbb{R}, \mathcal{B})$, we say **$f_n$ converges to f μ-almost everywhere** if $\mu(\{x \in E : f_n(x) \not\to f(x)\}) = 0$.

For r.v.s, we say $X_n \to X$ **$\mathbb{P}$-almost surely** if $\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$.

Definition 2.11 (Convergence in Measure)

We say f_n **converges to f in μ-measure** if for all $\varepsilon > 0$

$$\mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) \to 0,$$

as $n \to \infty$.

We say $X_n \to X$ in $\mathbb{P}$-probability if $\forall\, \varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0$$

as $n \to \infty$.

Theorem 2.3

Let $f_n \colon (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable functions.

1. If $\mu(E) < \infty$, then $f_n \to 0$ a.e. $\implies$ $f_n \to 0$ in measure;
2. If $f_n \to 0$ in measure, $\exists$ subsequence n_k s.t. $f_{n_k} \to 0$ a.e.

Example 2.5

Let $f_n = 1_{(n,\infty)}$ and the Lebesgue measure, then $f_n \to 0$ a.e. but $\mu(|f_n| > \varepsilon) = \infty\ \forall\, n$.

Proof. Fix $\varepsilon > 0$. Suppose $f_n \to 0$ a.e., then for every n,

$$\mu(E) \geq \mu(|f_n| \leq \varepsilon) \geq \mu\left(\bigcap_{m\geq n} \{|f_m| \leq \varepsilon\}\right)$$

Let $A_n = \bigcap_{m\geq n} \{|f_m| \leq \varepsilon\}$ which is increasing to $\bigcup_n \bigcap_{m\geq n} \{|f_m| \leq \varepsilon\}$. So by the countable additivity of μ,

$$\begin{aligned}
\mu\left(\bigcap_{m\geq n} \{|f_m| \leq \varepsilon\}\right) &\to \mu\left(\bigcup_n \bigcap_{m\geq n} \{|f_m| \leq \varepsilon\}\right) \\
&= \mu(|f_n| \leq \varepsilon \text{ eventually}) \\
&\geq \mu(|f_n| \to 0) \\
&= \mu(E) \text{ as } f_n \to 0 \text{ a.e. and } \mu \text{ finite.}
\end{aligned}$$

Hence,

$$\liminf_{n\to\infty} \mu(|f_n| \leq \varepsilon) = \mu(E) \implies \limsup_{n\to\infty} \mu(|f_n| > \varepsilon) \leq 0 \implies \mu(|f_n| > \varepsilon) \to 0$$

□

Proof. Suppose $f_n \to 0$ in measure, choosing $\varepsilon = \frac{1}{k}$ we have

$$\mu\left(|f_n| > \frac{1}{k}\right) \to 0.$$

So we can choose n_k s.t. $\mu\left(|f_n| > \frac{1}{k}\right) \leq \frac{1}{k^2}$. We can choose n_{k+1} in the same way s.t. $n_{k+1} > n_k$. So we get a subsequence n_k s.t. $\mu\left(|f_{n_k}| > \frac{1}{k}\right) < \frac{1}{k^2}$. Also $\sum_k \frac{1}{k^2} < \infty$, so $\sum_k \mu\left(|f_{n_k}| > \frac{1}{k}\right) < \infty$. So by the first Borel–Cantelli lemma, we have

$$\mu\left(\underbrace{|f_{n_k}| > \frac{1}{k} \text{ infinitely often}}_{f_{n_k} \not\to 0}\right) = 0$$

so $f_{n_k} \to 0$ a.e. □

Remark 19. The first statement is false if $\mu(E)$ is infinite: consider $f_n = 1_{(n,\infty)}$ on $(\mathbb{R}, \mathcal{B}, \mu)$, since $f_n \to 0$ almost everywhere but $\mu(f_n) = \infty$.

The second statement is false if we do not restrict to subsequences: consider independent events A_n such that $\mathbb{P}\left(A_n\right) = \frac{1}{n}$, then $1_{A_n} \to 0$ in probability since $\mathbb{P}\left(1_{A_n} > \varepsilon\right) = \mathbb{P}\left(A_n\right) = \frac{1}{n} \to 0$, but $\sum_n \mathbb{P}\left(A_n\right) = \infty$, and by the second Borel–Cantelli lemma, $\mathbb{P}\left(1_{A_n} > \varepsilon \text{ infinitely often}\right) = 1$, so $1_{A_n} \nrightarrow 0$ almost surely.

Definition 2.12 (Convergence in Distribution)

For X and X_n a sequence of r.v.s, we say $X_n \xrightarrow{d} X$[a] if $F_{X_n}(t) \to F_X(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ which are continuity points of F_X.

[a] X_n converges to X in distribution

Remark 20. This definition does not require X_n to be defined on the same probability space.

Remark 21. If $X_n \to X$ in probability, then $X_n \xrightarrow{d} X$, see Sheet 2 for proof.

Example 2.6

Let $(X_n)_{n\in\mathbb{N}}$ be iid Exp(1), i.e. $\mathbb{P}\left(X_n > x\right) = e^{-x}$ for $x \geq 0$.

Question

Find a deterministic fcn $g : \mathbb{N} \to \mathbb{R}$ s.t. a.s. $\limsup \frac{X_n}{g(n)} = 1$.

Define $A_n = \{X_n \geq \alpha \log n\}$ where $\alpha > 0$, so $\mathbb{P}(A_n) = n^{-\alpha}$, and in particular, $\sum_n \mathbb{P}(A_n) < \infty$ if and only if $\alpha > 1$. By the Borel–Cantelli lemmas, we have for all $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1 \text{ infinitely often}\right) = 1; \quad \mathbb{P}\left(\frac{X_n}{\log n} \geq 1 + \varepsilon \text{ infinitely often}\right) = 0$$

In other words, $\mathbb{P}(\limsup_n \frac{X_n}{\log n} = 1) = 1$.

§2.7 Kolmogorov's zero-one law

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of r.v.s. We can define $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$[5]. Let $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n$ be the **tail σ-algebra**, which contains all events in $\mathcal{F}$ that depend only on the 'limiting behaviour' of (X_n).

Theorem 2.4 (Kolmogorov 0-1 Law)
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent r.v.s. Let $A \in \mathcal{T}$ be an event in the tail σ-algebra. Then $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$.
If $Y\colon (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B})$ is measurable, it is constant almost surely.

Proof. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ to be the σ-algebra generated by the first n elements of (X_n). This is also generated by the π-system of sets $A = (X_1 \leq x_1, \dots, X_n \leq x_n)$ for any $x_i \in \mathbb{R}$. Note that the π-system of sets $B = (X_{n+1} \leq x_{n+1}, \dots, X_{n+k} \leq x_{n+k})$, for arbitrary $k \in \mathbb{N}$ and $x_i \in \mathbb{R}$, generates $\mathcal{T}_n$. By independence of the sequence, we see that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all such sets A, B, and so the σ-algebras $\mathcal{T}_n, \mathcal{F}_n$ generated by these π-systems are independent.

Let $\mathcal{F}_\infty = \sigma(X_1, X_2, \dots)$. Then, $\bigcup_n \mathcal{F}_n$ is a π-system that generates $\mathcal{F}_\infty$. If $A \in \bigcup_n \mathcal{F}_n$, we have $A \in \mathcal{F}_n$ for some n, so there exists $\overline{n}$ such that $B \in \mathcal{T}_{\overline{n}}$ is independent of A. In particular, $B \in \bigcap_n \mathcal{T}_n = \mathcal{T}$. By uniqueness, $\mathcal{F}_\infty$ is independent of $\mathcal{T}$.

Since $\mathcal{T} \subseteq \mathcal{F}_\infty$, if $A \in \mathcal{T}$, A is independent from A. So $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$, so $\mathbb{P}(A)^2 - \mathbb{P}(A) = 0$ as required.

Finally, if $Y\colon (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B})$, the preimages of $\{Y \leq y\}$ lie in $\mathcal{T}$, which give probability one or zero. Let $c = \inf\{y : F_Y(y) = 1\}$, so $Y = c$ almost surely. $\square$

[5] The smallest σ-algebra s.t. $X_{n+1}, \dots$ are measurable.