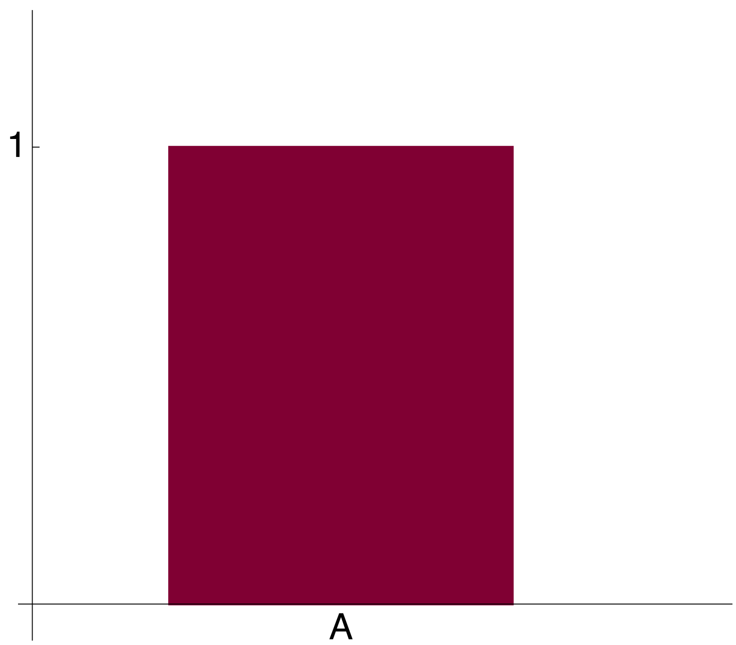
Measure Theory

Keith Ball



Contents

Introduction

Measure theory was developed in the late 19^{th} and early 20^{th} centuries to cope with problems that arose with the existing Riemann and other integrals. Suppose $f_n : [0,1] \to \mathbf{R}$ and $f : [0,1] \to \mathbf{R}$ satisfy $0 \le f_n(x) \le 1$ for all x and n and for all x

$$f_n(x) \to f(x).$$

We would like to know that if each f_n is integrable then so is f and

$$\int_0^1 f_n \to \int_0^1 f.$$

But f may not even be integrable.

For example, suppose (r_i) is an enumeration of the rationals in $[0,1]$. Set

$$f_n(x) = \begin{cases} 1 & \text{if } x = r_i \text{ for some } i \le n \\ \\ 0 & \text{otherwise.} \end{cases}$$

Clearly we can find very short intervals surrounding the first n rationals so f_n is Riemann integrable with integral 0. But $\lim f_n$ is equal to 1 at all rationals so its upper sums are all equal to 1. (**HW**)

The solution required a more coherent idea of length, area and volume than our simple intuitive one. The result turned out to be very much more useful than anyone could have predicted. It provided the tools to formalise probability theory and is fundamental to modern functional and harmonic analysis and to the theory of PDE.

The paradox of probability

Toss a coin repeatedly and for each n set H_n to be the number of heads obtained in the first n tosses. We expect $H_n/n \to 1/2$ as $n \to \infty$. So we certainly won't get the infinite sequence $(H, H, H, \ldots)$ with every toss a head. But no other sequence is any more likely

$$(H, T, H, H, T, H, T, T, \ldots).$$

Whatever sequence you think of, it won't happen. But something happens.

The goal

Our first goal will be to build a function λ (called Lebesgue measure) defined on as many subsets of $\mathbf{R}$ as we can, which measures their lengths. So for example

$$\lambda([a,b]) = b - a.$$

We would like to know that if A and B are disjoint then

$$\lambda(A \cup B) = \lambda(A) + \lambda(B).$$

One of the most crucial ideas in the theory was to ask for more than this. Suppose $A_1, A_2, \ldots$ is a countable sequence of disjoint sets: then

$$\lambda\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty \lambda(A_i).$$

For example

$$\begin{aligned} 1 &= \lambda\left([0,1)\right) = \lambda\left([0,1/2) \cup [1/2,3/4) \cup [3/4,7/8) \cup \cdots\right) \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \cdots . \end{aligned}$$

However the interval $[0,1]$ has length 1 and it is the union of singleton sets $\{0\}$, $\{1/3\}$, $\{1/2\}$ and so on, each of which has length 0. So we don't want Lebesgue measure to add up when we have **uncountably** many sets. This resolves the paradox of probability: there are uncountably many sequences of heads and tails: each one has measure zero but together they have measure (probability) 1.

A continuous function is defined on a topological space, a linear function on a vector space and so on. In a similar way a measure is defined on a collection $\mathcal{F}$ of sets with certain properties: in particular that if $(A_i)_1^\infty$ are disjoint members of $\mathcal{F}$ then so is their union. We shall begin by defining what are called σ-algebras, which are the appropriate collections, and then define measures. We then move on to the construction of Lebesgue measure: the measure we want on lots of subsets of $\mathbf{R}$.

Using Lebesgue measure we shall build a new integral more flexible than the Riemann integral and most importantly we shall prove the convergence theorems: for example

Theorem (An example)**.** *Let $f_n : [0,1] \to \mathbf{R}$ be a sequence of Lebesgue integrable functions which converge at each point and set*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for each x. Suppose that there is a Lebesgue integrable function g with $|f_n(x)| \leq g(x)$ for all n and x. Then f is Lebesgue integrable and

$$\int f_n \to \int f.$$

The rest of the course will consist of applications: for example, a new more general Fubini Theorem for double integrals.

Course plan

1. **σ-algebras and measures** Definitions and simple properties. The Borel sets.
2. **Lebesgue measure** Outer measures. The Key Lemma. Caratheodory's construction. Regularity. Uniqueness.
3. **Measurable functions** Definition and basic properties. Convergence.
4. **The integral** The construction. The convergence theorems.
5. **Product measures and Fubini** The construction. The $\pi - d$ Theorem. Uniqueness of extensions. Fubini's Theorem.
6. **L_p spaces** Definition and simple properties. Hölder's and Minkowski's Inequalities. Completeness.
7. **Approximation of measurable functions** Approximation in L_p. Lusin's Theorem.
8. **The Radon-Nikodym Theorem** Absolute continuity. Signed measures. The H-J decomposition. The R-N Theorem.

Chapter 1. σ-algebras and measures

Let Ω be a set: for example $\mathbf{R}$. We want to describe the families of subsets on which we can make sense of measures.

Definition (**Algebra**). *A collection $\mathcal{A}$ of subsets of a set Ω is an* **algebra** *if*

- $\emptyset \in \mathcal{A}$ *and* $\Omega \in \mathcal{A}$
- *If* $A \in \mathcal{A}$ *then* $A^c = \Omega \backslash A \in \mathcal{A}$
- *If* $A, B \in \mathcal{A}$ *then* $A \cup B \in \mathcal{A}$

Note that an algebra is also closed under finite intersections because

$$A \cap B = (A^c \cup B^c)^c.$$

As you would expect a σ-algebra has the souped up property: it is closed under **countable** unions instead of just finite ones

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

- $\emptyset \in \mathcal{F}$ *and* $\Omega \in \mathcal{F}$
- *If* $A \in \mathcal{F}$ *then* $\Omega \backslash A \in \mathcal{F}$
- *If* $A_1, A_2, A_3, \ldots \in \mathcal{F}$ *then* $\bigcup A_i \in \mathcal{F}$

If $\mathcal{F}$ is a σ-algebra and (A_i) is a countable sequence of members of $\mathcal{F}$ then the intersection $\bigcap A_i$ is also a member.

Example. *For any set Ω the collection of all subsets is a σ-algebra.*

Example. *For any set Ω and any subset A, the family*

$$\{\emptyset, A, A^c, \Omega\}$$

is a σ-algebra.

Example. *For $\Omega = \mathbf{R}$ the family of finite unions of half open intervals of the form $[a, b)$, $[a, \infty)$, $(-\infty, b)$ or $\mathbf{R}$ is an algebra but not a σ-algebra.* (**HW**)

Observe that if $\mathcal{F}$ is a σ-algebra and A and B are members. Then $B^c \in \mathcal{F}$ and hence

$$A \backslash B = A \cap B^c \in \mathcal{F}.$$

Generating σ-algebras and the Borel sets

As with almost all mathematical structures the properties of σ-algebras are automatically transferred to intersections.

Lemma (**Intersection of σ-algebras**)**.** *If $(\mathcal{F}_\alpha)_\alpha$ is an arbitrary collection of σ-algebras then*

$$\bigcap_\alpha \mathcal{F}_\alpha$$

is a σ-algebra.

Proof. For example, if (A_i) are in the intersection they belong to each $\mathcal{F}_\alpha$. So their union belongs to each $\mathcal{F}_\alpha$. □

This enables us to talk about σ-algebras generated by families.

Definition (**Generated σ-algebra**)**.** *For any family $\mathcal{U}$ of subsets of a set Ω there is a smallest σ-algebra on Ω including $\mathcal{U}$, known as the σ-algebra* **generated by $\mathcal{U}$** *and denoted $\sigma(\mathcal{U})$.*

Proof. The family of all subsets of Ω is a σ-algebra including $\mathcal{U}$. The intersection of all σ-algebras that include $\mathcal{U}$ is a σ-algebra that includes $\mathcal{U}$ and is clearly the smallest. □

Example. *For any set Ω and any subset A, the family*

$$\{\emptyset, A, A^c, \Omega\}$$

is the σ-algebra generated by $\{A\}$.

For our purposes the most important σ-algebra will be the σ-algebra generated by the open sets: the **Borel** σ-algebra. It contains all the open sets: and all the closed sets. It can also be generated by much smaller families: for example the family of open intervals.

Lemma (**Open sets and intervals**)**.** *Every open subset of $\mathbf{R}$ can be written as a countable union of open intervals.*

Proof. The family of all open intervals with rational ends (p, q) is countable. If U is an open set then for each point $x \in U$ we can find such an interval I with $x \in I \subset U$. The union of all these is U. □

So any σ-algebra that contains the open intervals must contain all open sets and hence the Borel σ-algebra. The Borel σ-algebra contains closed sets because they are the complements of open sets. In particular it contains singletons $\{x\}$. So it also contains countable sets like $\mathbf{Q}$.

Measures

Definition (Measure)**.** *Let $\mathcal{F}$ be a σ-algebra on a set Ω. A function $\mu : \mathcal{F} \to [0, \infty]$ is called a* **measure** *if*

- $\mu(\emptyset) = 0$
- *If $A_1, A_2, A_3, \ldots \in \mathcal{F}$ are disjoint then*

$$\mu\left(\bigcup A_i\right) = \sum \mu(A_i).$$

The last property is known as "countable additivity". The restriction to countable unions of sets is crucial. The real line is a union of singleton sets $\{x\}$ which have no measure but the line has infinite measure.

Once we define Lebesgue measure we will see that all countable sets have measure zero: for example the rationals. Note that the measure may take infinite values. We want the measure of $\mathbf{R}$ to be ∞ for example. Throughout the course we will repeatedly write sums like

$$\sum_1^\infty \lambda(A_i) = \infty.$$

As long as all the terms in the sum are non-negative there is no ambiguity or contradiction. It simply means that the terms of the sum are too big to add up. What we shall avoid doing is "cancelling infinities". We cannot write $\infty - \infty = 0$.

Example (**Exercise**)**.** *For any set Ω we may consider the σ-algebra of all subsets and define the measure of a set A to be the number of points in A.*

Example. *For any set Ω and any $x \in \Omega$ we may consider the σ-algebra of all subsets and define the measure of a set A as follows*

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

Example. *For* $\Omega = [0, 1)$ *and the* σ*-algebra*

$$\{\emptyset, [0, 1/2), [1/2, 1), [0, 1)\}$$

we may define the measures of the sets individually as $0, 1/2, 1/2, 1$.

Example. *For* $\Omega = \{H, T\}$ *and the* σ*-algebra of all subsets we may define the measures of the sets individually as* $\mu(\{H\}) = 1/2$, $\mu(\{T\}) = 1/2$ *(and* $\mu(\{H, T\}) = 1$*).*

More generally we refer to any function defined on a family of sets as a **set function**. A set function μ is called **monotone** if whenever $A \subset B$ we have $\mu(A) \leq \mu(B)$. A set function μ is called **countably subadditive** if for every sequence of sets (not necessarily disjoint)

$$\mu\left(\bigcup A_i\right) \leq \sum \mu(A_i).$$

Lemma (**Monotonicity and subadditivity of measures**)**.** *Every measure on a* σ*-algebra is monotone and countably subadditive.*

Proof. Let $\mathcal{F}$ be the σ-algebra. Suppose A and B are members of $\mathcal{F}$ and $B \subset A$. Then $A \backslash B \in \mathcal{F}$ and is disjoint from B. So

$$\mu(A) = \mu(B) + \mu(A \backslash B) \geq \mu(B).$$

Suppose (A_i) is a sequence of sets. For each i let B_i be the set A_i with all earlier sets removed:

$$B_i = A_i \backslash \left(\bigcup_1^{i-1} A_j\right).$$

Then $B_i \subset A_i$ for each i and the B_i are disjoint. So

$$\mu\left(\bigcup A_i\right) = \mu\left(\bigcup B_i\right) = \sum \mu(B_i) \leq \sum \mu(A_i).$$

□

Continuity of measure

You will see in **Assignment 1 Q 3** a key fact

Lemma (**The continuity of measure**)**.** *Suppose*

$$E_1 \subset E_2 \subset E_3 \subset \cdots$$

is an increasing sequence of measurable sets in $(\Omega, \mathcal{F}, \mu)$ *with* $E = \bigcup_1^\infty E_n$. *Then*

$$\mu(E) = \lim_n \mu(E_n).$$

Remark We sometimes write $E = \lim_n E_n$.

Formally our objects of study will be triples

$$(\Omega, \mathcal{F}, \mu)$$

where μ is a measure on a σ-algebra $\mathcal{F}$ of subsets of Ω. Such a triple is called a **measure space**. The most useful example we have so far is that in which Ω is any set, $\mathcal{F}$ is the family of all subsets and $\mu(A)$ is the number of points in A. This is called **counting measure**. A pair $(\Omega, \mathcal{F})$ of a set and a σ-algebra is sometimes known as a **measurable space**.

Our first major goal will be to define a measure λ on the Borel sets which has the property that assigns each interval its length.

$$\lambda([a, b]) = b - a.$$

It turns out that such a measure cannot be defined on **all** subsets of $\mathbf{R}$. Instead we define a function on all subsets that takes the correct value on open sets but which is not a measure. We then show that when restricted to the Borel sets it **is** a measure.

Chapter 2. Lebesgue measure

We shall carry out the following programme:

1. We define $\lambda(U)$ for each open set U.
2. For each $A \subset \mathbf{R}$ we set
$$\lambda^*(A) = \inf_{U \supset A} \lambda(U)$$
where the infimum is taken over all open sets in $\mathbf{R}$ that include A.
3. We prove that λ^* is (monotone and) countably subadditive.
4. We show that if you restrict λ^* to the Borel sets then it is measure.

The measure of open sets

We saw that each open set is a countable union of open intervals. Actually we can be much more precise.

Lemma (Open sets and disjoint intervals). *Every open subset of* $\mathbf{R}$ *can be written in an unique way as a countable disjoint union of open intervals.*

Proof. We know the set is a countable union of open intervals. Suppose our set is $U = I_1 \cup I_2 \cup I_3 \cdots$. To find the decomposition we want we are going to expand I_1 as much as possible while staying in U. Pick some $x \in I_1$ and choose the largest number below x, not in U: call it a_1. (If all numbers below x belong to U then set $a_1 = -\infty$.) Similarly let b_1 be the least number of $[x, \infty) \cap U^c$. Then $I_1 \subset (a_1, b_1) \subset U$ while $a_1, b_1 \notin U$.

Each interval I_j must be inside (a_1, b_1) or disjoint from it since otherwise it would contain a_1 or b_1. Throw out all the intervals inside (a_1, b_1). Repeat the process with the first unused interval in the list. Continue in this way to produce a countable family of open intervals whose union includes all the I_j but are disjoint.

Now, to show uniqueness, observe that if $U = J_1 \cup J_2 \cup \cdots$ is a disjoint union of intervals then the end points of each interval are not in U. If an end point of J_7 were in U it would be in another interval which would not be disjoint from J_7. Now suppose we had two decompositions
$$I_1 \cup I_2 \cup \cdots = J_1 \cup J_2 \cup \cdots .$$
Each I_k must be a subset of some J_m or disjoint from it, otherwise it would contain an end point. But if $I_k \subset J_m$ then since they are not disjoint we also have $J_m \subset I_k$. So the decompositions are the same. □

Note that we can't prove this by choosing the interval "nearest to" the one we just did. There may be no such interval. If you remove the points 0 and $(1/n)$ for each natural number n from the interval $(-1, 1)$ you get an open set which is the union of $(-1, 0)$ and a sequence of open intervals decreasing to 0. The Cantor set that you saw last year and that we shall discuss later is obtained by removing countably many open intervals from $[0, 1]$. It consists of uncountably many points "between them".

If we **can** build the measure we want then it must be that the measure of an open set is the sum of the lengths of the disjoint intervals that make it up.

Definition (**The measure of an open set**)**.** *For the open set $U = I_1 \cup I_2 \cup I_3 \cdots$ written as a disjoint union of open intervals, we define*

$$\lambda(U) = \sum_{1}^{\infty} \mathit{length}(I_i).$$

Note that the measure can be infinity. We want to "extend" the definition to the Borel σ-algebra.

The outer measure

Definition (**Lebesgue outer measure**)**.** *For each subset A of $\mathbf{R}$ define the Lebesgue outer measure of A to be*

$$\lambda^*(A) = \inf_{U \supset A} \lambda(U)$$

where the infimum is taken over all open subsets of $\mathbf{R}$ that include A.

So we approximate the measure of a set by using open sets around it.

Our aim will be to show that when we restrict λ^* to the Borel sets we get a measure. But we also need to know that we get the right value for each interval. Could it be that there is an open set U containing $(0, 1)$ but with $\lambda(U) < 1$? No, because one of the disjoint intervals making up U must contain $(0, 1)$. So it is clear that for each open interval $\lambda^*(I) = \lambda(I) = |I|$.

What is not quite so clear is that $\lambda^*(U) = \lambda(U)$ for **all** open sets. It might be that you could have an open set U in $(0,1)$ but with $\lambda(U) > 1$. Ruling this out will be **HW** but it will be automatically ruled out once we know that λ^* is a measure on the Borel sets.

If we are able to extend our measure λ from open sets to the Borel sets then the result will be monotone and countably subadditive. So it would be good to know that these properties hold for the outer measure. To do this we need the following Key Lemma.

Lemma (**Key lemma I**)**.** *Suppose a compact interval $[a,b]$ is covered by the union of finitely many open intervals (a_i, b_i) for $1 \leq i \leq m$. Then*

$$b - a \leq \sum_1^m (b_i - a_i).$$

Proof. We shall use induction on m. If $m = 1$ there is nothing to prove. Assume that (a_m, b_m) is the interval that contains b. If $a_m < a$ then

$$b_m - a_m \geq b - a$$

and so

$$\sum_1^m (b_i - a_i) \geq b - a.$$

If not, then the interval $[a, a_m]$ is covered by the other $m-1$ intervals and by the inductive hypothesis we have

$$\sum_1^{m-1} (b_i - a_i) \geq a_m - a.$$

Putting back the m^{th} interval we get

$$\sum_1^m (b_i - a_i) \geq a_m - a + b_m - a_m = b_m - a \geq b - a.$$

□

The set **Q** of rationals can be covered with countably many open intervals whose lengths add up to less than 0.00001. In the preceding proof we did not use the completeness property of the reals so it would work equally well for the rationals. So we cannot cover the rationals in $[0,1]$ with **finitely many** intervals whose lengths add up to less than 1. But the fact that we **can** cover the rationals with countably many shows that we need completeness for the next really crucial fact.

Lemma (Key lemma II)**.** *If an open interval J is covered by the union of countably many open intervals $I_1, I_2, \ldots$ then*

$$\lambda(J) \leq \sum \lambda(I_i).$$

Proof. Suppose $J = (a, b)$ and choose $\varepsilon > 0$. The closed interval $[a + \varepsilon, b - \varepsilon]$ is covered by the open intervals (I_i) and hence by some finite number of them. So by the previous lemma

$$\sum_i \lambda(I_i) \geq b - \varepsilon - (a + \varepsilon) = b - a - 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary

$$\sum \lambda(I_i) \geq b - a.$$

If on the other hand J has infinite length then for each M we can find a compact interval of length M inside J and get

$$\sum \lambda(I_i) \geq M.$$

Now since M was arbitrary we have

$$\sum \lambda(I_i) = \infty.$$

□

We almost have enough to prove the subadditivity of the outer measure. The last version is this.

Lemma (Key lemma III)**.** *If $U = \bigcup I_i$ is the union of countably many open intervals $I_1, I_2, \ldots$ (not necessarily disjoint) then*

$$\lambda(U) \leq \sum \lambda(I_i).$$

Proof. Write $U = \bigcup D_k$ as a union of disjoint open intervals so that

$$\lambda(U) = \sum \lambda(D_k).$$

Since the endpoints of D_k are not in U, each I_j is either inside a D_k or disjoint from it. So each D_k is a union of some subfamily of the I_j. By relabelling we may write the sequence as a doubly indexed sequence $(I_{k,j})$ where for each k

$$D_k = \bigcup_{j=1}^{\infty} I_{k,j}.$$

By the Key Lemma II we have

$$\lambda(D_k) \le \sum_j \lambda(I_{k,j})$$

and hence

$$\lambda(U) = \sum \lambda(D_k) \le \sum_{k,j} \lambda(I_{k,j})$$

as required. □

We shall show that λ^* is monotone, countably subadditive and satisfies $\lambda^*(\emptyset) = 0$. Any non-negative set function with these properties is called an **outer measure**. So, somewhat oddly, we shall show that Lebesgue outer measure is an outer measure. We shall then prove a general theorem showing how to restrict an outer measure to a σ-algebra on which it is a measure and check that in the case of Lebesgue measure, this σ-algebra includes the Borel sets.

Definition (**Outer measure**). *Let Ω be a set. A function μ^* defined on the subsets of Ω and taking values in $[0, \infty]$ is called an* **outer measure** *if*

- $\mu^*(\emptyset) = 0$
- *If $A \subset B$ then $\mu^*(A) \le \mu^*(B)$*
- *If $A_1, A_2, A_3, \ldots \subset \Omega$ then*

$$\mu^*\left(\bigcup A_i\right) \le \sum \mu^*(A_i).$$

Theorem (**Lebesgue outer measure is an outer measure**). *λ^* is an outer measure.*

Proof. Since $\emptyset$ is open and $\lambda(\emptyset) = 0$ we have

$$\lambda^*(\emptyset) = 0.$$

Suppose $B \subset A$ and $\varepsilon > 0$. Find an open set $U \supset A$ with

$$\lambda(U) < \lambda^*(A) + \varepsilon.$$

Then since $B \subset U$ we have

$$\lambda^*(B) \le \lambda(U) < \lambda^*(A) + \varepsilon.$$

Since the outermost inequality holds for all $\varepsilon > 0$ we have $\lambda^*(B) \le \lambda^*(A)$.

Suppose A_i is a sequence of subsets of $\mathbf{R}$. Given $\varepsilon > 0$ choose open sets $U_1, U_2, \ldots$ so that for each i, $U_i \supset A_i$ but

$$\lambda(U_i) \le \lambda^*(A_i) + \frac{\varepsilon}{2^i}.$$

Let $U = \bigcup U_i$ so that U is an open set with $U \supset \bigcup A_i$. Now for each i write $U_i = \bigcup I_{i,j}$ as a union of disjoint open intervals so that

$$\lambda(U_i) = \sum \lambda(I_{i,j}).$$

Then

$$U = \bigcup_{i,j} I_{i,j}$$

and so by the Key Lemma

$$\lambda(U) \le \sum_{ij} \lambda(I_{i,j}) = \sum_i \lambda(U_i).$$

Therefore

$$\begin{aligned} \lambda^*\left(\bigcup A_i\right) &\le \lambda(U) \le \sum \lambda(U_i) \\ &\le \sum (\lambda^*(A_i) + \varepsilon/2^i) = \sum \lambda^*(A_i) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary.

$$\lambda^*\left(\bigcup A_i\right) \le \sum \lambda^*(A_i).$$

□

Note that the monotonicity of λ^* shows that for every interval R, whether it be open, closed, half-open,... we have that $\lambda^*(R)$ is the length of R. **Exercise**

The 4^{th} step of our programme was to show that λ^* restricted to the Borel sets is a measure. This step divides into two parts. The first part is an abstract restriction theorem and the second part consists in showing that for the outer measure λ^*, the Borel sets satisfy the conditions of the first part. The first part will consist of showing that for every outer measure there is a special σ-algebra of sets on which it is a measure. The idea for identifying these special sets comes from the ordinary integral.

Idea The outer measure is a bit like the upper integral in Riemann theory. You take the infimum of things above your function. In that theory a function is "nice" if the upper and lower integral are the same. $\lambda^*(A)$ approximates A from the outside. To approximate from the inside we could use the outer measure of the complement.

$$\text{Inside}(A) = \Omega - \text{Outside}(\Omega \backslash A).$$

So

$$\text{inner}(A) = \lambda^*(\Omega) - \lambda^*(\Omega \backslash A).$$

For this to be the same as the outer measure we need

$$\lambda^*(A) = \lambda^*(\Omega) - \lambda^*(\Omega \backslash A).$$

So good sets should satisfy

$$\lambda^*(\Omega) = \lambda^*(A) + \lambda^*(\Omega \backslash A).$$

In fact it is convenient to ask for a bit more.

Definition (**μ^*-measurability**). *Suppose μ^* is an outer measure on the subsets of Ω. A subset $A \subset \Omega$ is called* **μ^*-measurable** *if for every $T \subset \Omega$*

$$\mu^*(T) = \mu^*(T \cap A) + \mu^*(T \backslash A).$$

If $A \subset [0,1]$ had the property that

$$\lambda^*(A) + \lambda^*([0,1] \backslash A) > 1$$

we would not want to put A into our family of good sets.

Caratheodory's restriction theorem

Theorem (**Caratheodory's restriction theorem**). *Suppose μ^* is an outer measure on the subsets of Ω. Then the family $\mathcal{F}$ of μ^*-measurable sets is a σ-algebra and μ^* restricted to this σ-algebra is a measure.*

Proof. $\mathcal{F}$ is an algebra. Clearly $\emptyset$ and Ω are measurable since for example, for any T

$$\mu^*(T \cap \emptyset) + \mu^*(T \backslash \emptyset) = \mu^*(\emptyset) + \mu^*(T) = \mu^*(T).$$

If A is measurable then its complement is too because the definition of measurability is self-complementary. We start by checking that $\mathcal{F}$ is closed under finite unions.

Suppose A and B are measurable. We want to test the union $A \cup B$ with a set T. By subadditivity we already have

$$\mu^*(T) \leq \mu^*(T \cap (A \cup B)) + \mu^*(T \backslash (A \cup B)).$$

To go the other way, note that

$$T \cap (A \cup B) = (T \cap A) \cup [(T \backslash A) \cap B]$$

and that

$$T \backslash (A \cup B) = (T \backslash A) \backslash B.$$

Hence by subadditivity

$$\mu^*(T \cap (A \cup B)) + \mu^*(T \backslash (A \cup B)) \leq \mu^*(T \cap A) + \mu^*((T \backslash A) \cap B) + \mu^*((T \backslash A) \backslash B).$$

The last two terms are what you get when you test the measurability of B using $T \backslash A$ instead of T. So their sum is equal to $\mu^*(T \backslash A)$. So the full sum is at most

$$\mu^*(T \cap A) + \mu^*(T \backslash A)$$

and this is $\mu^*(T)$ by the measurability of A.

We now know that $\mathcal{F}$ is closed under finite unions and complements. That means it is also closed under finite intersections and hence differences $A \backslash B = A \cap (\Omega \backslash B)$.

Disjointification. Now suppose that (A_i) is sequence of measurable sets. We want to know that $\bigcup A_i$ is measurable. For each i we set

$$B_i = A_i \backslash \left(\bigcup_{j \leq i-1} A_j \right)$$

and then all the B_i belong to $\mathcal{F}$, the B_i are disjoint and $\bigcup B_i = \bigcup A_i$. So it suffices to prove the measurability of **disjoint** unions.

To complete the proof we shall show that if (B_i) is a sequence of disjoint measurable sets then $\bigcup B_i$ is measurable and

$$\mu^* \left(\bigcup B_i \right) = \sum \mu^*(B_i).$$

Finite additivity. Let $T \subset \Omega$. Set $B = \bigcup B_i$ and for each n set $F_n = \bigcup_1^n B_i$. We want to check that $\mu^*(T) = \mu^*(T \cap B) + \mu^*(T \backslash B)$. We shall begin by showing that for each n

$$\mu^*(T \cap F_n) = \sum_1^n \mu^*(T \cap B_k).$$

Since $F_1 = B_1$ we know that this is true when $n = 1$.

Assume inductively that $\mu^*(T \cap F_n) = \sum_1^n \mu^*(T \cap B_k)$ and test the measurability of F_n with $T \cap F_{n+1}$. We get

$$\begin{aligned} \mu^*(T \cap F_{n+1}) &= \mu^*(T \cap F_{n+1} \cap F_n) + \mu^*((T \cap F_{n+1}) \backslash F_n) \\ &= \mu^*(T \cap F_n) + \mu^*(T \cap B_{n+1}). \end{aligned}$$

Therefore

$$\mu^*(T \cap F_{n+1}) = \sum_1^{n+1} \mu^*(T \cap B_k)$$

giving the inductive step.

Countable additivity We have $\mu^*(T \cap F_n) = \sum_1^n \mu^*(T \cap B_k)$ for all n. Hence for each n

$$\mu^*(T \cap B) \geq \mu^*(T \cap F_n) = \sum_1^n \mu^*(T \cap B_k).$$

Letting $n \to \infty$ we have

$$\mu^*(T \cap B) \geq \sum_1^\infty \mu^*(T \cap B_k).$$

The reverse is true by subadditivity so we have

$$\mu^*(T \cap B) = \sum_1^\infty \mu^*(T \cap B_k).$$

If we put $T = \Omega$ we get the additivity property we want

$$\mu^*(B) = \sum_1^\infty \mu^*(B_k).$$

We just need to complete the proof that B is measurable.

Countable unions By the measurability of the finite unions F_n we have

$$\begin{aligned}\mu^*(T) &= \mu^*(T \cap F_n) + \mu^*(T \backslash F_n) \\ &\geq \mu^*(T \cap F_n) + \mu^*(T \backslash B) \\ &= \sum_1^n \mu^*(T \cap B_k) + \mu^*(T \backslash B).\end{aligned}$$

Since this is true for every n we have

$$\mu^*(T) \geq \sum_1^\infty \mu^*(T \cap B_k) + \mu^*(T \backslash B) = \mu^*(T \cap B) + \mu^*(T \backslash B).$$

□

In order to finish our programme we need to check that the Borel sets are λ^*-measurable. Note that in the end we don't really care whether the measurable sets form a σ-algebra: merely that Borel sets are among them. However we can't write down a generic Borel set and check whether it is measurable. So we need to use the fact that the measurable sets form a σ-algebra. Recall from **HW** that the Borel σ-algebra is generated by the family of half-infinite intervals of the form $(-\infty, a)$.

If you want to show that Borel sets are good, prove that the good sets form a σ-algebra and that intervals are good.

Theorem (Borel sets are λ^*-measurable). *Borel subsets of* **R** *are λ^*-measurable.*

Proof. The measurable sets form a σ-algebra so it suffices to prove that half-infinite intervals $(-\infty, a)$ are measurable. Suppose $T \subset \mathbf{R}$ and I is an interval of the form $(-\infty, a)$. Given $\varepsilon > 0$ choose an open set $V \supset T$ with

$$\lambda(V) < \lambda^*(T) + \varepsilon.$$

Now write $V = \bigcup J_j$ as a countable disjoint union of open intervals.

For each j, the sets $J_j \cap I$ and $J_j \backslash I$ are intervals so we know their measures are their lengths and so

$$\lambda(J_j) = \lambda(J_j \cap I) + \lambda^*(J_j \backslash I).$$

Therefore

$$\begin{aligned}\lambda(V) &= \sum \lambda(J_j) = \sum \lambda(J_j \cap I) + \sum \lambda^*(J_j \backslash I) \\ &\geq \lambda(V \cap I) + \lambda^*(V \backslash I).\end{aligned}$$

Therefore

$$\lambda^*(T \cap I) + \lambda^*(T \backslash I) \leq \lambda(V \cap I) + \lambda^*(V \backslash I) \leq \lambda(V) < \lambda^*(T) + \varepsilon.$$

Since ε was arbitrary we have

$$\lambda^*(T \cap I) + \lambda^*(T \backslash I) \leq \lambda^*(T)$$

as required. □

Theorem (Lebesgue measure). *There is a measure λ on the Borel sets with the property that for each interval*

$$\lambda((a, b)) = b - a.$$

Remark It is more conventional to define the outer measure as

$$\lambda^*(A) = \inf\{\sum |I_j| : A \subset \bigcup I_j\}$$

where the infimum is taken over all countable families of intervals whose union includes A. Moreover instead of open intervals the construction usually uses half-open ones, for reasons that will be clear later. In the conventional formulation it is quite a bit easier to prove subadditivity since you don't need the key lemma at this stage: and you don't need to prove that open sets have unique expressions as countable disjoint unions of open intervals (see Assignment I). But it is harder to prove that intervals are λ^*-measurable and harder to prove that intervals have the correct outer measure: this is where you need the key lemma. You also have to prove **regularity** (which we come to in the next section).

However the main advantage of the interval covering approach is that it works equally well for higher dimensions: $\mathbf{R}^d$. The main advantage of the open set approach is that it gives a more intuitive understanding of what the measure is. Defining Lebesgue measure is not the problem (at least in hindsight). The problem is checking that it **is** a measure on a σ-algebra containing intervals.

Properties of Lebesgue measure

We defined the measure of a set $B \subset \mathbf{R}$ to be

$$\inf\{\lambda(U) : U \supset B, U \text{ is open}\}.$$

Thus we can approximate a set (in measure) from the outside by open sets. An obvious question is whether we can approximate from the inside by closed or better still compact sets. A measure on a topological space with these two properties is called a **regular** measure.

Lemma (**Regularity of Lebesgue measure**)**.** *Let B be a λ^*-measurable set. Then*

$$\begin{aligned} \lambda(B) &= \inf\{\lambda(U) : U \supset B, U \textit{ is open}\} \\ &= \sup\{\lambda(K) : K \subset B, K \textit{ is compact}\}. \end{aligned}$$

Proof. The first is automatic from our definition. For the second observe that

$$\lambda(B) = \lim_{M\to\infty} \lambda(B \cap [-M, M])$$

by the continuity of measure. So it suffices to assume that $B \subset [-M, M]$ is included in some finite interval.

Let $A = [-M, M]\backslash B$ and given $\varepsilon > 0$ find an open subset U of $\mathbf{R}$ including A with

$$\lambda(U) < \lambda(A) + \varepsilon.$$

Then $K = [-M, M]\backslash U$ is a compact set inside B and

$$\lambda(B) = 2M - \lambda(A) < 2M - \lambda(U) + \varepsilon = \lambda(K) + \varepsilon.$$

□

It is easy to see that Lebesgue measure is the only measure on the Borel sets with the property that $\mu((a, b)) = b - a$ for every open interval. If μ has this property then it must measure each open set correctly. Hence for every Borel set B

$$\begin{aligned} \lambda(B) &= \inf\{\lambda(U) : U \supset B, U \text{ is open}\} \\ &= \inf\{\mu(U) : U \supset B, U \text{ is open}\} \geq \mu(B). \end{aligned}$$

Now if B is inside a bounded interval $(-M, M)$ then by applying this inequality to $(-M, M)\backslash B$ we get the reverse. Now use the continuity of measure again.

Lemma (**Translation invariance of Lebesgue measure I**)**.** *Let B be a λ^*-measurable set and $x \in \mathbf{R}$. Then $\lambda(B + x) = \lambda(B)$.*

Proof. This is obviously true for intervals, hence for open sets and hence for arbitrary sets. □

Lemma (**Translation invariance of Lebesgue measure II**)**.** *Suppose μ is a translation invariant measure on the Borel sets of $\mathbf{R}$ for which $\lambda([0, 1]) = 1$. Then $\mu = \lambda$.*

Proof. If any singleton had positive measure then they all would. But we can fit infinitely many inside $[0, 1]$. So singletons have measure 0. Now for each n we can express $[0, 1]$ as a disjoint union of n open intervals of length $1/n$ and some singletons whose total measure is zero.

$$[0, 1] = (0, 1/n) \cup (1/n, 2/n) \cup \cdots \cup (1 - 1/n, 1) \cup \cdots .$$

The intervals of length $1/n$ all have the same measure so it must be $1/n$. Now we can approximate any interval by short ones and check that it has the right measure. By the preceding uniqueness observation, $\mu = \lambda$. □

Lemma (**Scale invariance of Lebesgue measure**)**.** *Suppose $k > 0$ and A is a measurable subset of $\mathbf{R}$. Then*

$$\lambda(kA) = k\lambda(A).$$

Proof. This is obviously true for intervals, hence for open sets and hence for arbitrary sets. □

Push forward or image measures

Recall that if $f : \Omega \to \Psi$ and $A \subset \Psi$ then the pre-image $f^{-1}(A)$ of A (under f) is the set

$$f^{-1}(A) = \{x \in \Omega : f(x) \in A\}$$

of things that f sends into A.

Lemma (**Image measures**)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space. Suppose that* Ψ *is another set and* $f : \Omega \to \Psi$ *is a function. Let* $\mathcal{G}$ *be the family of subsets of* Ψ *whose pre-images are in* $\mathcal{F}$*:*

$$\mathcal{G} = \{G \subset \Psi : f^{-1}(G) \in \mathcal{F}\}.$$

Then $\mathcal{G}$ *is a* σ*-algebra on* Ψ*.*

If we set

$$\nu(G) = \mu(f^{-1}(G))$$

for each $G \in \mathcal{G}$ *then* ν *is a measure.*

Proof. Firstly $f^{-1}(\emptyset) = \emptyset \in \mathcal{F}$ so $\emptyset \in \mathcal{G}$. If (A_i) are sets in $\mathcal{G}$ then

$$f^{-1}\left(\bigcup A_i\right) = \bigcup f^{-1}(A_i) \in \mathcal{F}$$

and so $\bigcup A_i \in \mathcal{G}$. Finally if $A \in \mathcal{G}$ then

$$f^{-1}(\Psi \backslash A) = \Omega \backslash f^{-1}(A) \in \mathcal{F}$$

and so $\Psi \backslash A \in \mathcal{G}$.

Clearly $\nu \geq 0$ and

$$\nu(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0.$$

If (A_i) are disjoint sets in $\mathcal{G}$ then

$$\nu\left(\bigcup A_i\right) = \mu\left(f^{-1}\left(\bigcup A_i\right)\right) = \mu\left(\bigcup f^{-1}(A_i)\right).$$

If A and B are disjoint then $f^{-1}(A)$ and $f^{-1}(B)$ are disjoint because a point in their intersection would have to map into two different places: one in A and one in B. So the last expression is

$$\sum \mu\left(f^{-1}(A_i)\right) = \sum \nu(A_i).$$

□

Restriction to subsets and the circle

Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and $A \in \mathcal{F}$. Then we can consider a new measure space with A as the ground set, the σ-algebra consisting of those members of $\mathcal{F}$ which are subsets of A and μ restricted to this σ-algebra (**Exercise**). Thus $[0, 1]$ becomes a measure space equipped with Lebesgue measure: so does $(0, \infty)$ and so on.

By the Image Measure Lemma we can transfer Lebesgue measure on $[0,1]$ to the circle in $\mathbf{R}^2$ via the map

$$x \mapsto (\cos 2\pi x, \sin 2\pi x)$$

or to the circle in the complex plane via

$$x \mapsto e^{2\pi i x}.$$

This measure is rotation invariant because of the translation invariance of Lebesgue measure. **Exercise**

A non-measurable set of Vitali

I have repeatedly said that we cannot extend Lebesgue measure to all the subsets of $\mathbf{R}$. Let's see that there are some sets that are not λ^*-measurable (**Sketch**).

Define an equivalence relation on $\mathbf{R}$ by $x \sim y$ if $x - y \in \mathbf{Q}$. This breaks the reals into equivalence classes, $\mathbf{Q}$, $\mathbf{Q} + \sqrt{2}$ and so on. Choose a subset W of $[0,1)$ which contains one member of each equivalence class. No two points of W differ by a rational number. The translations of W by rationals are disjoint.

Consider the image $\tilde{W}$ of this in the circle. For each rational number $r \in [0,1)$, we consider the rotation of this set by an angle $2\pi r$. There are countably many of these sets, they are disjoint and their union is the circle. If one of them were measurable then they all would be and they would have equal measure. But if that were α we would have an infinite sum

$$\alpha + \alpha + \cdots = 1$$

which is impossible.

The Lebesgue measurable sets

When we have covered the Devil's staircase I will remark that there are sets that are Lebesgue measurable but not Borel sets. For the moment we shall content ourselves with a description of the measurable sets.

Lemma (**The Lebesgue measurable sets**)**.** *A subset of* $\mathbf{R}$ *is* λ^**-measurable if and only if it can be written as a union*

$$A = N \cup \bigcup_1^\infty F_n$$

where $\lambda^*(N) = 0$ and the F_n are closed.

Proof. Firstly if $\lambda^*(N) = 0$ and $T \subset \mathbf{R}$ then

$$T \cap N \subset N \quad \text{and} \quad T \backslash N \subset T$$

and so

$$\lambda^*(T \cap N) + \lambda^*(T \backslash N) \leq 0 + \lambda^*(T) = \lambda^*(T).$$

So N is measurable and hence so is any union of the form $N \cup \bigcup F_n$.

On the other hand by the regularity of Lebesgue measure, for each n we can find a closed set $F_n \subset A \cap [-n, n]$ with

$$\lambda((A \cap [-n, n]) \backslash F_n) \leq \frac{1}{2^n}.$$

Clearly $\bigcup F_n \subset A$. Let $N = A \backslash (\bigcup F_n)$. We know that $\lambda(N) = \lim_n \lambda(N \cap [-n, n])$ but for each n

$$N \cap [-n, n] \subset (A \cap [-n, n]) \backslash F_n)$$

so the limit is 0. □

Among other things what we see that any set of outer measure 0 is automatically a Lebesgue set (of measure 0). Countable sets are automatically Borel but it isn't too hard to construct uncountable sets of measure 0.

Cantor's middle thirds set

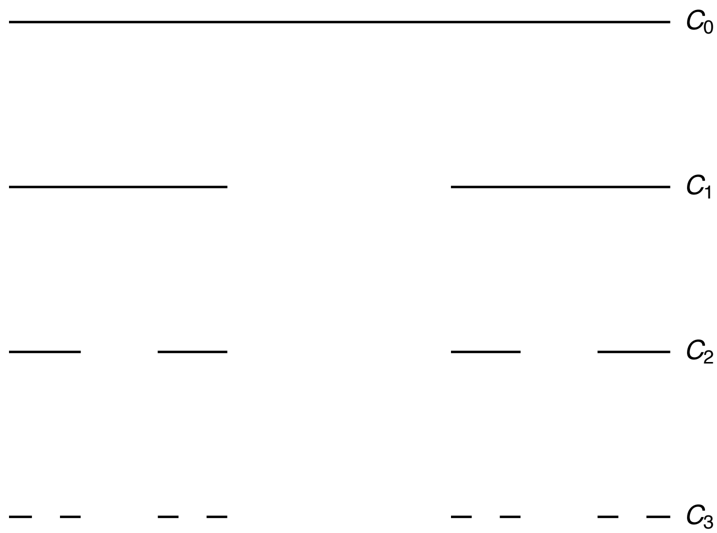
Recall from N, M and T. Suppose we start with $C_0 = [0, 1]$. Now to create C_1 remove open the middle third $(1/3, 2/3)$. So

$$C_1 = [0, 1/3] \cup [2/3, 1].$$

To find C_2 remove each middle third of the intervals in C_1.

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

If we continue in this way then for each n, we get a closed set C_n whose measure is $(2/3)^n$.

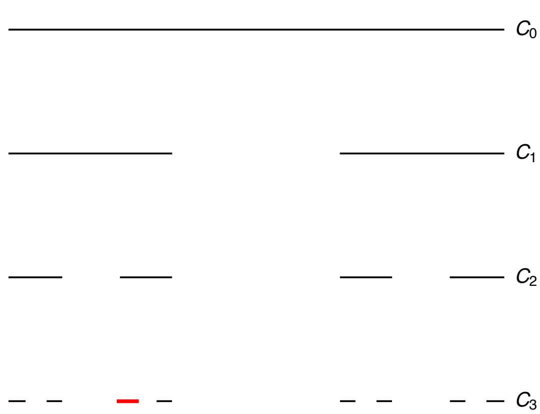


Their intersection C is a closed set of measure 0. We just need to check that the intersection is uncountable. Recall that any decreasing sequence of closed bounded intervals

$$F_0 \supset F_1 \supset F_2 \supset \cdots$$

has a point in the intersection.

The points of C are intersections of such sequences: one interval at each level.



At each stage we can choose left or right and the sequence of choices determines the point in the intersection.

Different choices lead to different points.

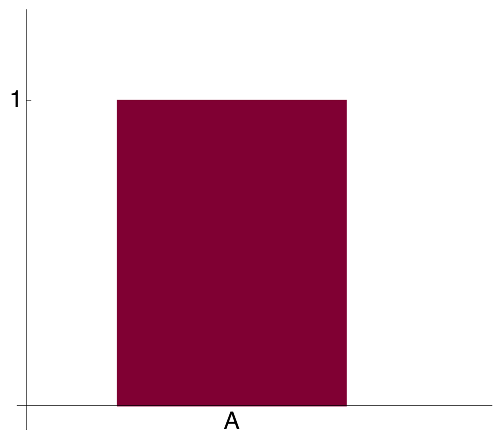
So there are as many points in C as there sequences

$$LRL\ldots.$$

This is an uncountable set by the Cantor diagonal argument.

Chapter 3. Measurable functions

Once we have a measure space $(\Omega, \mathcal{F}, \mu)$ we can consider functions $f : \Omega \to \mathbf{R}$ and try to integrate them. In particular if $A \in \mathcal{F}$ then the indicator function $\mathbf{1}_A$ should have integral $\mu(A)$.



$$\int_\Omega \mathbf{1}_A \, d\mu = \mu(A).$$

Note that if A were not measurable then we would not know how to integrate its indicator. So we need to restrict attention to functions that "fit" with the σ-algebra.

The definition and basic properties

The definition we need is somewhat similar to the definition of continuity for maps on a topological space.

Definition (Measurable functions)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \mathbf{R}$ a function. We say that f is measurable (or Borel-measurable) if for every Borel set $B \subset \mathbf{R}$ we have*

$$f^{-1}(B) \in \mathcal{F}.$$

Example. *The indicator function $f = \mathbf{1}_A$ of a measurable set A in a space $(\Omega, \mathcal{F})$ is a measurable function.*

Proof. Let B be a Borel set in $\mathbf{R}$. Then

$$\begin{cases} f^{-1}(B) = \Omega & \text{if } 0,1 \in B \\ f^{-1}(B) = A & \text{if } 1 \in B \text{ but } 0 \notin B \\ f^{-1}(B) = \Omega \backslash A & \text{if } 0 \in B \text{ but } 1 \notin B \\ f^{-1}(B) = \emptyset & \text{if } 0,1 \notin B \end{cases}$$

Lemma (**Inverse images and σ-algebras**)**.** *Suppose $f : \Omega \to \mathbf{R}$ as above. Then the family*

$$\mathcal{G} = \left\{B \subset \mathbf{R} : f^{-1}(B) \in \mathcal{F}\right\}$$

is a σ-algebra.

Proof. This is part of the Image Measure Lemma from Chapter 2. □

We shall use this to check that, for example, if you know that

$$\{x : f(x) < a\}$$

is measurable for all $a \in \mathbf{R}$ then you know that f is measurable. You don't have to check

$$\{x : f(x) \in B\}$$

is measurable for every Borel set B.

Recall from **HW 1** that the Borel σ-algebra is generated by sets of the form $(-\infty, a)$.

Lemma (**Measurable functions and open sets**)**.** *Suppose $f : \Omega \to \mathbf{R}$ as above. Then f is measurable if one of the following holds*

- $f^{-1}((-\infty, a)) = \{x : f(x) < a\} \in \mathcal{F}$ *for every* $a \in \mathbf{R}$.
- $f^{-1}((-\infty, a]) \in \mathcal{F}$ *for every* $a \in \mathbf{R}$.
- $f^{-1}((a, \infty)) \in \mathcal{F}$ *for every* $a \in \mathbf{R}$.
- $f^{-1}([a, \infty)) \in \mathcal{F}$ *for every* $a \in \mathbf{R}$.

Proof. Of the first. We know that $\Sigma = \{B \subset \mathbf{R} : f^{-1}(B) \in \mathcal{F}\}$ contains the intervals $(-\infty, a)$. Since the family is a σ-algebra it contains the Borel sets. □

Essential remark As usual we don't pick a Borel set B, try to write it in terms of intervals and then apply f^{-1} to show that B is "good". Instead we check that the family of good sets is a σ-algebra that contains intervals.

Corollary (Continuous functions are measurable). *Suppose Ω is a topological space and $\mathcal{F}$ consists of its Borel sets. If $f : \Omega \to \mathbf{R}$ is continuous then f is measurable.*

Proof. If $U \subset \mathbf{R}$ is open then $f^{-1}(U)$ is open and hence in $\mathcal{F}$. By the inverse image lemma, f is measurable □

Just as with continuous functions we want to be able to combine functions.

Corollary (Algebra of measurable functions). *Suppose $(\Omega, \mathcal{F})$ is a measurable space, $f, g : \Omega \to \mathbf{R}$ are measurable and $\phi : \mathbf{R} \to \mathbf{R}$ is measurable. Then so are $f + g$, fg, $\phi \circ f$, $|f|$, $\max(f, g)$ and $\min(f, g)$.*

In the spirit of probability theory we shall use the notation $(f < a)$ to mean $f^{-1}((-\infty, a)) = \{\omega \in \Omega : f(\omega) < a\}$.

Proof. We shall just do $f + g$. Suppose $r \in \mathbf{R}$. I claim that the set $(f + g < r)$ can be written as a countable union of sets built from f and g:

$$\bigcup_{q \in \mathbf{Q}} [(f < q) \cap (g < r - q)] .$$

Clearly if ω belongs to one of these intersections then $f(\omega) < q$ and $g(\omega) < r - q$ so $(f + g)(\omega) < r$.

On the other hand if $f(\omega) + g(\omega) = r - \delta$ then choose a rational q with

$$q - \delta < f(\omega) < q.$$

We have

$$g(\omega) = r - \delta - f(\omega) < r - \delta - (q - \delta) = r - q.$$

So for this particular q the point ω belongs to the intersection

$$(f < q) \cap (g < r - q).$$

□

We have some tricks for the others. Note that if we take $\phi(x) = x^2$ then we get that f^2 is measurable and hence $2fg = (f+g)^2 - f^2 - g^2$. Also $|f| = \max(f, -f)$. Note that

$$(\max(f,g) < r) = (f < r) \cap (g < r).$$

When we studied the Riemann integral the integrability of sums was not quite trivial because we had to use common refinement but it looks easier than the proof here using countable unions. I will comment on this when we have discussed simple functions.

Analytic combinations work for measurable functions whereas they do not work for continuous ones.

Corollary (**Analysis of measurable functions**). *Suppose* $(\Omega, \mathcal{F})$ *is a measurable space, and for each* n*, the function* $f_n : \Omega \to \mathbf{R}$ *is measurable. Then so are* $\sup_n f_n$*,* $\inf_n f_n$*,* $\limsup_n f_n$*,* $\liminf_n f_n$ *and* $\lim_n f_n$ *if it exists.*

Proof. We shall do the sup. Observe that $\sup f_n(x) > r$ if and only if one of the f_n is bigger than r. Hence

$$\{x : \sup f_n(x) > r\} = \bigcup_k \{x : f_k(x) > r\}$$

and this is a countable union of measurable sets. □

We used

$$\{x : \sup f_n(x) > r\} = \bigcup_k \{x : f_k(x) > r\}$$

Warning If you tried to do this with $\{x : \sup f_n(x) \geq r\}$ you would get into trouble because the sup could be equal to r even if all the functions are $< r$. **Choose your generating family carefully.**

Convergence of measurable functions

Suppose (f_n) is a sequence of measurable functions $f_n : \Omega \to \mathbf{R}$ and f another such function. The set

$$\{x : f_n(x) \to f(x)\}$$

of points where the f_n converge to f is measurable.

To see this observe that $f_n(x) \to f(x)$ if and only if for every k there is an m with $|f_n(x) - f(x)| < \frac{1}{k}$ for all $n \geq m$. So the set is

$$\bigcap_k \bigcup_m \bigcap_{n \geq m} \left\{ x : |f_n(x) - f(x)| < \frac{1}{k} \right\}.$$

Here you see we really use the σ-algebra properties.

In the same way the set where the (f_n) converge (to something) is also measurable: **HW**.

We often have functions converging at all but a handful of points. Since we are interested in the integral we normally don't care about a few points.

Definition (**Almost everywhere convergence**)**.** *Let (f_n) be a sequence of measurable functions defined on $(\Omega, \mathcal{F}, \mu)$. We say that the sequence converges almost everywhere to f and write $f_n \to f$* **a.e.** *if*

$$\mu(\{\omega : f_n(\omega) \text{ does not converge to } f(\omega)\}) = 0.$$

Example. *The functions $x \mapsto x^n$ converge to 0 on $[0, 1)$ but not at 1 itself. So they converge a.e. to 0 on $[0, 1]$.*

There is a second type of convergence that is extremely useful.

Definition (**Convergence in measure**)**.** *Let (f_n) be a sequence of measurable functions defined on $(\Omega, \mathcal{F}, \mu)$. We say that the sequence* **converges in measure** *to f if for every $\varepsilon > 0$*

$$\mu(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \to 0, \quad \text{as } n \to \infty.$$

Convergence in measure is weaker than convergence a.e. as long as the measure space is finite.

Lemma (**Convergence a.e. implies convergence in measure, if...**)**.** *Let (f_n) be a sequence of measurable functions defined on $(\Omega, \mathcal{F}, \mu)$ with $\mu(\Omega) < \infty$. If $f_n \to f$ a.e. then $f_n \to f$ in measure.*

Proof. For each ω where $f_n(\omega)$ does **not** converge to $f(\omega)$ we know that there is an $\varepsilon > 0$ so that for every n there is some $m > n$ for which $|f_m(\omega) - f(\omega)| > \varepsilon$. So the set where the f_n do **not** converge to f is

$$\bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} (|f_m - f| > 1/k).$$

This set has measure zero if $f_n \to f$ a.e. and hence so does each set

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} (|f_m - f| > \varepsilon).$$

Since $\mu(\Omega) < \infty$ that means that

$$\mu\left(\bigcup_{m=n}^{\infty} (|f_m - f| > \varepsilon)\right) \to 0$$

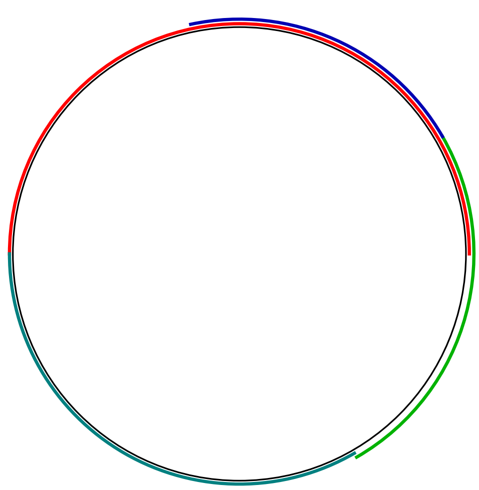
as $n \to \infty$. This is obviously stronger than

$$\mu\left(|f_n - f| > \varepsilon\right) \to 0$$

as $n \to \infty$. □

The hypothesis that the measure space is finite is clearly needed. For example we choose $f_n : \mathbf{R} \to \mathbf{R}$ to be the indicator function of $[n, n+1]$. $f_n \to 0$ a.e. (actually everywhere) but $\lambda(|f_n - 0| > 1/2) = 1$ for all n.

It is also fairly easy to see that convergence in measure does **not** imply convergence a.e. even if the measure is finite. Take intervals $[0, 1/2)$, $[1/2, 5/6)$, $[5/6, 13/12)$ of lengths $1/2$, $1/3$ and so on and "wrap them around" the circle.



Clearly the indicators of these sets converge to 0 in measure because their measures converge to 0. But because $\sum \frac{1}{n} = \infty$ the sequence wraps around the circle infinitely often. So the indicators do not converge to 0 anywhere. Notice that a subsequence of the indicators **does** converge to 0 a.e. Choose only those intervals that contain a particular point on the circle.

In the **HW** I will ask you to prove the first Borel-Cantelli Lemma.

Lemma (**Borel-Cantelli I**)**.** *Let* (E_k) *be a family of measurable sets in* $(\Omega, \mathcal{F}, \mu)$ *with*

$$\sum \mu(E_k) < \infty.$$

Then

$$\mu(\{x : x \text{ belongs to infinitely many of the } E_k\}) = \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 0.$$

Using this we can obtain a partial converse to the implication convergence a.e. implies convergence in measure **if the measure space is finite**.

Lemma (**Convergence in measure implies a.e. convergence of a subsequence**)**.** *Let* (f_n) *be a sequence of measurable functions defined on* $(\Omega, \mathcal{F}, \mu)$ *with* $\mu(\Omega) < \infty$*. If* $f_n \to f$ *in measure then there is a subsequence* (f_{n_k}) *with* $f_{n_k} \to f$ *a.e.*

Proof. For each $k \in \mathbf{N}$ the

$$\mu(|f_n - f| > 1/k) \to 0$$

so for some n_k we have

$$\mu(|f_{n_k} - f| > 1/k) < 2^{-k}.$$

Since

$$\sum_{k=1}^{\infty} \mu(|f_{n_k} - f| > 1/k) < \infty$$

the first Borel-Cantelli Lemma guarantees that for almost every $x \in \Omega$ there are only finitely many k for which

$$|f_{n_k}(x) - f(x)| > 1/k.$$

So for a.e. x we have that from some point on

$$|f_{n_k}(x) - f(x)| \leq 1/k$$

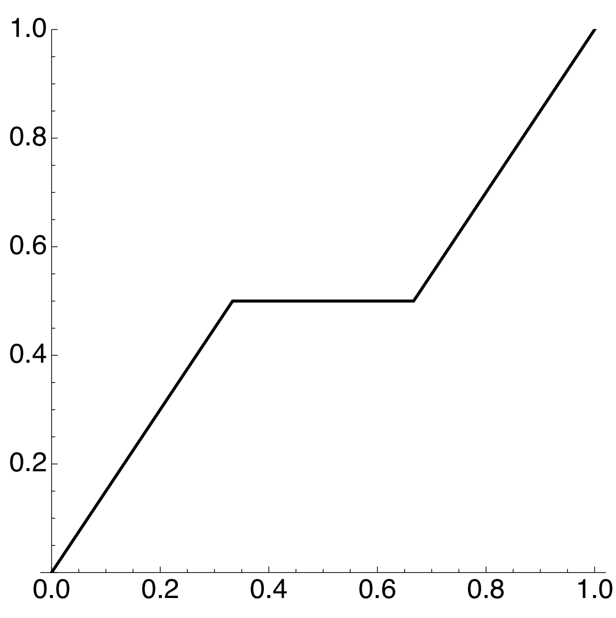
implying convergence (actually at a specified rate). □

There is a second Borel-Cantelli Lemma that concerns independent sets so it is mainly of relevance in probability theory.

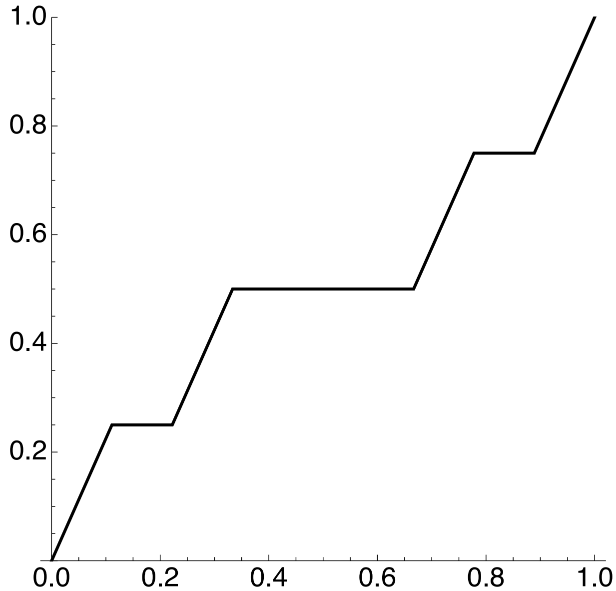
The Devil's staircase

Recall the construction of Cantor's middle thirds set.

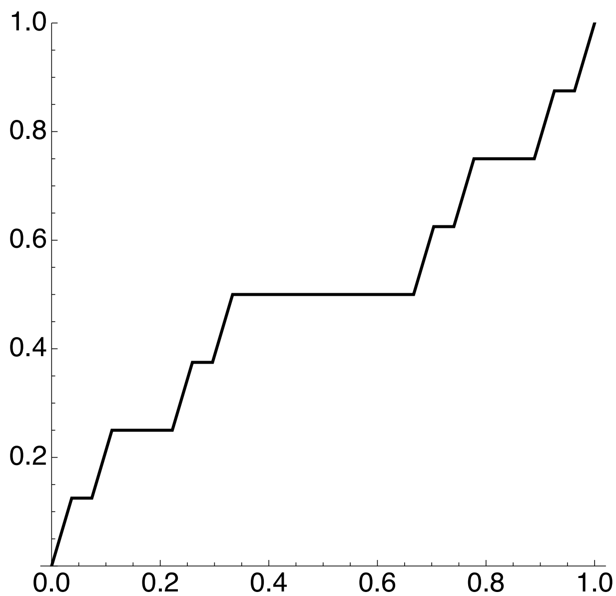
We shall build a function in a similar way. $c_0(x) = x$. On the middle third we set $c_1(x) = 1/2$ and join linearly elsewhere.



On the middle of the first third we set $c_2(x)$ it to be $1/4$ and on the middle of the last third to be 3/4 and so on.



For $c_3(x)$ we put in horizontal pieces at the next level.

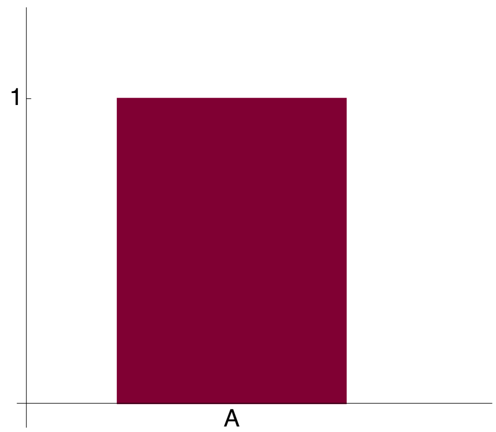


HW It is easy to check that these functions converge uniformly so their limit c is continuous. On each middle third, at every stage, the function is horizontal. So, almost everywhere, the function is differentiable with derivative 0. But the function rises from 0 to 1 so it certainly isn't the integral of its derivative.

Using the Devil's staircase we can check that there are sets of measure 0 that are not Borel sets: Lebesgue but not Borel.

Chapter 4. The integral

We already remarked that if we have a measure space $(\Omega, \mathcal{F}, \mu)$ and $A \in \mathcal{F}$ then the indicator function $\mathbf{1}_A$ should have integral $\mu(A)$.



The definition

We will build the integral in 4 steps.

1. If $E \in \mathcal{F}$ then we set
$$\int_\Omega \mathbf{1}_E \, d\mu = \mu(E).$$

2. A function of the form $\sum c_j \mathbf{1}_{E_j}$ for $c_j \in \mathbf{R}$ will be called a **simple function**.

   As long as all $c_j \geq 0$ we set
$$\int_\Omega \sum c_j \mathbf{1}_{E_j} \, d\mu = \sum c_j \mu(E_j).$$

   The sets are allowed to have infinite measure.

3. If f is a non-negative measurable function we put
$$\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega g \, d\mu : 0 \leq g \leq f, \ g \text{ simple} \right\}.$$

   We approximate f from below by simple functions.

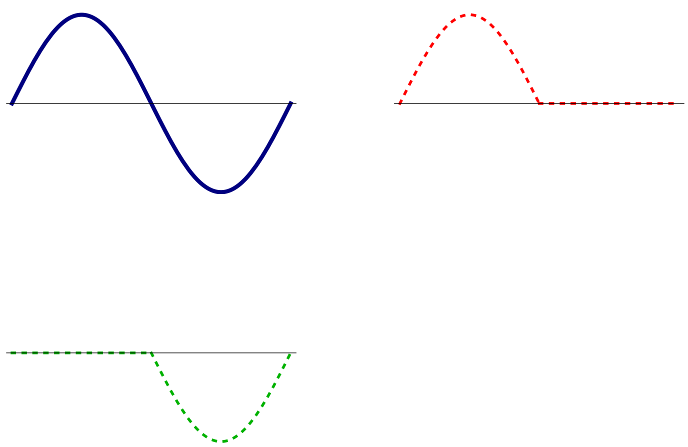
   For a non-negative f we don't mind an integral of ∞.

4. If f is measurable but not necessarily non-negative we can write it as

$$f = f_+ - f_-$$

where $f_+(x) = \max(f(x), 0)$ and $f_-(x) = -\min(f(x), 0)$.

f_+ is the positive part of f and f_- is the negative part.



Now, if $\int f_+ \, d\mu$ and $\int f_- \, d\mu$ are both finite then we say f is **integrable** and we set

$$\int_\Omega f = \int_\Omega f_+ - \int_\Omega f_-.$$

Note that if a function is integrable then $\int |f| < \infty$.

We need to check that our integral has certain properties such as linearity and monotonicity: for example if f and g are integrable

$$\int (f+g) = \int f + \int g.$$

We also want to check the convergence theorems such as

Theorem (**Monotone convergence**)**.** *Suppose $f_n : \Omega \to \mathbf{R}$ is an increasing sequence of non-negative measurable functions. Set*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for each x. Then f is measurable and

$$\int f_n \to \int f.$$

Exercise One thing we can check already is that if $f : \Omega \to \mathbf{R}$ is non-negative and $\int f = 0$ then $f = 0$ a.e.

Small remark Recall Step 2. If $f = \sum c_j \mathbf{1}_{E_j}$ is a non-negative simple function then

$$\int_\Omega \sum c_j \mathbf{1}_{E_j}\, d\mu = \sum c_j \mu(E_j).$$

A simple function can have many different representations. If $F \cap G = \emptyset$ then

$$\mathbf{1}_{F\cup G} = \mathbf{1}_F + \mathbf{1}_G.$$

Strictly speaking we need to check that all these presentations give the same integral: **Exercise**

Lemma (**Basic properties for simple functions**)**.** *If f and g are non-negative a simple functions then so is $f+g$ and if $\lambda \geq 0$ then*

1. $\int \lambda f = \lambda \int f$
2. $\int (f+g) = \int f + \int g$
3. *If $f \geq g$ then $\int f \geq \int g$.*

Proof. The first is obvious since if $f = \sum c_j \mathbf{1}_{F_j}$ then $\lambda f = \sum \lambda c_j \mathbf{1}_{F_j}$. For the second, suppose that $f = \sum c_j \mathbf{1}_{F_j}$ and $g = \sum d_k \mathbf{1}_{G_k}$. By considering all possible intersections $F_j \cap G_k$ we can express both functions using a single decomposition into multiples of $\mathbf{1}_{H_r}$.

$$f = \sum u_r \mathbf{1}_{H_r} \quad \text{and} \quad g = \sum v_r \mathbf{1}_{H_r}.$$

Then

$$f + g = \sum (u_r + v_r) \mathbf{1}_{H_r}$$

and so

$$\int (f+g) = \sum (u_r + v_r)\mu(H_r) = \int f + \int g.$$

The same argument works for the third statement. □

Note that this double decomposition looks like the common refinement that we used to prove additivity for the Riemann integral.

We now move on to more general measurable functions.

Lemma (**Basic properties for non-negative functions**). *If f and g are non-negative measurable functions and if $\lambda \geq 0$ then*

1. $\int \lambda f = \lambda \int f$
2. $\int (f+g) \geq \int f + \int g$
3. *If $f \geq g$ then $\int f \geq \int g$.*

Note that we only prove super-additivity at this stage.

Proof. The first is obvious again since if g is a simple function with $0 \leq g \leq f$ then λg is simple and satisfies $0 \leq \lambda g \leq \lambda f$. Similarly the third.

For the second suppose that u and v are simple functions with $0 \leq u \leq f$ and $0 \leq v \leq g$. Then $u + v \leq f + g$ and hence

$$\int (f+g) \geq \int (u+v) = \int u + \int v.$$

Now if we take the sup over all choices of u and v we get

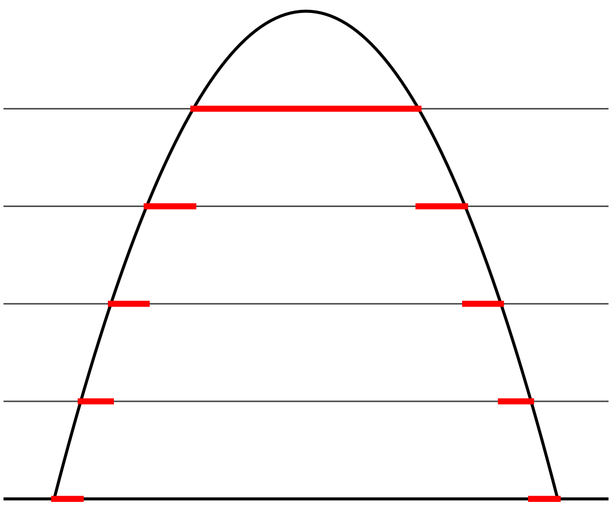
$$\int (f+g) \geq \int f + \int g.$$

□

In order to get $\int (f+g) = \int f + \int g$ and to do much more we need to get our hands on $\int f$ much more directly using simple functions. We shall observe that if f is a non-negative measurable function then there is an increasing sequence of simple functions (f_n) with $f_n \to f$ for all $x \in \Omega$. Once we have proved the Monotone Convergence Theorem stated above, we will then get $\int f_n \to \int f$ and similarly for g and $f + g$.

Approximation by simple functions

For each n we build a grille with spacing 2^{-n} and height 2^n. Now, given a non-negative measurable function f we "drop it on the grille".



For each n define

$$f_n(x) = \begin{cases} m2^{-n} & \text{if } m2^{-n} \leq f(x) < (m+1)2^{-n} \text{ and } 0 \leq m \leq 2^{2n} - 1 \\ \\ 2^n & \text{if } f(x) \geq 2^n. \end{cases}$$

This is a simple function because the set

$$\{x : m2^{-n} \leq f(x) < (m+1)2^{-n}\}$$

is measurable. It is easy to check that $f_n(x) \uparrow f(x)$ for each x. **Exercise** "If you drop onto a finer grille then the function doesn't fall as far."

The Monotone Convergence Theorem

In this section we prove the Monotone Convergence Theorem. The continuity of measure will be crucial.

Lemma (**The continuity of measure**)**.** *Suppose*

$$E_1 \subset E_2 \subset E_3 \subset \cdots$$

is an increasing sequence of measurable sets in $(\Omega, \mathcal{F}, \mu)$ *with* $E = \bigcup_1^\infty E_n$. *Then*

$$\mu(E) = \lim_n \mu(E_n).$$

Theorem (**Monotone convergence**)**.** *Suppose* $f_n : \Omega \to \mathbf{R}$ *is an increasing sequence of non-negative measurable functions. Set*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for each x. *Then* f *is measurable and*

$$\int f_n \to \int f.$$

We use the "standard machine": indicator, simple, general

Proof. Note that since $f_n \leq f$ we have $\int f_n \leq \int f$ for all n. So it suffices to check that

$$\lim \int f_n \geq \int f.$$

Indicator Suppose $f = \mathbf{1}_F$ for a measurable F. Given $0 < \varepsilon < 1$ let

$$E_n = \{x \in \Omega : f_n(x) > 1 - \varepsilon\} \subset F.$$

Since $f_n(x) \uparrow 1$ for each $x \in F$ we have

$$E_1 \subset E_2 \subset E_3 \cdots$$

and $F = \bigcup E_n$. Therefore $\mu(E_n) \to \mu(F)$ by continuity.

Now $f_n \geq (1-\varepsilon)\mathbf{1}_{E_n}$ and so

$$\int f_n \geq (1-\varepsilon) \int \mathbf{1}_{E_n} = (1-\varepsilon)\mu(E_n) \to (1-\varepsilon)\mu(F).$$

So

$$\lim \int f_n \geq (1-\varepsilon) \int f$$

and since ε was arbitrary we have

$$\lim \int f_n \geq \int f.$$

Simple Suppose $f = \sum c_j \mathbf{1}_{F_j}$ is a non-negative simple function. We can assume that the F_j are disjoint. For each n $f_n = \sum f_n \mathbf{1}_{F_j}$. Now for each j

$$\int f_n \mathbf{1}_{F_j} \to \int c_j \mathbf{1}_{F_j} = c_j \mu(F_j)$$

by the result for indicators.

By the superadditivity of the integral

$$\int f_n \geq \sum \int f_n \mathbf{1}_{F_j} \to \sum c_j \mu(F_j) = \int f.$$

Therefore

$$\lim \int f_n \geq \int f.$$

General Now suppose that f is a general non-negative measurable function and g is a simple function with $0 \leq g \leq f$. The functions $\min(f_n, g)$ increase to g. By the previous step

$$\int \min(f_n, g) \to \int g.$$

Since $\int f_n \geq \int \min(f_n, g)$ for each n we have

$$\lim \int f_n \geq \int g$$

and since g was arbitrary we get

$$\lim \int f_n \geq \int f.$$

□

Corollary (**Additivity**)**.** *Suppose $f, g : \Omega \to \mathbf{R}$ are non-negative measurable functions. Then*

$$\int (f+g) = \int f + \int g.$$

Proof. Choose increasing sequences of simple functions $f_n \uparrow f$ and $g_n \uparrow g$. Then $f_n + g_n \uparrow f + g$ and so

$$\int (f+g) = \lim \int (f_n + g_n) = \lim \left(\int f_n + \int g_n \right) = \int f + \int g.$$

□

Exercise Check that additivity works for integrable functions $f = f_+ - f_-$ and $g = g_+ - g_-$. Notice that $|f| = f_+ + f_-$ and so

$$\int |f| = \int f_+ + \int f_-.$$

Slight warning The way MON was stated the limit f could take the value ∞. This does not cause problems but strictly speaking if we want to allow this possibility we should extend the Borel σ-algebra to one on $(-\infty, \infty]$. If a function takes the value ∞ on a set of measure 0 we effectively use the convention $0.\infty = 0$.

The most important measure spaces are $\mathbf{R}$, $\mathbf{R}^n$, $[0,1]$, $[0,\infty)$ and so on equipped with Lebesgue measure. But we have also seen another: the set of natural numbers $\mathbf{N}$, the σ-algebra consisting of all subsets and counting measure:

$$\mu(\{1,3,4\}) = 3.$$

A function $f : \mathbf{N} \to \mathbf{R}$ is just a sequence of real numbers

$$(f(1), f(2), \ldots).$$

An indicator function $\mathbf{1}_A$ for $A \subset \mathbf{N}$ is the sequence x_i with $x_i = 1$ if $i \in A$ and 0 if not.

For example if $A = \{1,3,4\}$ then

$$\mathbf{1}_A = (1,0,1,1,0,0,\ldots).$$

For any A

$$\int \mathbf{1}_A \, d\mu = \mu(A) = \sum_{n=1}^{\infty} \mathbf{1}_A(n).$$

So for a simple function $f = \sum_1^m c_j \mathbf{1}_{A_j}$ we have

$$\begin{aligned} \int f &= \sum_1^m c_j \mu(A_j) = \sum_1^m c_j \sum_{n=1}^{\infty} \mathbf{1}_{A_j}(n) \\ &= \sum_{n=1}^{\infty} \left(\sum_{j=1}^m c_j 1_{A_j}(n) \right) = \sum_1^{\infty} f(n). \end{aligned}$$

The same is true for other non-negative f. (**Exercise**)

For a function which takes negative values we only call it integrable if $\int |f| < \infty$ and thus if

$$\sum_1^\infty |f(n)| < \infty.$$

So only the absolutely convergent series are integrable. The Lebesgue integral is an absolute integral. The function $x \mapsto \sin x/x$ is improperly Riemann integrable

$$\int_0^\infty \frac{\sin x}{x}\, dx = \frac{\pi}{2}$$

but it is not Lebesgue integrable.

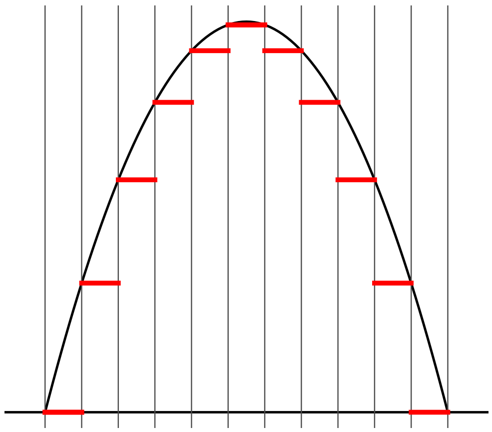
Agreement with the Riemann integral

It is possible to show that every Riemann integrable function is Lebesgue measurable. However it is not true that Riemann integrable functions are necessarily Borel measurable. The advantage of the Lebesgue integral is not really that it integrates more functions although it has to do so in order for the convergence theorems to make sense.

The main advantages are the convergence theorems and the fact that we can consider all sorts of measure spaces for which Riemann integration makes no sense. This is especially true in probability theory.

What matters relative to the Riemann integral is that if a measurable function **is** Riemann integrable then the integrals are the same which means that the Fundamental Theorem of Calculus holds equally well for continuous integrands. This is easy to see. If f is Riemann integrable then for each $\varepsilon > 0$ we can choose a partition for which the lower and upper sums differ by less than ε from the integral. These sums are the Lebesgue integrals of simple functions based on the intervals of the partition, lying above and below the function. (It doesn't matter what you do at the end-points of the subintervals.) So the Lebesgue integral is sandwiched between the upper and lower sums. So it within ε of the Riemann integral.

This discussion draws attention to the fact that the Riemann integral does not involve approximating functions: we approximate the integral directly without actually approximating the function. In the Lebesgue theory we really approximate the function.



The second form of the Fundamental Theorem of Calculus can be proved in greater generality with the Lebesgue integral. If F is differentiable everywhere then its derivative is automatically Borel measurable (**HW**) but it might not be Lebesgue integrable. If the derivative **is** Lebesgue integrable then its integral is F. As the Devil's staircase shows, you can't really relax the hypothesis that the derivative exists everywhere.

The first form of the Fundamental Theorem of Calculus has a subtle and important extension to the Lebesgue Theory.

Theorem (Lebesgue's differentiation theorem). *If $f : [0, 1] \to \mathbf{R}$ is integrable then the function*

$$F : x \mapsto \int_0^x f(t)\, d\lambda(t)$$

is differentiable almost everywhere and $F' = f$ almost everywhere.

An example of convergence

Example.

$$\int_0^\infty \frac{x}{e^x - 1}\, dx = \sum_1^\infty \frac{1}{n^2}.$$

The function $x/(e^x - 1)$ is non-negative for $x > 0$. It is not defined at 0 but this doesn't matter: set it to be whatever you like. Actually the function approaches 1 as $x \to 0$ so it would be most natural to choose this value.

For each $x > 0$

$$\frac{1}{1-e^{-x}} = \sum_0^\infty e^{-kx}$$

and hence

$$\frac{x}{e^x - 1} = \sum_0^\infty x e^{-(k+1)x}.$$

The partial sums

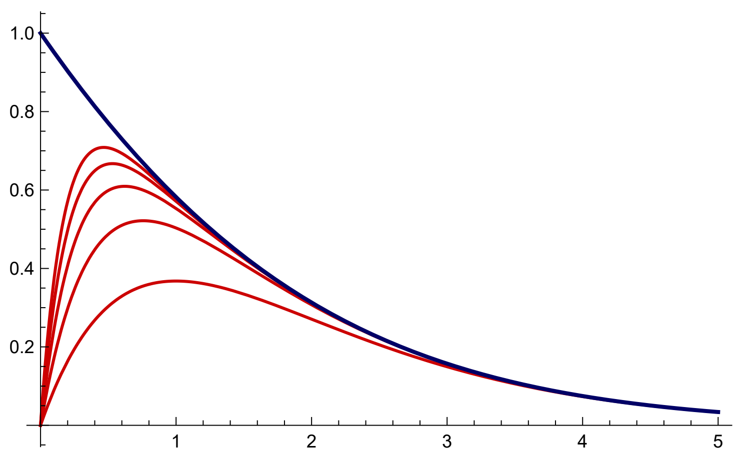
$$\sum_1^m x e^{-(k+1)x}$$

increase to $x/(e^x - 1)$ on $(0, \infty)$ so

$$\int_0^\infty \frac{x}{e^x - 1} = \lim_{m \to \infty} \sum_0^m \int_0^\infty x e^{-(k+1)x}\, dx.$$

Note that the convergence is **not** uniform because the partial sums are continuous but at 0 the sum converges to 0 which is not the limit of $x/(e^x - 1)$ at 0.

It is quite easy to see that the convergence is "dominated".



The convergence is not uniform but it is "dominated" because $x/(e^x - 1)$ is (pretty obviously) integrable.

Exercise Show that $\frac{x}{e^x-1} \le e^{-x/2}$ if $x > 0$.

Returning to the example, for each k the function $xe^{-(k+1)x}$ is the monotone limit of the cut off functions

$$x \mapsto xe^{-(k+1)x}\mathbf{1}_{[0,N]}(x).$$

By the FTC

$$\begin{aligned}\int_0^N xe^{-(k+1)x}\,dx &= \frac{1}{(k+1)}Ne^{-(k+1)N} - \frac{1}{(k+1)^2}e^{-(k+1)N} + \frac{1}{(k+1)^2} \\ &\to \frac{1}{(k+1)^2}\end{aligned}$$

as $N \to \infty$. So

$$\int_0^\infty xe^{-(k+1)x}\,dx$$

exists and equals $1/(k+1)^2$.

Therefore

$$\int_0^\infty \frac{x}{e^x - 1} = \lim_{m\to\infty} \sum_0^m \frac{1}{(k+1)^2} = \sum_1^\infty \frac{1}{n^2}.$$

The Dominated Convergence Theorem

We now come to the original goal: Lebesgue's Dominated Convergence Theorem (DOM). The first step is Fatou's Lemma (FAT).

Theorem (Fatou)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_n : \Omega \to \mathbf{R}$ be a sequence of non-negative measurable functions. Then*

$$\int (\liminf f_n)\,d\mu \leq \liminf \int f_n\,d\mu.$$

Reminder If (x_n) is a sequence of reals then the bottoms of tails are $b_m = \inf_{n\geq m} x_n$. These increase and we set

$$\liminf x_n = \lim b_m.$$

Proof. For each m set $g_m = \inf_{n \geq m} f_n$ which is measurable. The g_m increase to $\liminf f_n$ so by Monotone Convergence

$$\int (\liminf f_n) = \lim \int g_m.$$

For each m we have $g_m \leq f_m$ and so $\int g_m \leq \int f_m$. Hence

$$\int (\liminf f_n) = \lim \int g_m = \liminf \int g_m \leq \liminf \int f_m.$$

□

Example. *Suppose $f_n = \mathbf{1}_{[n,n+1]}$ on $\mathbf{R}$. Then for each n we have $\int f_n = 1$ but $\liminf f_n = \lim f_n = 0$ at every point.*

So we need $\leq$ in Fatou not $=$. Note that these functions are not dominated by an integrable function because $\int \mathbf{1}_{[0,\infty)} = \infty$. However Fatou is half way to Dominated Convergence.

Theorem (Dominated convergence). *Let $f_n : [0,1] \to \mathbf{R}$ be a sequence of integrable functions which converge at each point and set*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for each x. Suppose that there is an integrable function g with $|f_n(x)| \leq g(x)$ for all n and x.

Then f is integrable and

$$\int f_n \to \int f.$$

Proof. For each n we have $f_n + g = f_n - (-g) \geq 0$. By FAT

$$\int \liminf(f_n + g) \leq \liminf \int (f_n + g) = \liminf \left(\int f_n + \int g \right).$$

$$\int \liminf(f_n + g) \leq \liminf \int (f_n + g) = \liminf \left(\int f_n + \int g \right).$$

But $\lim f_n = f$ so the left side is $\int (f + g) = \int f + \int g$. Therefore

$$\int f + \int g \leq \liminf \left(\int f_n + \int g \right) = \liminf \int f_n + \int g.$$

Consequently

$$\int f \leq \liminf \int f_n.$$

On the other hand applying Fatou to the sequence $g - f_n$ we get

$$\int f \geq \limsup \int f_n.$$

So together we get

$$\int f \leq \liminf \int f_n \leq \limsup \int f_n \leq \int f$$

and this means that $\lim \int f_n$ exists and is equal to $\int f$. □

The next section contains a striking application.

Differentiation under the integral

In this section we prove something very useful that becomes surprisingly easy with Dominated Convergence.

Example. *For $u \in (0, \infty)$ we have*

$$\int_0^1 u^x \, dx = \begin{cases} \frac{u-1}{\log u} & \text{if } u \neq 1 \\ 1 & \text{if } u = 1. \end{cases}$$

For $u \neq 1$ we can compute the derivative by the machine. What about when $u = 1$?

You could use l'Hôpital. But we expect

$$\frac{d}{du} \int_0^1 u^x \, dx = \int_0^1 x u^{x-1} \, dx$$

because the integral is a kind of sum and the derivative of the sum is the sum of the derivatives. If $u = 1$ we get $\int_0^1 x \, dx = 1/2$.

Just as we had to be careful when differentiating power series because we were interchanging the order of limits so we need an argument when differentiating integrals.

Recall from the first year:

Theorem (**Continuous and sequential limits**)**.** *If $f : I\backslash\{c\} \to \mathbf{R}$ is defined on the interval I except at $c \in I$ then*

$$\lim_{x\to c} f(x) = L$$

if and only if for every sequence (x_n) in $I\backslash\{c\}$ with $x_n \to c$ we have

$$f(x_n) \to L.$$

So for a **fixed** function we can compute its limit by looking at countable sequences which is what measure theory deals with.

Theorem (**Differentiation under the integral**)**.** *Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and $I \subset \mathbf{R}$ is an open interval. Let $f : I \times \Omega \to \mathbf{R}$ be a function with the property that $f(t,\omega)$ is differentiable wrt t for all ω and integrable wrt ω for every t. Suppose that there is an integrable function g on Ω so that*

$$\left|\frac{\partial f(t,\omega)}{\partial t}\right| \le g(\omega)$$

for all $t \in I$. Then the function $\omega \mapsto \frac{\partial f(t,\omega)}{\partial t}$ is integrable, the function

$$F : t \mapsto \int_\Omega f(t,\omega)\, d\mu(\omega)$$

is differentiable and

$$\frac{\partial}{\partial t}\int_\Omega f(t,\omega)\, d\mu(\omega) = \int_\Omega \frac{\partial f(t,\omega)}{\partial t}\, d\mu(\omega).$$

Proof. We know that

$$\frac{\partial f(t,\omega)}{\partial t} = \lim_{n\to\infty} \frac{f(t+1/n,\omega) - f(t,\omega)}{1/n}$$

so it is a limit of measurable functions. Since it is dominated by g it is integrable.

We want to prove that if δ_n is any sequence converging to zero then

$$\frac{F(t+\delta_n) - F(t)}{\delta_n} \to \int_\Omega \frac{\partial f(t,\omega)}{\partial t}\, d\mu(\omega).$$

The expression on the left is

$$\int_\Omega \frac{f(t+\delta_n,\omega) - f(t,\omega)}{\delta_n}\, d\mu(\omega).$$

We know that

$$\frac{f(t+\delta_n,\omega)-f(t,\omega)}{\delta_n} \to \frac{\partial f(t,\omega)}{\partial t}$$

for each ω by the differentiability of f so we just need to check that the convergence is dominated. But by the MVT

$$\frac{f(t+\delta_n,\omega)-f(t,\omega)}{\delta_n} = \left.\frac{\partial f(u,\omega)}{\partial u}\right|_s$$

for some s between t and $t+\delta_n$ and the latter is at most $g(\omega)$. □

Example (doi.org/10.1112/mtk.12172)**.** *The function*

$$s \mapsto \int_{-\infty}^{\infty} \frac{(1/2+iy)^{1-s}}{\cosh^2 \pi y}\, dy$$

is an entire function.

Sketch We want to show that we can differentiate with respect to s. The derivative of the integrand is

$$-\log(1/2+iy)\frac{(1/2+iy)^{1-s}}{\cosh^2 \pi y}.$$

The factors $\log(1/2+iy)$ and $(1/2+iy)^{1-s}$ grow at most as fast as a power of y. So the derivative is dominated by an integrable function.

Chapter 5. Product measures and Fubini's Theorem

Rectangles

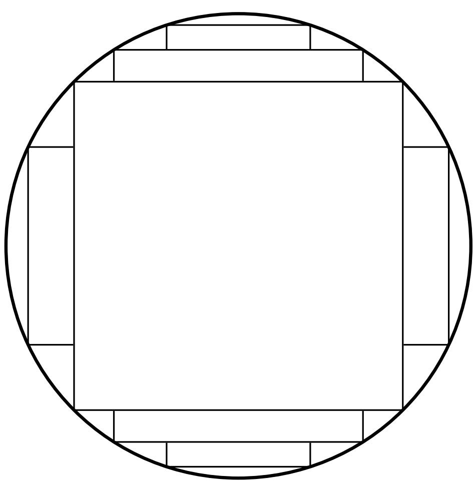
We constructed Lebesgue measure λ on the line. What about the plane or more generally $\mathbf{R}^n$? In this chapter we shall build a **product measure** $\mu \otimes \nu$ from two measure spaces $(\Omega, \mathcal{F}, \mu)$ and $(\Phi, \mathcal{G}, \nu)$.

Our underlying space will be the Cartesian product

$$\Omega \times \Phi = \{(x, y) : x \in \Omega, y \in \Phi\}.$$

We have to find the appropriate σ-algebra of subsets of $\Omega \times \Phi$ and then put the right measure on it.

If $F \subset \mathcal{F}$ and $G \subset \mathcal{G}$ we can form the Cartesian product $F \times G$ in $\Omega \times \Phi$ and we know its measure is supposed to be the product $\mu(F)\nu(G)$. We call the set $F \times G$ a **rectangle set**. Our σ-algebra should contain at least these sets and hence the σ-algebra they generate: σ(rectangles). For most purposes this σ-algebra is sufficient and our aim is to extend the measure $\mu \otimes \nu$ from the rectangles to the generated σ-algebra.



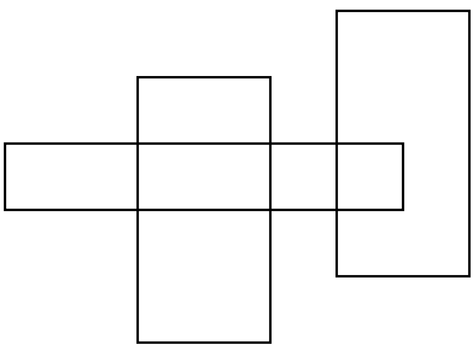
This σ-algebra will include the sort of sets you want.

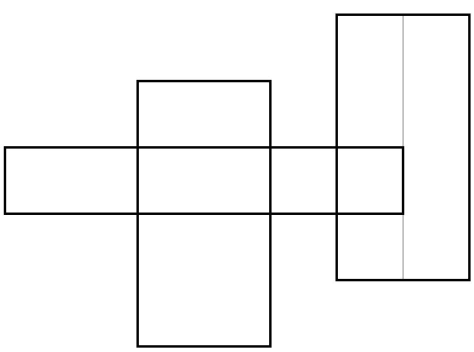
Let's check that it is really enough in the important case. Suppose we look at $\mathbf{R} \times \mathbf{R}$ and the Borel rectangles: $A \times B$ for $A, B \in \mathcal{B}$. Does the generated σ-algebra contain all the Borel subsets of the plane? It suffices to show that every open set in the plane is a countable union of open rectangles. Suppose $U \subset \mathbf{R}^2$ is open. For each point of U we can find a small open rectangle $(a, b) \times (c, d)$ containing that point, inside U, and for which $a, b, c, d \in \mathbf{Q}$. The union of all these rectangles is U and there are only countably many because the rationals are countable.

We can easily extend the measure to the family of **finite** disjoint unions of rectangles just by adding. It is not too hard to check that this family is an algebra.

Lemma (Finite unions of rectangles). *Suppose $(\Omega, \mathcal{F}, \mu)$ and $(\Phi, \mathcal{G}, \nu)$ are measure spaces. The family of finite disjoint unions of rectangles $F \times G$ for $F \in \mathcal{F}$ and $G \in \mathcal{G}$ forms an algebra on $\Omega \times \Phi$.*

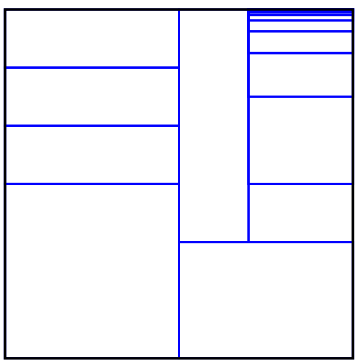
Proof. The empty set and $\Omega \times \Phi$ are rectangle sets. If $R_1, R_2, \ldots, R_n$ is any collection of rectangles (not necessarily disjoint) then there is a finite disjoint family $S_1, S_2, \ldots, S_m$ of rectangles so that each R_i is a union of some of the S_j.





If we are given a finite disjoint union of rectangles $A = \bigcup W_i$ then we can find a finite number of disjoint rectangles that decompose all the W_i and $\Omega \times \Phi$. Then $(\Omega \times \Phi)\backslash A$ is a union of all these rectangles with some omitted. So it is a finite disjoint union of rectangles. Similarly a finite union of finite unions of rectangles can be decomposed into a finite **disjoint** union. □

We shall see that if you start with a measure on an algebra then you can extend it to the generated σ-algebra. But we need the original to be a measure in the fullest possible sense. It is possible that we have a rectangle equal to a countable disjoint union of rectangles. If the measures didn't add up then we would have no hope of extending.



We want the measure to be countably additive even though we don't know that countable unions stay in the algebra. So we insist on countable additivity when they do.

Definition (Measure on an algebra)**.** *Let $\mathcal{A}$ be an algebra on a set Ω. A function $\mu : \mathcal{A} \to \mathbf{R} \cup \{\infty\}$ is called a* **measure** *if*

- $\mu(A) \geq 0$ *for all* $A \in \mathcal{A}$
- $\mu(\emptyset) = 0$
- *If* $A_1, A_2, A_3, \ldots \in \mathcal{A}$ *are disjoint and* $\bigcup A_i \in \mathcal{A}$ *then*

$$\mu\left(\bigcup A_i\right) = \sum \mu(A_i).$$

Caratheodory's extension theorem

Theorem (**Caratheodory's extension theorem**)**.** *Suppose μ is a measure on an algebra $\mathcal{A}$ of subsets of Ω. Then μ can be extended to a measure on $\sigma(\mathcal{A})$. Under appropriate conditions the extension is unique.*

Remark The conditions will be described later when we look at the uniqueness.

We start with a small lemma.

Lemma (**Sub-additivity on algebras**)**.** *If μ is a measure on an algebra $\mathcal{A}$ and we have $A_1, A_2, A_3 \ldots \in \mathcal{A}$, $A \in \mathcal{A}$ and $A \subset \bigcup A_i$ then*

$$\mu(A) \leq \sum \mu(A_i).$$

Proof. We disjointify: set $B_n = A_n \backslash \left(\bigcup_1^{n-1} A_i\right)$ for each n. The B_n belong to $\mathcal{A}$, their union is $\bigcup A_i$ and they are disjoint. **We have $A \subset \bigcup B_i$ but we don't know that the union is in $\mathcal{A}$.** But, $A \cap B_n \in \mathcal{A}$ for each n and $A = \bigcup(A \cap B_n)$. So

$$\mu(A) = \sum_1^\infty \mu(A \cap B_n) \leq \sum_1^\infty \mu(B_n) \leq \sum_1^\infty \mu(A_n).$$

□

For the proof of the Extension Theorem we use the Restriction Theorem. We build an outer measure μ^* on the subsets of Ω using the measure μ on $\mathcal{A}$. We check that the sets in $\mathcal{A}$ are μ^*-measurable. We check that if A is in the algebra then $\mu^*(A) = \mu(A)$. This parallels what we did when constructing Lebesgue measure: see later.

Proof. (**Of Caratheodory's extension**)
The outer measure For any set $E \subset \Omega$ set

$$\mu^*(E) = \inf\{\sum_1^\infty \mu(A_i) : E \subset \bigcup A_i \text{ and } A_i \in \mathcal{A}\}.$$

Clearly $\mu^*(\emptyset) = 0$ since $\emptyset \in \mathcal{A}$. If $E \subset F$ then any covering of F is a covering of E so μ^* is monotone.

Suppose $E \subset \bigcup E_i$. Given $\varepsilon > 0$, for each i, choose a family (A_{ij}) of sets in $\mathcal{A}$ so that $E_i \subset \bigcup_j A_{ij}$ and

$$\sum_j \mu(A_{ij}) \leq \mu^*(E_i) + \varepsilon/2^i.$$

Since $E \subset \bigcup_{ij} A_{ij}$ we have

$$\mu^*(E) \leq \sum_{ij} \mu(A_{ij}) \leq \sum_i \mu^*(E_i) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary we have

$$\mu^*(E) \leq \sum \mu^*(E_i)$$

as required.

Measurability Suppose $A \in \mathcal{A}$ and $T \subset \Omega$. We want that

$$\mu^*(T) \geq \mu^*(T \cap A) + \mu^*(T \backslash A).$$

Given $\varepsilon > 0$ choose A_i in $\mathcal{A}$ with $T \subset \bigcup A_i$ and

$$\sum \mu(A_i) \leq \mu^*(T) + \varepsilon.$$

Then $T \cap A \subset \bigcup (A_i \cap A)$, $T \backslash A \subset \bigcup (A_i \backslash A)$ and for each i

$$\mu(A_i) = \mu(A_i \cap A) + \mu(A_i \backslash A)$$

because μ is additive on $\mathcal{A}$. Therefore

$$\begin{aligned} \mu^*(T) &\geq \sum \mu(A_i) - \varepsilon = \sum \mu(A_i \cap A) + \sum \mu(A_i \backslash A) - \varepsilon \\ &\geq \mu^*(T \cap A) + \mu^*(T \backslash A) - \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we get

$$\mu^*(T) \geq \mu^*(T \cap A) + \mu^*(T \backslash A).$$

$\mu^*(A) = \mu(A)$ if $A \in \mathcal{A}$. This is where we really use the fact that μ is a measure. Plainly $\mu^*(A) \leq \mu(A)$ since A covers itself. If $A \subset \bigcup A_i$ then by the lemma $\mu(A) \leq$

$\sum \mu(A_i)$. So the infimum in the definition of $\mu^*(A)$ does not record anything less than $\mu(A)$. $\square$

To finish the construction we need to check that the product measure on the algebra really is a measure: that it is countably additive. The key step will be to show that if a rectangle is decomposed as a countable disjoint union of rectangles then the measures add up.

Theorem (**The product measure is a measure on the algebra**)**.** *Suppose* $(\Omega, \mathcal{F}, \mu)$ *and* $(\Phi, \mathcal{G}, \nu)$ *are measure spaces. The product measure* $\mu \otimes \nu$ **is** *a measure on the algebra of finite disjoint unions of rectangles.*

Proof. To begin with suppose that

$$E \times F = \bigcup_{1}^{\infty} E_i \times F_i$$

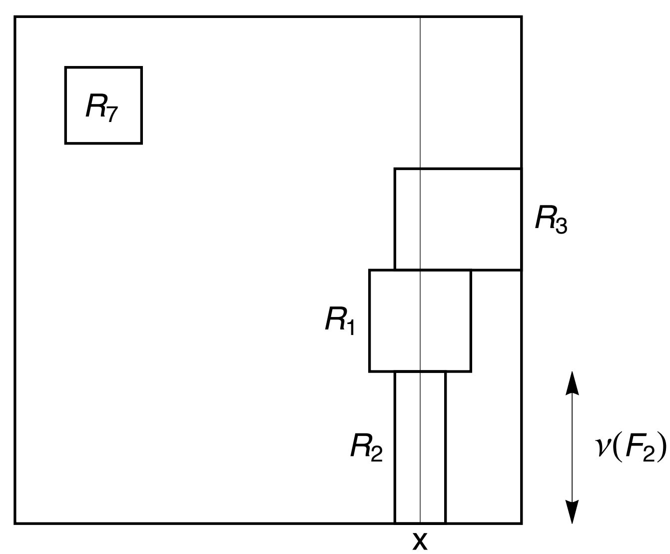
is a rectangle written as a countable disjoint union of rectangles. For each i define a function $f_i : \Omega \to \mathbf{R}$ by

$$f_i(x) = \nu(F_i)\mathbf{1}_{E_i}(x)$$

and similarly $f(x) = \nu(F)\mathbf{1}_E(x)$.

$f_i(x)$ tells us which R_i sit above x and how much they contribute to the slice above x.

$f_7(x) = 0$.



Observe that for each i the integral is

$$\int_\Omega f_i = \nu(F_i)\mu(E_i)$$

and similarly

$$\int_\Omega f = \nu(F)\mu(E).$$

So our aim is to show that

$$\int_\Omega f = \sum_i \int_\Omega f_i.$$

Now for each x there are some E_i that contain x and since the rectangles together cover $E \times F$ we have

$$F = \bigcup_{i:x\in E_i} F_i.$$

Therefore for each $x \in E$

$$f(x) = \nu(F) = \sum_{i:x\in E_i} \nu(F_i) = \sum_i f_i(x).$$

By MON we have

$$\int_\Omega f = \sum_i \int_\Omega f_i.$$

as required.

Now suppose that we have an element of the algebra, a finite disjoint union of rectangles $\bigcup R_i$ equal to a countable disjoint union of finite disjoint unions of rectangles

$$\bigcup_{k=1}^{\infty} \bigcup_{j=1}^{n_k} S_{kj}.$$

We can rewrite the latter as a countable disjoint union of rectangles

$$\bigcup_m S'_m$$

with a single indexing sequence and

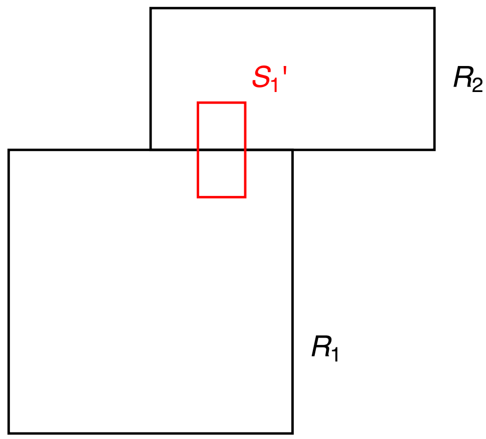
$$\sum_k \mu \otimes \nu \left(\bigcup_{j=1}^{n_k} S_{kj} \right) = \sum_{kj} \mu \otimes \nu(S_{kj}) = \sum_m \mu \otimes \nu(S'_m).$$

So we have

$$\bigcup_i R_i = \bigcup_m S'_m$$

and we want to check that

$$\sum_i \mu \otimes \nu(R_i) = \sum_m \mu \otimes \nu(S'_m).$$



Now for each i and m, the set $R_i \cap S'_m$ is a rectangle and these are all disjoint. So by the first part of the argument

$$\begin{aligned}\sum_m \mu \otimes \nu(S'_m) &= \sum_m \sum_i \mu \otimes \nu(S'_m \cap R_i) \\ &= \sum_i \sum_m \mu \otimes \nu(S'_m \cap R_i) = \sum_i \mu \otimes \nu(R_i).\end{aligned}$$

□

We have now constructed the product measure $\mu \otimes \nu$ on a product space $\Omega \times \Phi$. This will give us Lebesgue measure on (a bit more than) the Borel subsets of the plane $\mathbf{R} \times \mathbf{R}$. We can repeat the process and get Lebesgue measure on $\mathbf{R}^d$. This is a measure on the Borel sets which assigns measure

$$(b_1 - a_1)(b_2 - a_2) \ldots (b_d - a_d)$$

to each cuboid

$$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d).$$

Another way to do it is as follows.

Digression: a second construction of Lebesgue measure

Consider the family of finite disjoint unions of half-open intervals $[a, b)$, $[a, \infty)$, $(-\infty, b)$ and $(-\infty, \infty)$. As mentioned in the first chapter this is an algebra. We define Lebesgue measure on this algebra in the obvious way as the sum of the lengths. We then extend to the Borel σ-algebra using the restriction and extension theorems.

On the face of it this looks easier than what we did because "we don't need the crucial lemma". The catch is that we need to check that the "sum of the lengths" **is** a measure on the algebra. So we need to show that if

$$[a, b) = \bigcup [a_i, b_i)$$

disjointly then

$$b - a = \sum (b_i - a_i).$$

This looks intuitively obvious: but it is essentially the crucial lemma.

However one advantage of this over our construction is that this works equally well for $\mathbf{R}^d$. We use the algebra of disjoint unions of half-open blocks

$$[a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_d, b_d)$$

with λ(block) equal to the product of the lengths of its sides. For this measure, whichever way we create it, there is something we would like to check. Suppose M is a $d \times d$ matrix and we consider its action on a set $A \subset \mathbf{R}^d$. We want that $\lambda(M(A)) = |\det M|.\lambda(A)$.

It is easy to see that if M is a diagonal matrix we scale the sides of each box and so we multiply the measure by the product of the diagonal entries. We can write M as a product UDV where D is diagonal and U and V are orthogonal. So we need to check that an orthogonal matrix doesn't change the volume. This will be **HW**.

Uniqueness

Definition. *A measure space $(\Omega, \mathcal{F}, \mu)$ (or the measure μ) is called σ-**finite** if there are sets $A_1, A_2, \ldots \in \mathcal{F}$ of finite measure with*

$$\Omega = \bigcup A_i.$$

For example, Lebesgue measure on the line is σ-finite because $\mathbf{R}$ is the union of the finite intervals $[n, n+1)$ for $n \in \mathbf{Z}$. It turns out that the extension of a measure from an algebra to a σ-algebra is unique if the measure is σ-finite. This can be proved using something called the Monotone Class Theorem or, as we shall see, Dynkin's $\pi - d$ Lemma.

These two principles are proved using spooky magic. Hold tight!

Definition. *A family of subsets $\mathcal{P}$ of a set Ω is called a* π**-system** *if whenever A and B are in $\mathcal{P}$, so is $A \cap B$.*

Note that any algebra is a π-system.

Definition. *A family of subsets $\mathcal{D}$ of a set Ω is called a* d**-system** *if*

- $\emptyset, \Omega \in \mathcal{D}$
- *If $A, B \in \mathcal{D}$ and $B \subset A$ then $A \backslash B \in \mathcal{D}$*
- *If $A_1 \subset A_2 \subset \cdots$ are in $\mathcal{D}$ then so is their union $\bigcup A_i$*

Note that if a family is both a π-system and a d-system then it is a σ-algebra: since the system is closed under finite intersections and complements, it is closed under finite unions. Therefore an arbitrary countable union can be turned into an increasing union.

Note also that if we are given two measures on Ω with the same finite total mass, then the family of sets on which they agree is automatically a d-system. **Exercise**.

So if they agree on a family $\mathcal{P}$ then they automatically agree on the d-system generated by $\mathcal{P}$. The magical lemma will state that if $\mathcal{P}$ is a π-system then its generated d-system will be the same as its generated σ-algebra. In other words, if you start with a system closed under intersections and generate the d-system, then even though the sets of the d-system have almost nothing to do with those of the original π-system, the intersection property somehow passes to the much larger system.

Theorem (**Dynkin's $\pi - d$ Lemma**)**.** *Suppose $\mathcal{P}$ is a π-system of subsets of Ω. Then the d-dystem generated by $\mathcal{P}$ is also the σ-algebra generated by $\mathcal{P}$.*

We have repeatedly seen that if you want to check some property of a set system you don't try to write down typical sets and check that they have the property: you look at the things that have the property and check that they form a ?-system. Here we have to do it twice.

Proof. Clearly a σ-algebra is a d-system so $\sigma(\mathcal{P})$ includes the d-system generated by $\mathcal{P}$. What matters is the other direction.

It suffices to show that the d-system $d(\mathcal{P})$ generated by $\mathcal{P}$ is closed under intersections since as we saw, a d-system that is a π-system is a σ-algebra. Consider the family

$$\Sigma = \{E : E \cap F \in d(\mathcal{P}) \text{ for all } F \in \mathcal{P}\}.$$

Note that $\Sigma \supset \mathcal{P}$ because $\mathcal{P}$ is closed under intersections. I claim that Σ is a d-system ensuring that $\Sigma \supset d(\mathcal{P})$. Clearly $\emptyset$ and Ω are in Σ.

Suppose $A, B \in \Sigma$ and $F \in \mathcal{P}$. Then

$$(A \backslash B) \cap F = (A \cap F) \backslash (B \cap F) \in d(\mathcal{P})$$

and so $A \backslash B \in \Sigma$. Suppose $A_1 \subset A_2 \subset \cdots$ are in Σ and $F \in \mathcal{P}$. Then

$$\left(\bigcup A_i\right) \cap F = \bigcup (A_i \cap F)$$

and the union is increasing so it belongs to $d(\mathcal{P})$. Therefore $\bigcup A_i \in \Sigma$. So we now know that if $E \in d(\mathcal{P})$ and $F \in \mathcal{P}$ then $E \cap F \in d(\mathcal{P})$.

Now we consider

$$\Sigma' = \{E : E \cap F \in d(\mathcal{P}) \text{ for all } F \in d(\mathcal{P})\}.$$

We know by the first part that $\mathcal{P} \subset \Sigma'$. Using the same argument as before we get that $d(\mathcal{P}) \subset \Sigma'$. But that tells us that $d(\mathcal{P})$ is closed under finite intersections. □

We are now in a position to check the uniqueness of extensions.

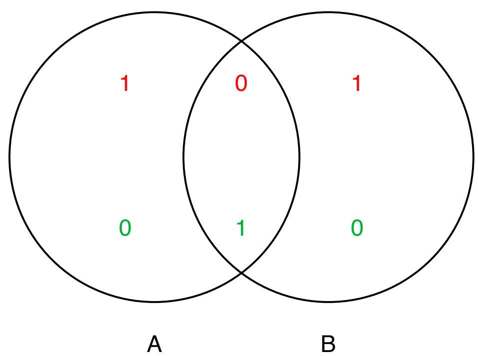
Theorem (**Caratheodory's extension theorem II**)**.** *Suppose μ is a measure on an algebra $\mathcal{A}$ of subsets of Ω. Then μ can be extended to a measure on $\sigma(\mathcal{A})$. If μ is σ-finite then the extension is unique.*

Proof. What remains is to prove uniqueness. We may assume that μ is finite since we can break the space up into countably many pieces on which this is true and then just add up.

Suppose we have two measures μ and ν that agree on the algebra. The family on which they agree is a d-system containing the algebra $\mathcal{A}$ which is a π-system. So they agree on the generated d-system which is $\sigma(\mathcal{A})$. □

I mentioned that the uniqueness can be done using the $\pi - d$ lemma or something called the Monotone Class Theorem. The first is a bit simpler than the second. But the real value of the first is that it can be used for proving independence of σ-algebras in probability theory: because independence tells you about intersections.

Remark If you were given two measures on a σ-algebra it would be extremely hard to check that the family where they agree is a σ-algebra or even an algebra. Suppose we have two sets A and B as shown with two measures, red and green, assigning the given masses to $A \backslash B$, $B \backslash A$ and $A \cap B$.



Then red and green agree on A and B but not on $A \cup B$ (or $A \cap B$).

Fubini's Theorem

In this section we are going to prove the Fubini Theorem for the Lebesgue integral. We have a pair of measure spaces $(\Omega, \mathcal{F}, \mu)$ and $(\Phi, \mathcal{G}, \nu)$ and the product measure $\mu \otimes \nu$ on the product space $\Omega \times \Phi$. We have an integrable function $f : \Omega \times \Phi$ and we want to know that we can compute the integral

$$\int_{\Omega \times \Phi} f(x, y)\, d\mu \otimes \nu$$

by integrating in the separate variables:

$$\int_\Omega \left(\int_\Phi f(x, y) d\nu(y) \right) d\mu(x)$$

or

$$\int_\Phi \left(\int_\Omega f(x, y) d\mu(x) \right) d\nu(y).$$

In other words that we can make sense of these two repeated integrals, that they give the same value and that this value is the integral with respect to the product measure. As usual we need some conditions on f: a bit like the MON and DOM conditions. The two versions of the theorem are often called Tonelli's Theorem and Fubini's Theorem.

Without such conditions Fubini does not hold. Suppose the two spaces are both $\mathbf{N}$ with counting measure. Then consider the function whose values on $\mathbf{N} \times \mathbf{N}$ are as shown.

$$
\begin{array}{cccccc|c}
1 & -1 & 0 & 0 & \ldots & & 0 \\
0 & 1 & -1 & 0 & \ldots & & 0 \\
0 & 0 & 1 & -1 & 0 & \cdots & 0 \\
\vdots & \vdots & & & & & \\
\vdots & \vdots & & & & & \\
\hline
1 & 0 & 0 & \cdots & & &
\end{array}
$$

If you integrate first along the rows you get 0 in each. But if you integrate first down the columns and then add the results you get 1.

In order to prove the theorem we are going to need that if we start with a product measurable f then the **sections**

$$x \mapsto f(x, y)$$

are measurable for each y (and similarly the other way around) and that the **marginals**

$$y \mapsto \int f(x, y) d\mu(x)$$

are measurable as a function of y (and the other way around).

The proofs of these will go via the standard machine in the first case and the $\pi - d$ lemma in the second

Lemma (**Measurability of sections**). *If $f : \Omega \times \Phi \to \mathbf{R}$ is measurable wrt $\sigma(\mathcal{F} \times \mathcal{G})$ then for each fixed y the map*

$$x \mapsto f(x, y)$$

is $\mathcal{F}$-measurable and similarly the other way around.

Proof. **We use the "standard machine": indicator, simple, general.**

Indicator Suppose that f is the indicator of a rectangle $\mathbf{1}_{F \times G}$. Then the section $x \mapsto \mathbf{1}_{F \times G}(x, y)$ is either $\mathbf{1}_F(x)$ if $y \in G$ or 0 if $y \notin G$.

Now we move on to the indicators of arbitrary sets A in $\sigma(\mathcal{F} \times \mathcal{G})$. As usual we can't write down such a set. So we ask whether the family of good sets is a σ-algebra. Note that for each y the function $x \mapsto \mathbf{1}_A(x, y)$ is the indicator of

$$A_y = \{x : (x, y) \in A\}.$$

Let $\mathcal{H}$ be the family of sets A for which A_y is $\mathcal{G}$-measurable for a fixed y (or for all y). Clearly $\Omega \backslash A_y = (\Omega \times \Phi \backslash A)_y$ so $\mathcal{H}$ is closed under complements. In a similar way it is closed under countable unions. So $\mathcal{H}$ is a σ-algebra that includes the rectangles so it includes $\sigma(\mathcal{F} \times \mathcal{G})$.

Simple If f is simple we just add the sections for the indicators it is built from to get the section of the function.

General For a general function f we express it as a limit of simple functions. Their sections are measurable and converge to the sections of f. □

We now want to look at the measurability of marginals such as

$$x \mapsto \int_\Phi f(x, y)\, d\nu(y).$$

This raises again the "infinite value" issue. Even if f is finite and integrable, it could be that the marginal is infinite for some values of x. As long as this only happens on a set of measure 0 we can end up with a finite integral over the product. By **HW 6 Q 1**, if it happens on a set of positive measure then the integral **is** infinite.

Lemma (**Measurability of marginals**). *If* $(\Omega, \mathcal{F}, \mu)$ *and* $(\Phi, \mathcal{G}, \nu)$ *are* σ*-finite measure spaces,* $f : \Omega \times \Phi \to \mathbf{R}$ *is measurable wrt* $\sigma(\mathcal{F} \times \mathcal{G})$ *and* $f \geq 0$ *then the map*

$$x \mapsto \int_\Phi f(x, y)\, d\nu(y)$$

is $\mathcal{F}$*-measurable and similarly the other way around.*

Proof. **This time we need the spooky magic.** Suppose first that μ and ν are finite measures.

Indicator If $A = F \times G$ is a rectangle then

$$\int_\Phi \mathbf{1}_A(x, y)\, d\nu(y) = \nu(G)\mathbf{1}_F(x)$$

which is measurable. Now let $\mathcal{H}$ be the family of sets $A \subset \Omega \times \Phi$ for which

$$x \mapsto \int_\Phi \mathbf{1}_A(x, y)\, d\nu(y)$$

is $\mathcal{F}$-measurable. $\mathcal{H}$ includes the rectangles which form a π-system so once we show that $\mathcal{H}$ is a d-system we are guaranteed that it includes all of $\sigma(\Omega \times \Phi)$.

The full space is in $\mathcal{H}$ because it is a rectangle. Suppose $A, B \in \mathcal{H}$ and $B \subset A$. Then

$$\int_\Phi \mathbf{1}_{A \setminus B}(x, y)\, d\nu(y) = \int_\Phi \mathbf{1}_A(x, y)\, d\nu(y) - \int_\Phi \mathbf{1}_B(x, y)\, d\nu(y)$$

and is measurable. (The integrability follows from the fact that both measures are finite.)

Now suppose that (A_i) is an increasing sequence of sets in $\mathcal{H}$ and A is their union. Then the sequence $\mathbf{1}_{A_i}$ increases to $\mathbf{1}_A$ and so by MON

$$\int_\Phi \mathbf{1}_{A_i}(x, y)\, d\nu(y) \to \int_\Phi \mathbf{1}_A(x, y)\, d\nu(y)$$

and the latter is a limit of measurable functions.

Simple If f is simple we just add the integrals for the indicators it is built from to get the integral of the function.

General For a general non-negative function write it as an increasing limit of simple functions and apply MON.

Finally if μ and ν are σ-finite we can write $\Omega = \bigcup \Omega_i$ as a union of sets of finite measure and similarly Φ. For each i the function

$$x \mapsto \int_\Phi f(x,y)\,d\nu(y)$$

is measurable on Ω_i because it is the sum of the positive measurable functions

$$x \mapsto \int_{\Phi_j} f(x,y)\,d\nu(y).$$

We can extend each of these restricted functions to Ω by setting it to be 0 outside Ω_i and the extensions are measurable. Now the function we want is the sum of all these and so is measurable.

In other words

$$\int_\Phi f(x,y)\,d\nu(y) = \sum_i \sum_j \int_{\Phi_j} \mathbf{1}_{\Omega_i}(x) f(x,y)\,d\nu(y).$$

□

We now come to the first version of Fubini's Theorem, Tonelli's Theorem.

Theorem (**Tonelli: Fubini for positive functions**). *If $(\Omega, \mathcal{F}, \mu)$ and $(\Phi, \mathcal{G}, \nu)$ are σ-finite measure spaces, and $f : \Omega \times \Phi \to [0, \infty)$ is measurable wrt $\sigma(\mathcal{F} \times \mathcal{G})$ then*

$$\begin{aligned}\int_{\Omega\times\Phi} f(x,y)\,d\mu\otimes\nu &= \int_\Omega \left(\int_\Phi f(x,y) d\nu(y)\right) d\mu(x) \\ &= \int_\Phi \left(\int_\Omega f(x,y) d\mu(x)\right) d\nu(y).\end{aligned}$$

Proof. By cutting up the space we may assume that the measures are finite. The repeated integrals make sense by the two preceding theorems.

Indicator If f is the indicator of a rectangle $F \times G$ then

$$\mathbf{1}_{F\times G}(x,y) = \mathbf{1}_F(x)\mathbf{1}_G(y)$$

so we can do the integrals.

$$\int_{\Omega\times\Phi} \mathbf{1}_{F\times G}(x,y)\,d\mu\otimes\nu = \mu\otimes\nu(F\times G) = \mu(F)\nu(G).$$

$$
\begin{aligned}
\int_\Omega \left(\int_\Phi \mathbf{1}_F(x)\mathbf{1}_G(y) d\nu(y) \right) d\mu(x) &= \int_\Omega \mathbf{1}_F(x) \left(\int_\Phi \mathbf{1}_G(y) d\nu(y) \right) d\mu(x) \\
&= \int_\Omega \mathbf{1}_F(x) \nu(G) d\mu(x) \\
&= \nu(G)\mu(F).
\end{aligned}
$$

Now for $A \in \sigma(\mathcal{F} \times \mathcal{G})$ let

$$
\theta(A) = \int \left(\int \mathbf{1}_A \, d\mu(x) \right) \, d\nu(y).
$$

Let $\mathcal{H}$ be the family of sets $A \in \sigma(\mathcal{F} \times \mathcal{G})$ for which $\theta(A) = \mu \otimes \nu(A)$. As in the previous theorem it suffices to prove that $\mathcal{H}$ is a d-system. For complements this is just linearity.

If (A_i) is an increasing sequence of sets in $\mathcal{H}$ and $A = \bigcup A_i$ then by applying the continuity of measure to $\mu \otimes \nu$ and MON to the integrals twice we get

$$
\theta(A) = \lim_n \theta(A_n) = \lim_n \mu \otimes \nu(An) = \mu \otimes \nu(A).
$$

Simple and general For simple functions we just use the additivity of the integral and for general functions we approximate by simple functions and use MON again twice. □

By now you realise that we need σ-finiteness whenever the argument uses the spooky magic. The reason is pretty much always the same: we need to check measures of complements and we can't subtract infinities. In the case of the measurability of sections we work only with sets: no measures or integrals. So we can make do with σ-algebras directly.

The full Fubini Theorem works for integrable functions.

Theorem (**Fubini for integrable functions**). *If $(\Omega, \mathcal{F}, \mu)$ and $(\Phi, \mathcal{G}, \nu)$ are σ-finite measure spaces, and $f : \Omega \times \Phi \to \mathbf{R}$ is integrable wrt $\sigma(\mu \otimes \nu)$ then its sections are integrable a.e., its marginals are integrable and*

$$
\begin{aligned}
\int_{\Omega \times \Phi} f(x,y) \, d\mu \otimes \nu &= \int_\Omega \left(\int_\Phi f(x,y) d\nu(y) \right) d\mu(x) \\
&= \int_\Phi \left(\int_\Omega f(x,y) d\mu(x) \right) d\nu(y).
\end{aligned}
$$

Proof. By the Theorem for positive functions we know that

$$\begin{aligned}\int_{\Omega\times\Phi} |f(x,y)|\, d\mu\otimes\nu &= \int_\Omega \left(\int_\Phi |f(x,y)| d\nu(y)\right) d\mu(x) \\ &= \int_\Phi \left(\int_\Omega |f(x,y)| d\mu(x)\right) d\nu(y)\end{aligned}$$

and we are given that these quantities are finite.

Since

$$\int_\Omega \left(\int_\Phi |f(x,y)| d\nu(y)\right) d\mu(x) < \infty$$

we know that

$$\int_\Phi |f(x,y)| d\nu(y) < \infty$$

for almost all x and similarly the other way around.

Now if we write $f = f_+ - f_-$ as a difference of positive functions we have

$$\int_\Phi f(x,y) d\nu(y) = \int_\Phi f_+(x,y) d\nu(y) - \int_\Phi f_-(x,y) d\nu(y)$$

whenever the section is integrable. So the marginal is measurable because we know this for the marginals of f_+ and f_-.

Finally

$$\begin{aligned}\int_{\Omega\times\Phi} f\, d\mu\otimes\nu &= \int_{\Omega\times\Phi} f_+\, d\mu\otimes\nu - \int_{\Omega\times\Phi} f_-\, d\mu\otimes\nu \\ &= \int_\Omega \left(\int_\Phi f_+ d\nu\right) d\mu - \int_\Omega \left(\int_\Phi f_- d\nu\right) d\mu \\ &= \int_\Omega \left(\int_\Phi f d\nu\right) d\mu\end{aligned}$$

and similarly the other way around. □

The use of Fubini Suppose that $f : \Omega\times\Phi \to \mathbf{R}$ is measurable on a σ-finite product space and

$$\int \left(\int |f(x,y)|\, d\mu(x)\right) d\nu(y) < \infty.$$

Then by Fubini for positive functions

$$\int |f|\, d\mu\otimes\nu = \int \left(\int |f(x,y)|\, d\mu(x)\right) d\nu(y).$$

So f is integrable and we can apply Fubini for integrable functions to interchange the order of integration. In particular

$$\int |f(x,y)|\, d\nu(y)$$

is finite for almost every x.

A classic example

$$\lim_{N\to\infty} \int_0^N \frac{\sin t}{t}\, dt = \frac{\pi}{2}.$$

We remarked that the integral on $[0,\infty)$ does not exist in the Lebesgue sense. We can compute the limit using our machinery.

Recall from **HW** that for $t > 0$

$$\frac{1}{t} = \int_0^\infty e^{-xt}\, dx.$$

For each fixed N

$$\begin{aligned} \int_0^N \left(\int_0^\infty |\sin t| e^{-xt}\, dx \right) dt &= \int_0^N \frac{|\sin t|}{t}\, dt \\ &\leq \int_0^N 1\, dt = N < \infty. \end{aligned}$$

So by Fubini for integrable functions we can interchange the order of integration:

$$\begin{aligned} \int_0^N \frac{\sin t}{t}\, dt &= \int_0^N \sin t \int_0^\infty e^{-xt}\, dx\, dt \\ &= \int_0^\infty \left(\int_0^N e^{-xt} \sin t\, dt \right) dx \end{aligned}$$

The inner integral is a standard integration by parts trick.

$$\int_0^N e^{-xt} \sin t\, dt = \frac{1 - e^{-Nx}(\cos N - x \sin N)}{1 + x^2}.$$

If $x > 0$ then as $N \to \infty$ this converges to $\frac{1}{1+x^2}$.

You can check by DOM that $t \mapsto e^{-xt}\sin t$ is integrable over $[0,\infty)$ and that

$$\int_0^\infty e^{-xt}\sin t\,dt = \frac{1}{1+x^2},$$

but we don't need that here.

We want to confirm the remaining (red) step:

$$\begin{aligned}
\lim_{N\to\infty}\int_0^N \frac{\sin t}{t}\,dt &= \lim_{N\to\infty}\int_0^\infty \left(\int_0^N e^{-xt}\sin t\,dt\right)dx \\
&= \lim_{N\to\infty}\int_0^\infty \frac{1-e^{-Nx}(\cos N - x\sin N)}{1+x^2}\,dx \\
&{\color{red}=} \int_0^\infty \lim_{N\to\infty}\frac{1-e^{-Nx}(\cos N - x\sin N)}{1+x^2}\,dx \\
&= \int_0^\infty \frac{1}{1+x^2}\,dx
\end{aligned}$$

since the latter is $\pi/2$ by using $\tan^{-1}$ and MON.

We want to use DOM to prove

$$\lim_{N\to\infty}\int_0^\infty \frac{1-e^{-Nx}(\cos N - x\sin N)}{1+x^2}\,dx$$
$$= \int_0^\infty \lim_{N\to\infty}\frac{1-e^{-Nx}(\cos N - x\sin N)}{1+x^2}\,dx.$$

Now $|\cos N| \le 1$, $|\sin N| \le 1$ and for $x \ge 0$ we have $0 < e^{-Nx} \le 1$ and $0 \le xe^{-Nx} \le 1/N$. So the convergence is dominated by the integrable function

$$\frac{1+1+1}{1+x^2}.$$

Remark In principle our theory only works for limits along countable sequences. In this case we could just observe that we can compute what we want by restricting to integer values of N. But as remarked when we differentiated under the integral, our theory tells us that the limit along reals exists because we get the same limit along every sequence of reals approaching ∞.

Chapter 6. L_p

The definition

In the norms, metrics and topologies course you met the ℓ_p spaces. For each p with $1 \leq p < \infty$, ℓ_p is the space of sequences (x_n) of real (or complex) numbers for which

$$\sum_1^\infty |x_n|^p < \infty$$

equipped with the norm

$$\|(x_n)_1^\infty\|_p = \left(\sum_1^\infty |x_n|^p\right)^{1/p}.$$

The fact that this is a norm follows from Minkowski's Inequality.

If $p = \infty$ then the norm is simply

$$\|(x_n)\|_\infty = \sup_n |x_n|.$$

Exercise Note that if the sequence is finite then

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p.$$

In the case of infinite sequences this need not be true: the sequence might be bounded but with infinite ℓ_p norm for all $p < \infty$:

$$(1, 1, 1, \ldots).$$

We are going to describe a generalisation of these spaces. Given a measure space $(\Omega, \mathcal{F}, \mu)$ we consider measurable functions $f : \Omega \to \mathbf{R}$ for which

$$\int_\Omega |f|^p < \infty.$$

The map $x \to |x|^p$ is continuous on $\mathbf{R}$ and hence measurable, so $|f|^p$ is measurable. (See the algebra of measurable functions in Chapter 3.) If $\Omega = \mathbf{N}$ and the measure is counting measure on all the subsets of $\mathbf{N}$ then we get ℓ_p. For a general measure space we have to be a bit careful.

If f and g differ but only on a set of measure 0 then

$$\int_\Omega |f-g|^p = 0.$$

This means that $\|f-g\| = 0$ and according to the rules for a norm we need $f = g$. For this reason we consider not functions themselves, but equivalence classes of functions. Two functions are equivalent if they agree a.e.

It is not difficult to check that the space of equivalence classes **is** a vector space with the inherited operations. You just need to check that the operations are well-defined: if $f_1 \sim f_2$ and $g_1 \sim g_2$ and $\lambda \in \mathbf{R}$ then $f_1 + g_1 \sim f_2 + g_2$ and $\lambda f_1 \sim \lambda f_2$. We tend to abuse notation by writing f for the equivalence class of f. If f and g agree a.e. then for $1 \le p < \infty$ we get

$$\|f\|_p = \|g\|_p.$$

It is easy to check that the space is closed under scalar multiplication and that

$$\|\lambda f\|_p = |\lambda|.\|f\|_p.$$

As in the case of ℓ_p the tricky thing is to check the triangle inequality. (See later.)

For $p = \infty$ we have to be doubly careful. If $f : [0,1] \to \mathbf{R}$ is defined by

$$f(x) = \begin{cases} 0 & \text{if } x > 0 \\ \\ 1 & \text{if } x = 0 \end{cases}$$

then $\sup |f(x)| = 1$ but f is equivalent to the function that is 0 everywhere. Consequently we define the norm slightly differently.

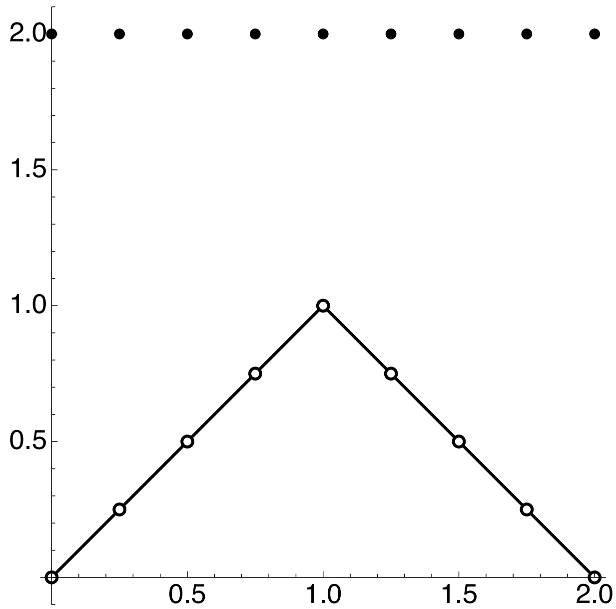
Let $f : \Omega \to \mathbf{R}$ be measurable. For each $a \in \mathbf{R}$ we ask whether there is a set of positive measure on which $f > a$. If so then the **essential supremum** of f will be at least a. If $(f > a)$ has measure 0 then the $\operatorname{ess\,sup}$ of f is at most a.

Formally

$$\operatorname{ess\,sup} f = \inf \{a : \mu(f > a) = 0\}$$

or $\operatorname{ess\,sup} f = \infty$ if the set is empty. If $b < \operatorname{ess\,sup} f$ then $\mu(f > b) > 0$. If $b \ge \operatorname{ess\,sup} f$ then $\mu(f > b) = 0$.

In this example $\sup f = 2$ but $\operatorname{ess\,sup} f = 1$.



Note that the function could be unbounded but with finite essential supremum.

We now define the L_∞ norm for a measure space $(\Omega, \mathcal{F}, \mu)$. If $f : \Omega \to \mathbf{R}$ is measurable then

$$\|f\|_\infty = \operatorname{ess\,sup} |f|.$$

Note that if $\operatorname{ess\,sup} |f| = S$ then we can find a function g with $\sup |g| = S$ and $f = g$ a.e. Conversely, if $f = g$ a.e. then

$$\|f\|_\infty = \|g\|_\infty.$$

The inequalities

There are a couple of important inequalities related to L_p spaces. We already mentioned Minkowski's Inequality which amounts to the triangle inequality in L_p. The other is Hölder's inequality. The indices $p > 1$ and $q > 1$ are called conjugate if

$$\frac{1}{p} + \frac{1}{q} = 1.$$

The pair 1 and ∞ are also called conjugate.

Theorem (**Hölder's Inequality II**)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* p *and* q *be conjugate indices. Suppose* $f \in L_p(\mu)$ *and* $g \in L_q(\mu)$*. Then* $f.g$ *is integrable and*

$$\int_\Omega fg\, d\mu \leq \|f\|_p.\|g\|_q.$$

If $p = q = 2$ this is the Cauchy-Schwarz inequality:

$$\int_\Omega fg\, d\mu \leq \left(\int_\Omega f^2\, d\mu\right)^{1/2} \left(\int_\Omega g^2\, d\mu\right)^{1/2}.$$

Note the homogeneity. If you double f then you double both sides of the inequality.

Suppose someone claimed that

$$\int fg \leq \left(\int f^2\right)\left(\int g^2\right).$$

Choose a pair of functions with say $\int f^2 = \int g^2 = 1$ and $\int fg = 1/2$. No contradiction. Now replace f by $f/100$. Then $\int fg$ becomes $1/200$ but $\int f^2$ becomes $1/10,000$ and the inequality is false. So the inequality is unlikely to be correct. The need for homogeneity corresponds to dimension analysis in applied maths: you can't have length on one side and area on the other.

To prove Hölder we use the concavity of the logarithm. If $u, v > 0$ and $\lambda \in [0, 1]$ then

$$(1 - \lambda)\log u + \lambda \log v \leq \log((1 - \lambda)u + \lambda v).$$

HW

Taking exponentials (using the fact that the exponential is increasing) we get

$$u^{1-\lambda} v^\lambda \leq (1 - \lambda)u + \lambda v.$$

Theorem (**Hölder's Inequality I**). *If u and v are non-negative measurable functions on $(\Omega, \mathcal{F}, \mu)$ and $\lambda \in (0,1)$ then*

$$\int_\Omega u^{1-\lambda} v^\lambda \, d\mu \le \left(\int_\Omega u \, d\mu \right)^{1-\lambda} \left(\int_\Omega v \, d\mu \right)^{\lambda}.$$

Proof. If $\int u = 0$ then $u = 0$ a.e. so both sides are 0. Similarly v. If $\int u = \infty$ then the inequality is automatically true. So we may assume that $\int u$ and $\int v$ are positive numbers. The inequality doesn't change if we scale u or v by a positive number. So we may assume that $\int u = \int v = 1$ and we want to prove that

$$\int_\Omega u^{1-\lambda} v^\lambda \, d\mu \le 1.$$

By the convexity statement

$$\begin{aligned} \int_\Omega u^{1-\lambda} v^\lambda \, d\mu &\le \int_\Omega [(1-\lambda)u + \lambda v] \, d\mu \\ &= (1-\lambda) \int_\Omega u \, d\mu + \lambda \int_\Omega v \, d\mu = 1. \end{aligned}$$

□

Note that the intermediate inequality is not homogeneous. By restricting to functions with integral 1 we dehomogenise and can then use the linearity of the integral.

Theorem (**Hölder's Inequality II**). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and p and q be conjugate indices. Suppose $f \in L_p(\mu)$ and $g \in L_q(\mu)$. Then $f.g$ is integrable and*

$$\int_\Omega fg \, d\mu \le \|f\|_p . \|g\|_q.$$

Proof. If $p \in (1, \infty)$ set $u = |f|^p$, $v = |g|^q$ and $1 - \lambda = 1/p$ so that $\lambda = 1/q$. Then

$$\int_\Omega |f.g| = \int_\Omega u^{1-\lambda} v^\lambda \le \left(\int_\Omega u \, d\mu \right)^{1-\lambda} \left(\int_\Omega v \, d\mu \right)^{\lambda} = \|f\|_p . \|g\|_q.$$

If $p = 1$ and $q = \infty$ then set $M = \operatorname{ess\,sup} |g| = \|g\|_\infty$. Then

$$\int_\Omega |f.g| \le M \int_\Omega |f| = \|f\|_1 . \|g\|_\infty.$$

□

In the N, M and T course you saw a couple of ways to prove Minkowski's Inequality. One approach is to use Hölder. We know that $\int f.g \leq \|f\|_p.\|g\|_q$. If $1 \leq p < \infty$, then for a given f I claim that there is equality if

$$g(x) = |f(x)|^{p-1}\text{sign}(x)$$

for each $x \in \Omega$.

To see it for $p > 1$ observe that

$$f.g = |f|^p$$

and

$$|g|^q = |f|^{(p-1)q} = |f|^p.$$

Therefore

$$\|f\|_p.\|g\|_q = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p} \left(\int_\Omega |f|^p \, d\mu\right)^{1/q} = \int_\Omega |f|^p = \int_\Omega f.g.$$

We also have

$$\|g\|_q^q = \|f\|_p^p.$$

In the case $p = 1$ we have $g(x) = \text{sign}(x)$ so clearly $\|g\|_\infty = 1$ and

$$\int_\Omega f.g = \int_\Omega |f| = \|f\|_1.$$

So, for each f and for $1 \leq p < \infty$ we can find a function g for which there is equality in Hölder and $\|g\|_q \neq 0$ if $\|f\|_p \neq 0$.

Lemma (**Hölder's Inequality III**)**.** *If f is a measurable function on $(\Omega, \mathcal{F}, \mu)$ and p and q are conjugate then*

$$\|f\|_p = \sup_{\|g\|_q = 1} \int_\Omega f.g.$$

We can realise the L_p norm as an integral against an appropriate unit function in L_q.

Proof. If $1 \leq p < \infty$ then firstly

$$\int f.g \leq \|f\|_p.\|g\|_q = \|f\|_p$$

whenever $\|g\|_q = 1$ by Hölder.

In the other direction we may assume that $\|f\|_p \neq 0$ and choose a function g with $\|g\|_q > 0$ as above for which

$$\int f.g = \|f\|_p.\|g\|_q.$$

This inequality remains unchanged if we rescale g by a positive number. So we can assume that $\|g\|_q = 1$.

The case $p = \infty$ and $q = 1$ is **HW**. □

For $p = q = 2$ the function g is a multiple of

$$|f(x)|^{p-1}\text{sign}(x) = f(x).$$

Compare with **HW 3 Q 5**

Theorem (**Minkowski's Inequality**)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $p \in [1, \infty]$*. Suppose* f *and* h *belong to* $L_p(\mu)$*. Then* $f + h \in L_p(\mu)$ *and*

$$\|f + h\|_p \leq \|f\|_p + \|h\|_p.$$

Proof. For $p = \infty$ (and $p = 1$) this is obvious. If $1 \leq p < \infty$ then at each point of Ω the convexity of the function $x \mapsto |x|^p$ shows that

$$\left|\frac{f+h}{2}\right|^p \leq \frac{|f|^p + |h|^p}{2}$$

so $\|f + h\|_p < \infty$.

Now take q to be the index conjugate to p and choose g with $\|g\|_q = 1$ and $\|f + h\|_p = \int(f + h)g$. Then

$$\|f + h\|_p = \int(f + h)g = \int f.g + \int h.g \leq \|f\|_p.\|g\|_q + \|h\|_p.\|g\|_q = \|f\|_p + \|h\|_p.$$

By writing $\|f + h\|$ as a linear expression in $f + h$ we can again use the linearity of the integral. □

We don't really need the first part. We could apply Hölder to simple functions (which are automatically in L_p) and then take limits to get the general case. Minkowski's inequality shows that $L_p(\mu)$ is a vector space, it is closed under addition, and that $\|.\|_p$ is a norm on the space.

Definition (L_p)**.** *Given a measure space* $(\Omega, \mathcal{F}, \mu)$ *and* $p \in [1, \infty]$ *we define* $L_p(\mu)$ *to be the vector space of equivalence classes of measurable functions* f *for which* $\|f\|_p < \infty$*, equipped with this norm, where*

$$\|f\|_p = \begin{cases} \left(\int_\Omega |f|^p \, d\mu\right)^{1/p} & \text{if } 1 \le p < \infty \\ \\ \operatorname{ess\,sup} |f| & \text{if } p = \infty. \end{cases}$$

Completeness

There is one further fact that we need, whose importance becomes clear if you do the Functional Analysis I (or II) course: completeness. Suppose (f_n) is a sequence in L_p with

$$\sup_{m,n>N} \|f_m - f_n\|_p \to 0$$

as $N \to \infty$. We want to show that (f_n) converges in L_p. As usual, if we can find a convergent subsequence then the Cauchy property will force the full sequence to converge to the same limit.

Choose $n_1 < n_2 < \cdots$ so that if $m > n_k$ then

$$\|f_m - f_{n_k}\|_p < 2^{-k}.$$

Now we have a subsequence (f_{n_k}) such that $\|f_{n_{k+1}} - f_{n_k}\|_p \le 2^{-k}$ which implies by the triangle inequality in L_p

$$\left\| \sum_1^\infty |f_{n_{k+1}} - f_{n_k}| \right\|_p \le \sum_1^\infty \|f_{n_{k+1}} - f_{n_k}\|_p \le 1.$$

So the sum $\sum_1^\infty |f_{n_{k+1}} - f_{n_k}|$ is finite a.e. That in turn means that the sequence (f_{n_k}) is Cauchy a.e. Let f be its limit. We have that $f_{n_k} \to f$ a.e. and we want

$$\|f_{n_k} - f\|_p \to 0.$$

Now for $1 \le p < \infty$ by Fatou

$$\int |f_{n_k} - f|^p = \int |f_{n_k} - \lim_{m\to\infty} f_{n_m}|^p \le \liminf_m \int |f_{n_k} - f_{n_m}|^p \le 2^{-pk}.$$

So for each k

$$\|f_{n_k} - f\|_p \leq 2^{-k}.$$

For $p = \infty$

$$|f_{n_k} - f| = |f_{n_k} - \lim_{m\to\infty} f_{n_m}| = \lim_{m} |f_{n_k} - f_{n_m}| \leq 2^{-k}$$

a.e. □

Naturally certain choices of measure space are more important than others: $\mathbf{N}$, $\mathbf{R}$, $[0,1]$, $(0,\infty)$, $\mathbf{R}^d$ and the circle $\mathbf{T}$. We already mentioned that if $\Omega = \mathbf{N}$ we get the ℓ_p spaces. Recall from last year that if $p < q$ then $\ell_p \subset \ell_q$ and in fact

$$\|(x_n)\|_q \leq \|(x_n)\|_p.$$

This depends upon the fact that every non-empty set has measure at least 1.

If $\Omega = [0,1]$ (or $\mu(\Omega) = 1$) then the inclusions and inequalities go the opposite way. For example by Cauchy-Schwarz

$$\int_0^1 |f| \; = \; \int_0^1 |f|.1 \leq \left(\int_0^1 |f|^2\right)^{1/2} \left(\int_0^1 1^2\right)^{1/2} = \left(\int_0^1 |f|^2\right)^{1/2}.$$

The general case is **HW**.

We can define complex normed spaces using complex scalars instead of real ones and the integral by

$$\int f\, d\mu = \int \Re f\, d\mu + i \int \Im f\, d\mu.$$

The space $L_2(\mathbf{T})$ with complex scalars is especially notable. It makes sense to scale the measure and so if $f : \mathbf{T} \to \mathbf{C}$ we set

$$\|f\|^2 = \frac{1}{2\pi} \int_0^{2\pi} \left|f(e^{i\theta})\right|^2 \, d\theta.$$

The functions $e^{i\theta} \mapsto e^{in\theta}$ form an orthogonal sequence of unit functions.

We can express each function in the space as a sum

$$f\left(e^{i\theta}\right) = \sum_{-\infty}^{\infty} \hat{f}(n) e^{in\theta},$$

its Fourier series. The sum converges to f in the normed space $L_2(\mathbf{T})$.

In fact the Fourier series of a function in L_p converges in L_p for $1 < p < \infty$ but this is harder to prove. It need not converge in L_1 or L_∞ even if f is continuous. A deep theorem of Carleson guarantees that if $f \in L_2$ then its Fourier series converges a.e.

Chapter 7. Approximation of measurable functions

In this chapter we shall prove some theorems suggesting that measurable functions are not too much worse than continuous functions. But this isn't really saying what it appears to. It isn't really that measurable functions are nicer than you think: it's that approximation is a looser relationship than you think.

We know that if f is a non-negative measurable function on $(\Omega, \mathcal{F}, \mu)$ then we can find simple functions $f_n \uparrow f$ and in this case $\int f_n \to \int f$. If $\int f < \infty$ then

$$\int |f - f_n| = \int (f - f_n) = \int f - \int f_n \to 0.$$

So we can approximate f in the space L_1 by simple functions. By splitting into positive and negative parts we see that the same is true for arbitrary functions in L_1.

We can do the same in $L_p(\mu)$ for each $p < \infty$ because if $0 \leq f_n \leq f$

$$(f - f_n)^p \leq f^p - f_n^p$$

HW.

It is also easy to check that we can approximate functions in L_∞ by simple functions.

Now suppose that $\Omega = [0, 1]$ and we are working with Lebesgue measure. Can we approximate by **continuous** functions in $L_p([0, 1])$? Certainly not if $p = \infty$. If f has a jump then any continuous function must differ from f by a large amount on a set of positive measure. L_∞ is a huge space. The space of continuous functions is much smaller.

However for other L_p the continuous functions are dense: you can approximate a function in L_p by continuous ones. It suffices to approximate the indicator of a measurable set: $E \subset [0, 1]$. Given $\varepsilon > 0$ choose a closed set K and an open set U with

$$K \subset E \subset U$$

and $\lambda(U \backslash K) < \varepsilon$.

if C is a closed set then set

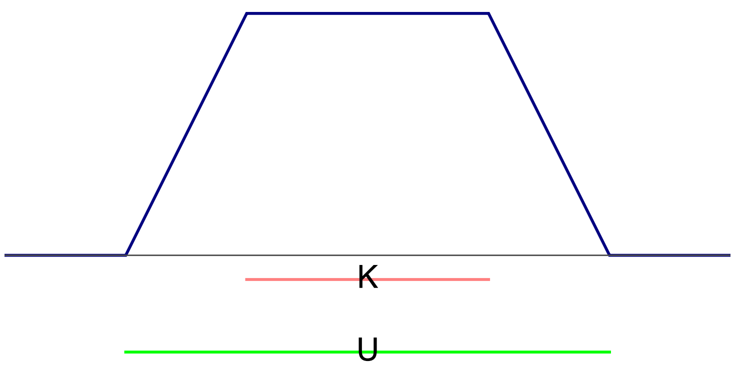
$$d(x, C) = \inf\{|x - y| : y \in C\}.$$

In **HW 2 Q 5** you are asked to show that the function $x \mapsto d(x, C)$ is continuous and if $x \notin C$ then $d(x, C)$ is strictly positive.

Now choose

$$c(x) = \frac{d(x, U^c)}{d(x, K) + d(x, U^c)}.$$

If $x \in U^c$ then $c(x) = 0$ while if $x \in K$ we have $c(x) = 1$. Thus, c is equal to 1 on K and vanishes outside U. So $c(x)$ agrees with (equals) $\mathbf{1}_E$ except on a small set.



That ensures that their difference has small L_p norm.

This construction can be souped up to something that looks rather surprising:

Theorem (**Lusin+Tietze**)**.** *Suppose $f : [0, 1] \to \mathbf{R}$ is measurable. Then for each $\varepsilon > 0$ there is a continuous $g : [0, 1] \to \mathbf{R}$ which agrees with f except on a set of measure at most ε.*

Proof. We can assume that f is non-negative. If f is unbounded then the sets

$$E_k = \{x \in [0, 1] : f(x) > k\}$$

form a decreasing sequence with empty intersection so their measures $\to 0$. So by discarding a set of measure $< \varepsilon/2$ we may assume that f is bounded. Now approximate by f_n as before

$$f_n(x) = m2^{-n} \quad \text{if } m2^{-n} \le f(x) < (m + 1)2^{-n} \ .$$

Note that the functions f_1, $f_2 - f_1$, $f_3 - f_2$ and so on are simple functions and that if $n \geq 1$ we have $0 \leq f_{n+1} - f_n < 2^{-n-1}$. So the sum

$$f = f_1 + \sum_{n=1}^{\infty} (f_{n+1} - f_n)$$

converges uniformly.

For each n choose a continuous function c_n which agrees with $f_{n+1} - f_n$ on a set of measure at least $1 - \varepsilon/2^{n+1}$ and c_0 agreeing with f_1 except on a set of measure $\varepsilon/2$. We may assume that for $n > 0$ we have $0 \leq c_n \leq 2^{-n-1}$ by considering $\min(c_n, f_{n+1} - f_n)$. Therefore the sum

$$c_0 + \sum c_n$$

also converges uniformly and is continuous. It agrees with f except on a set of measure at most

$$\varepsilon/2 + \varepsilon/4 + \varepsilon/8 + \ldots = \varepsilon.$$

□

Classically the Theorems of Lusin and Tietze are the following.

Theorem (**Lusin**)**.** *Suppose $f : [0,1] \to \mathbf{R}$ is measurable. Then for each $\varepsilon > 0$ there is a set $F \subset [0,1]$ with $\lambda(F) > 1 - \varepsilon$ such that $f|_F$ is continuous (with respect the topology on F).*

Theorem (**Tietze**)**.** *Suppose X is a compact Hausdorff space, $Y \subset X$ and $f : Y \to \mathbf{R}$ is continuous. Then there is a continuous function $g : X \to \mathbf{R}$ so that $g|_Y = f$.*

On a compact **metric** space this is easier to prove than in general: we can soup up the $d(x, C)$ argument that we used for indicators. By combining the two we get the previous theorem.

Chapter 8. The Radon-Nikodym Theorem

Signed measures

Suppose f is an integrable function on $(\Omega, \mathcal{F}, \mu)$. Then the map

$$A \mapsto \int_A f \, d\mu = \int_\Omega f \mathbf{1}_A \, d\mu$$

defines a finite set-function which is countably additive but not necessarily non-negative.
Exercise using DOM

Such a function is called a **signed measure**. It makes sense to ask whether we can go the other way. Are all signed measures given by integrals?

It is easy to see that this is not the case even for measures. Suppose ν is the δ measure on $\mathbf{R}$:

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

Then ν is not given by the integral of any function with respect to Lebesgue measure because $\nu(\{0\}) = 1$ but

$$\int_{\{0\}} f \, d\lambda = 0.$$

.

Definition (Signed measures)**.** *Let $\mathcal{F}$ be a σ-algebra on a set Ω. A function $\nu : \mathcal{F} \to \mathbf{R}$ is called a* **finite signed measure** *if*

- $\nu(\emptyset) = 0$
- *If $A_1, A_2, A_3, \ldots \in \mathcal{F}$ are disjoint then*

$$\nu\left(\bigcup A_i\right) = \sum \nu(A_i).$$

I will not deal with the case in which ν can take the value ∞ or $-\infty$. What property will guarantee that a signed measure **is** an integral?

Definition (Absolutely continuous measures)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A finite signed measure $\nu : \mathcal{F} \to \mathbf{R}$ is said to be* **absolutely continuous** *wrt μ if whenever $\mu(A) = 0$ we also have $\nu(A) = 0$.*

Notice that an integral

$$A \mapsto \int_A f \, d\mu$$

is necessarily absolutely continuous wrt μ. Our aim will be to prove the converse:

Theorem (**Radon-Nikodym**). *Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and ν a finite signed measure on $\mathcal{F}$ which is absolutely continuous wrt μ. Then there is an integrable function $f : \Omega \to \mathbf{R}$ with*

$$\nu(A) = \int_A f \, d\mu$$

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

The function f is called the Radon-Nikodym derivative of ν wrt μ and written

$$\frac{d\nu}{d\mu}.$$

The first step is a bit surprising.

Lemma (**Boundedness of finite signed measures**). *Let $(\Omega, \mathcal{F})$ be a measurable space and ν a finite signed measure on $\mathcal{F}$. Then ν is bounded: there is an M for which $|\nu(A)| \leq M$ for all $A \in \mathcal{F}$.*

Proof. If not then without loss of generality we can assume that

$$\sup\{\nu(A) : A \in \mathcal{F}\} = \infty.$$

Set

$$\theta(A) = \sup\{\nu(B) : B \subset A\}$$

for each $A \in \mathcal{F}$. θ tells us "where" the unboundedness happens.

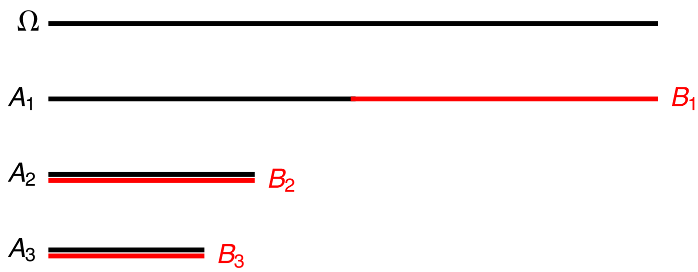
You can check that θ is a measure but all we need is that if $\theta(E \cup F) = \infty$ and $E \cap F = \emptyset$ then either $\theta(E) = \infty$ or $\theta(F) = \infty$ or both. To see this choose sets $B_n \subset E \cup F$ with $\nu(B_n) > n$. Then $\nu(B_n \cap E) + \nu(B_n \cap F) > n$ and hence either $\nu(B_n \cap E) > n/2$ or $\nu(B_n \cap F) > n/2$. For at least one of E and F we have $\nu(B_n \cap .)$ infinitely often and then $\theta(.) = \infty$.

Now we construct inductively a sequence $\Omega = A_0 \supset A_1 \supset A_2 \supset \cdots$ and $B_1, B_2, \ldots$, with for each n

$$B_n = A_n \quad \text{or} \quad B_n = A_{n-1} \backslash A_n$$

and

$$\theta(A_n) = \infty \quad \text{and} \quad \theta(B_n) > n.$$



Start with $A_0 = \Omega$. Choose $B_1 \subset A_0$ with $\nu(B_1) > 1$ which we can do since $\theta(A_0) = \infty$. Either $\theta(B_1) = \infty$ or $\theta(A_0 \backslash B_1) = \infty$ and we choose A_1 to be one of them with infinite θ-measure. Now choose $B_2 \subset A_1$ with $\nu(B_2) > 2$ and then choose A_2 to be whichever of B_2 and $A_1 \backslash B_2$ has infinite θ-measure. Continue.

There are two possibilities: for infinitely many indices n_k we have $B_{n_k} = A_{n_k - 1} \backslash A_{n_k}$ or for infinitely many we have $B_{n_k} = A_{n_k}$. In the first case these B_{n_k} are disjoint and hence

$$\nu\left(\bigcup B_{n_k}\right) = \sum \nu(B_{n_k} > \sum n_k = \infty$$

contradicting the finiteness of ν. In the second case $B_{n_1} \supset B_{n_2} \supset \cdots$ since the A_n are nested. But that means that

$$\nu\left(\bigcap B_{n_k}\right) = \lim_{k \to \infty} \nu(B_{n_k}) > \lim_{k \to \infty} n_k = \infty.$$

□

The Hahn-Jordan decomposition

The next step will be to decompose the signed measure into positive and negative parts (and then prove R-N just for measures). But there is a twist at the end of the story...

Theorem (**Hahn-Jordan decomposition**)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and ν a finite signed measure on $\mathcal{F}$. Then Ω can be partitioned into two sets P and N so that if $E \subset P$ then $\nu(E) \geq 0$ while if $E \subset N$ then $\nu(E) \leq 0$. As a result ν can be written as $\nu = \nu_+ - \nu_-$ where ν_+ and ν_- are measures.*

Proof. "Obviously" P will be the set with the largest υ-measure. So we hunt that set. Since ν is bounded we can set $\theta = \sup\{\nu(A) : A \in \mathcal{F}\}$. For each n choose A_n with

$\nu(A_n) \geq \theta - 2^{-n}$ and set

$$D_m = \bigcap_{n \geq m} A_n.$$

Note that $D_1 \subset D_2 \subset \cdots$ because as we increase m we intersect fewer sets. We shall prove that

$$\nu(D_m) \geq \theta - 2^{1-m}. \tag{1}$$

Since A_n loses only 2^{-n} of the measure the intersection cannot lose more than $\sum_{n \geq m} 2^{-n} = 2^{1-m}$. Once this is done we set $P = \bigcup D_n$. Then $\nu(P) = \lim \nu(D_n) = \theta$. Since this is the maximum possible measue there cannot be any positive measure outside P to add or any negative measure inside P to remove.

To establish (1) it is enough to prove that for each $p \geq m$

$$\nu\left(\bigcap_m^p A_n\right) \geq \theta - \sum_m^p 2^{-n}.$$

If $p = m$ then we want $\nu(A_m) \geq \theta - 2^{-m}$ which is how we chose A_m. Suppose inductively that the statement holds for some p and set $B = \bigcap_m^p A_n$. Then

$$\begin{aligned} \nu(B \cap A_{p+1}) &= \nu(B) + \nu(A_{p+1}) - \nu(B \cup A_{p+1}) \\ &\geq \theta - \sum_m^p 2^{-n} + \theta - 2^{-(p+1)} - \theta \\ &= \theta - \sum_m^{p+1} 2^{-n}. \end{aligned}$$

This completes the inductive step.

Finally we can set $\nu_+ = (E) = \nu(E \cap P)$ and $\nu_-(E) = \nu(E \cap N)$ to express ν as a difference of measures. □

H-J means that we only need to prove R-N for (positive) measures. But it also "proves" R-N. Suppose we are hunting for the function f. If we know, for each $r \in \mathbf{R}$, where $f \geq r$ then we "know" f. But if $\nu(E) = \int_E f \, d\mu$ then

$$\nu(E) - r\mu(E) = \int_E (f - r) d\mu.$$

So the set where $f \geq r$ or $f - r \geq 0$ is the P-set for the signed measure $\nu - r\mu$.

The Radon-Nikodym Theorem

We are now ready to prove:

Theorem (Radon-Nikodym)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a* σ*-finite measure space and* ν *a finite signed measure on* $\mathcal{F}$ *which is absolutely continuous wrt* μ*. Then there is an integrable function* $f : \Omega \to \mathbf{R}$ *with*

$$\nu(A) = \int_A f \, d\mu$$

for every $A \in \mathcal{F}$*.*

Proof. By H-J we can assume that ν is a measure. We can also assume that μ is finite by breaking into countably many pieces.

Let $\mathcal{M}$ be the class of non-negative measurable functions h for whcih

$$\int_E h \, d\mu \leq \nu(E)$$

for all $E \in \mathcal{F}$. The biggest h in $\mathcal{M}$ should work. Since $\nu(\Omega) < \infty$ we know that

$$\alpha = \sup \left\{ \int_\Omega h \, d\mu : h \in \mathcal{M} \right\}$$

is finite. For each n choose $h_n \in \mathcal{M}$ with $\int h_n \geq \alpha - 1/n$ and then set

$$g_n = \max(h_1, h_2, \ldots, h_n).$$

(g_n) is an increasing sequence and I claim that $g_n \in \mathcal{M}$.

We can write $\Omega = \bigcup_1^n U_j$ disjointly where U_j is a set on which $g_n = f_j$. For each $E \in \mathcal{F}$

$$\begin{aligned} \nu(E) &= \sum_1^n \nu(E \cap U_j) \geq \sum_1^n \int_{E \cap U_j} f_j \, d\mu \\ &= \sum_1^n \int_{E \cap U_j} g_n \, d\mu = \int_E g_n \, d\mu. \end{aligned}$$

So $g_n \in \mathcal{M}$ and if we set $f = \lim_n g_n$ then by MON

$$\int_E f \, d\mu = \lim_n \int g_n \, d\mu \leq \nu(E)$$

for every $E \in \mathcal{F}$ and so $f \in \mathcal{M}$. Since $g_n \geq f_n$ we have $\int g_n \geq \alpha - 1/n$ and so $\int f = \alpha$. We have found a function in $\mathcal{M}$ with the largest integral. We need to check that $\int_E f \geq \nu(E)$ on every set.

It suffices to show that for each n and every E

$$\nu(E) \leq \int_E f \, d\mu + \frac{1}{n}\mu(E)$$

because $\mu(E) < \infty$. Consider the signed measure

$$E \mapsto \nu(E) - \int_E f \, d\mu - \frac{1}{n}\mu(E).$$

We need to prove that it is non-positive. By H-J we can find sets P_n and N_n on which it is respectively non-negative and non-positive. To show that P_n isn't really there I use it to build a function with bigger integral than f.

Define

$$\hat{f}(x) = \begin{cases} f(x) + 1/n & \text{if } x \in P_n \\ \\ f(x) & \text{if } x \in N_n. \end{cases}$$

For any $E \subset P_n$

$$\nu(E) \geq \int_E f \, d\mu + \frac{1}{n}\mu(E) = \int_E \hat{f} \, d\mu.$$

On the other hand if $E \subset N_n$

$$\nu(E) \geq \int_E f \, d\mu = \int_E \hat{f} \, d\mu$$

because $f \in \mathcal{M}$. Hence $\hat{f} \in \mathcal{M}$ but

$$\int_\Omega \hat{f} = \int_\Omega f + \frac{1}{n}\mu(P_n) = \alpha + \frac{1}{n}\mu(P_n)$$

contradicting the definition of α unless $\mu(P_n) = 0$. Now since ν is absolutely continuous wrt μ we also get $\nu(P_n) = 0$ and so the signed measure in question is zero. □

Key example Suppose $\mathcal{F}$ and $\mathcal{G}$ are two σ-algebras on Ω with $\mathcal{G} \subset \mathcal{F}$ and let $X : \Omega \to \mathbf{R}$ be a function which is measurable wrt $\mathcal{F}$ and integrable. X may not be measurable wrt $\mathcal{G}$ but for each set $A \in \mathcal{G}$ we can define

$$\nu(A) = \int_A X \, d\mu$$

because $A \in \mathcal{F}$. According to R-N there is a function Y **which is measurable wrt** $\mathcal{G}$ so that

$$\nu(A) = \int_A Y \, d\mu$$

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

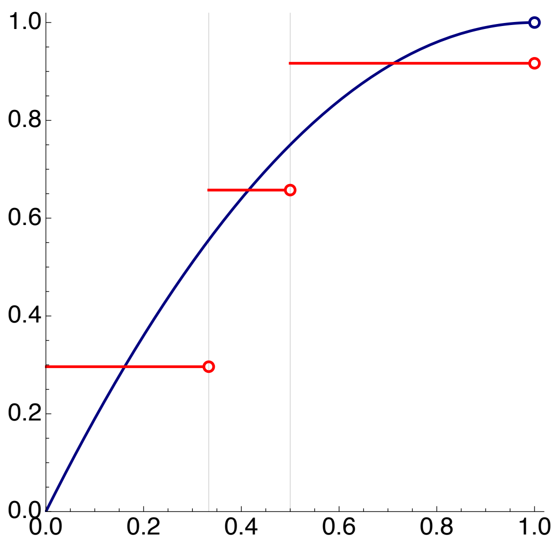
Y is measurable wrt $\mathcal{G}$ and

$$\int_A Y \, d\mu = \int_A X \, d\mu$$

for each $A \in \mathcal{G}$. Y is the "best $\mathcal{G}$-measurable approximation" to X. In the **HW** you will see that for example if $\mathcal{G}$ is the σ-algebra on $[0, 1)$ generated by say

$$[0, 1/3), \quad [1/3, 1/2), \quad [1/2, 1)$$

then Y is obtained from X by averaging over these sets.



This procedure is known as the **conditional expectation** in probability and is absolutely central to the study of random processes.

The term "best $\mathcal{G}$-measurable approximation" is genuinely appropriate if X is in L_2. In this case Y is the $\mathcal{G}$-measurable function which minimises

$$\|X - Y\|_2.$$

HW

Index