

Part II — Probability and Measure

Based on lectures by Dr Sarkar

Michaelmas 2023

Contents

0	Holes in classical theory	1
1	Introduction	3
1.1	Definitions	3
1.2	Rings and algebras	4
1.3	Uniqueness of extension	8
1.4	Borel measures	11
1.5	Lebesgue measure	12
1.6	Existence of non-measurable sets	14
1.7	Probability spaces	14
1.8	Borel–Cantelli lemmas	15

§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 $n \rightarrow \infty$, i.e. $f_n(x) \geq f_{n+1}(x) \forall x$.

So, $\lim_{n \rightarrow \infty} f_n(x)$ exists. But f is not Riemann integrable . We want a theory of integration s.t. f is integrable and $\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx = \int_0^1 f(x) dx$.
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. $X_i \sim \text{Ber}\left(\frac{1}{2}\right)$ and $A = \frac{\sum_{i=1}^n X_i}{n}$, the average number of heads. By strong law of large numbers $\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{n} \rightarrow \frac{1}{2}\right) = 1$.
 - 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 $\mathbb{E}X_n \rightarrow \mathbb{E}X$.

§1 Introduction

§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:

- $\emptyset \in \mathcal{E}$;
- $A \in \mathcal{E} \implies A^c = E \setminus A \in \mathcal{E}$;
- 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 $\mathcal{E} = \{\emptyset, E\}$. This is a σ -algebra. Also, $\mathcal{P}(E) = \{A \subseteq E\}$ is a σ -algebra.

Remark 1. Since $\bigcap_n A_n = (\bigcup_n A_n^c)^c$, 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 $\mu : \mathcal{E} \rightarrow [0, \infty]$, such that $\mu(\emptyset) = 0$, 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}, μ) 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 $m : E \rightarrow [0, \infty]$ 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 σ -algebra)

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

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

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 $\sigma(\mathcal{A})$ 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 $\emptyset \in \mathcal{A}$ 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 $\emptyset \in \mathcal{A}$ 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 $\sigma(\mathcal{A})$.

Definition 1.7 (Set Function)

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

- 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 $\bigcup_n A_n \in \mathcal{A}$ 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 $\tilde{A}_n = \bigcup_{j \leq n} A_j$, then $B_{n+1} = \tilde{A}_n \setminus \tilde{A}_{n-1}$. □

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 μ^* on $\sigma(\mathcal{A})$ such that $\mu^*|_{\mathcal{A}} = \mu$.

We will later prove that this extended measure is unique.

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

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

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

Definition 1.8 (μ^* measurable)

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

We define the class

$$\mathcal{M} = \{A \subseteq E \mid A \text{ is } \mu^* \text{ measurable}\}$$

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

Step 1. μ^* 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

$$\mu^*(B) \leq \sum_n \mu^*(B_n) \quad (\dagger)$$

We can assume without loss of generality that $\mu^*(B_n) < \infty$ 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

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

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

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

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

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

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

If μ^* is infinite, there is nothing to prove.

We need to prove the converse, that $\mu(A) \leq \mu^*(A)$. 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 $\mu(A) \leq \mu^*(A)$ as required.

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

We have $B \subseteq (B \cap A) \cup (B \cap A^c) \cup \emptyset \cup \dots$, hence by countable subadditivity (\dagger), $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$.

It now suffices to prove the converse, that $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$.

We can assume $\mu^*(B)$ is finite, and so $\forall \varepsilon > 0 \exists A_n \in \mathcal{A}$ s.t. $B \subseteq \bigcup_n A_n$ and $\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n)$. 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,

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

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

Step 4. \mathcal{M} is an algebra: Clearly \emptyset 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

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

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

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap A_1 \cap A_2) + \underbrace{\mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c)}_{\mu^*(B \cap (A_1 \cap A_2)^c) \text{ as } A_1 \in \mathcal{M}} \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c) \end{aligned}$$

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

Step 5. \mathcal{M} is a σ -algebra and μ^* 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 $\mu^*(A) = \sum_n \mu^*(A_n)$.

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

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

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

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

By (†),

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

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

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

So μ^* 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 $\emptyset \in \mathcal{A}$ 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

- $\emptyset \in \mathcal{A}$;
- $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- $A_n \in \mathcal{A}$ 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

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

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 $\mathcal{D} \subseteq \mathcal{D}'$ 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

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

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 $\emptyset \in \mathcal{D}$).

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

^aAs $\mathcal{D} \supseteq \mathcal{A}$ 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

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

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 $\mathcal{D} \subseteq \mathcal{E}$.

- $\emptyset \in \mathcal{D}$, since $\mu_1(\emptyset) = \mu_2(\emptyset) = 0$;
- $A \in \mathcal{D} \implies \mu_1(A) = \mu_2(A)$, 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 $\mu(\mathcal{A}) = \lim_{n \rightarrow \infty} \mu(A_n)$. 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{P} .

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/ \$\pi\$ - \$\lambda\$ / \$\pi\$ - \$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}, μ) 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

We will construct a unique Borel measure μ on \mathbb{R}^d such that

$$\mu\left(\prod_{i=1}^d [a_i, b_i]\right) = \prod_{i=1}^d |b_i - a_i|$$

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

$$a < b \implies \mu((a, b]) = b - a$$

Proof. Consider the subsets of \mathbb{R} of the form

$$A = (a_1, b_1] \cup \dots \cup (a_n, b_n]$$

where the intervals in question are disjoint. The set \mathcal{A} of such sets forms a ring and a π -system of Borel sets. This generates the same σ -algebra as that generated by finite unions of open intervals, by the first example sheet. Open intervals with rational endpoints generate \mathcal{B} , so $\sigma(\mathcal{A}) \supseteq \mathcal{B}$. We define the set function μ on \mathcal{A} by $\mu(A) = \sum_{i=1}^n (b_i - a_i)$. μ is additive, and well-defined since if $A = \bigcup_j C_j = \bigcup_k D_k$ for distinct disjoint unions, we can write $C_j = \bigcup_k (C_j \cap D_k)$ and $D_k = \bigcup_j (D_k \cap C_j)$, giving

$$\mu(A) = \mu\left(\bigcup_j C_j\right) = \sum_j \mu(C_j) = \sum_j \mu\left(\bigcup_k (C_j \cap D_k)\right) = \sum_j \sum_k \mu(C_j \cap D_k) = \mu\left(\bigcup_k D_k\right)$$

To prove the existence of μ on \mathcal{B} , we apply Carathéodory's extension theorem, and therefore must check that μ is countably additive on \mathcal{A} . Equivalently, by a question on an example sheet, it suffices to show that for all sequences $A_n \in \mathcal{A}$ such that A_n decreases to \emptyset , we have $\mu(A_n) \rightarrow 0$. Suppose this is not the case, so there exist $\varepsilon > 0$ and $B_n \in \mathcal{A}$ such that B_n decreases to \emptyset but $\mu(B_n) \geq 2\varepsilon$ for infinitely many n

(and so without loss of generality for all n). We can approximate B_n from within by a sequence C_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,

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

Since in addition $\mu(B_N) \geq 2\varepsilon$, 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 $K_N = \overline{C}_1 \cap \cdots \cap \overline{C}_N$. By Analysis I, K_N is a nested sequence of nonempty closed intervals and therefore there is a point $x \in \mathbb{R}$ such that $x \in K_N$ for all N . 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 new measures $\mu_n(A) = \mu(A \cap (n, n+1])$ and $\lambda_n(A) = \lambda(A \cap (n, n+1])$. These new measures are finite with total mass 1. Hence, we can use the uniqueness of extension theorem to show $\mu_n = \lambda_n$ on \mathcal{B} . We find

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

□

Definition 1.16

A Borel set $B \in \mathcal{B}$ is called a *Lebesgue null set* if $\mu(B) = 0$.

Remark 10. A singleton $\{x\}$ can be written as $\bigcap_n \left(x - \frac{1}{n}, x\right]$, hence $\mu(\{x\}) = \lim_n \frac{1}{n} = 0$. Hence singletons are null sets. In particular, $\mu((a, b)) = \mu((a, b]) = \mu([a, b)) = \mu([a, b])$. 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 $B + x = \{b + x \mid b \in B\}$ lies in \mathcal{B} if and only if $B \in \mathcal{B}$, and in this case, it satisfies $\mu(B+x) = \mu(B)$. We can define the translated Lebesgue measure $\mu_x(B) = \mu(B+x)$ for all $B \in \mathcal{B}$, but since the Lebesgue measure is unique, $\mu_x = \mu$.

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 can be shown to be

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

§1.6 Existence of non-measurable sets

Assuming the axiom of choice, there exists a non-measurable set of reals. 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 $x \sim y$ if $x - y \in Q$. Then, this gives equivalence classes $[x] = \{y \in E : x \sim y\}$ for all $x \in E$. Assuming the axiom of choice, we can select a representative of $[x]$ for each $x \in E$, and denote by S the set of such representatives. We can partition E into the union of its cosets, so $E = \bigcup_{q \in Q} (S + q)$ is a disjoint union.

Suppose S is a Borel set. Then $S + q$ is also a Borel set. We can therefore write

$$1 = \mu(E) = \mu\left(\bigcup_{q \in Q} (S + q)\right) = \sum_{q \in Q} \mu(S + q) = \sum_{q \in Q} \mu(S)$$

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

One can further show that μ 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 $x \in (0, 1]$.

§1.7 Probability spaces

Definition 1.17

If a measure space (E, \mathcal{E}, μ) has $\mu(E) = 1$, we call it a *probability space*, and 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. $\mathbb{P}(\Omega) = 1$;
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.

Definition 1.18

Events $A_i, i \in I$ 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 $\mathcal{A}_i, i \in I$ 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.

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

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

The *limit inferior* of A_n is

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

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

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

□

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, and $\sum_n \mathbb{P}(A_n) = \infty$. 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 $N \rightarrow \infty$, this approaches zero. Since $\bigcap_{m=n}^N A_m^c$ decreases to $\bigcap_{m=n}^{\infty} A_m^c$, by countable additivity we must have $\mathbb{P}(\bigcap_{m=n}^{\infty} A_m^c) = 0$. But then

$$\mathbb{P}(A_n \text{ infinitely often}) = \mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} A_m\right) = 1 - \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right) \geq 1 - \sum_n \mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right) = 1$$

Hence this probability is equal to one. □