Math 564: Real Analysis and Measure Theory, Fall 2025

Ezra Huang

The following is a set of notes taken for the course Math 564: Real Analysis and Measure Theory, taught by Anush Tserunyan.

Topics covered are: measures, their construction and properties, pocket tools like Borel–Cantelli lemmas, regularity, tightness, and measure exhaustion; measurable functions and integration, modes of convergence, Fubini–Tonelli theorem signed measures (Hahn decomposition), measure differentiation (Lebesgue decomposition, Radon–Nikodym derivative), Lebesgue density theorem in $\mathbb{R}^d$, and a.e. differentiable functions; and examples and theorems from ergodic theory interspersed with the main material.

For homework sets and the original lecture notes, see the course website.

Motivation for measure theory

Probability The field of probability was not developed until Kolmogorov laid an axiomatic and rigorous foundation using measure theory that was, at the time, being developed by Lebesgue, Borel, the French school, and the Polish school. Probability was one of the big motivations to understand what was happening in probability theory.

It is easy to understand the probability theory behind a coin toss when the probability of 1 is $p \in (0,1)$ and of 0 is $1-p$. For each word $w \in 2^n := \{0,1\}^n$, the probability of each coin toss resulting in this word w is:

$$\mathbb{P}(w) = p^{\text{\# of ones in w}}(1-p)^{\text{\# of zeros in w}}$$

What if $n = \infty$? In other words, we consider the space $2^{\mathbb{N}}$ of *infinite binary sequences* with the same probabilities of tossing a 1 or 0. How can we define the probability of "events" in this space?

Geometry We want to have a robust notion of volume in $\mathbb{R}^d, d \geq 1$, i.e. we would like to determine the volume of a large class of subsets of $\mathbb{R}^d$. We know what the volume of a box $B := I_1 \times \ldots \times I_d \subseteq \mathbb{R}^d$ where I_j is an interval:

$\text{Ln}(I) = \text{right endpoint} - \text{left endpoint}$

$$\text{Volume}(B) = \prod_j \text{Ln}(I_j)$$

We want to extent this to a class of sets that are closed under countable operations: complements, unions, intersections.

(Real) Analysis The class of Riemann-integrable functions is not closed under anything . We want a more stable class of integrable functions. Indeed, even a pointwise limit of continuous functions on $[0,1]$ is not in general Riemann-integrable. But the whole subject of analysis is about approximation and limits, so we would like to extend the class of integrable functions so that it is closed under pointwise limits. Clearly, for a subset $B \subseteq \mathbb{R}^d$, the integral of it's indicator function $\mathbb{1}_B$ will simply be the volume of B, so this task subsumes the previous task about volumes.

"you can sneeze on a Riemann integrable function and it becomes non-Riemann integrable."

Remember that continuous function on $[0,1]$ are Riemann integrable, since all continuous functions on closed and bounded intervals are Riemann integrable

Let $C \subseteq [0,1]$ be a positive measure Cantor set. Define $U := [0,1] \setminus C = (0,1) \setminus C$. C is closed, hence U is open (open set minus closed set is closed). $\mathbb{1}_U \notin \mathcal{R}([0,1])$ since it's set of discontinuity points is C, which is uncountable. We build a sequence of continuous functions that converge to $\mathbb{1}_U$; let I_s, $s \in 2^{\mathbb{N}}$ be the binary tree of the middle open intervals, as in the construction of C, so U is the (disjoint) union of all these I_s. For each $s \in 2^{\mathbb{N}}$, let $g_{s,n}(x) = 0$ if $x \in I_s^C$ and for $x \in I_s$, the graph of it's trapezoid of height 1 (in particular, continuous) which is very close to the indicator $\mathbb{1}_{I_s}$ so that $g_{n,s} \to \mathbb{1}_s$ as $n \to \infty$. Let $f_n := \sum_{s \in 2^{<n}} g_{s,n}$ i.e. trapezoids over all open middle intervals up to level n. Clearly the f_n are continuous and their limit is $\mathbb{1}_U$.

Contents

1 Measures, their constructions, and properties

1.1 Polish spaces

We now define a very robust class of metric spaces that we will be working with throughout the course and exist naturally in analysis and related fields. We generalize the most important properties of the reals: separability (admits rationals) and completeness.

Definition 1.1.1 (Polish space)**.** A metric space (X, d) is called Polish if d is complete and X is separable.

Recall: X is separable $\iff$ X admits a countable dense subset, d is complete $\iff$ every Cauchy sequence converges.

Definition 1.1.2 (Second countable)**.** A metric space (X, d) is second countable if it admits a countable basis of open sets.

Definition 1.1.3 (Basis)**.** In a metric space X, a basis is a collection U of open subsets of X such that every open set is a union (maybe uncountable) of sets in U.

Proposition 1.1.1. For any metric space X, X is separable $\iff$ X is second countable.

Proof. HW □

Example 1.1.1 (Examples of Polish spaces)**.**

(a) $\mathbb{R}$ and in general $\mathbb{R}^d$ with the metric $d_\infty = \max_i |x_i - y_i|$.

   (i) We know from MATH 255 that this is a complete metric.

   (ii) Rationals are dense and countable, so $\mathbb{Q}^d \subseteq \mathbb{R}^d$ is dense and countable.

   Hence $\mathbb{R}^d$ is Polish. We can also equip $\mathbb{R}^d$ with other equivalent metrics . An example is any p-norm, which is bi-Lipschitz equivalent to d_∞, i.e. $\exists C_p$ such that:

Two metrics are equivalent if they induce the same open sets.

$$\frac{1}{C_p} d_\infty \leq d_p \leq C_p d_\infty$$

   In particular, the spaces $(\mathbb{R}^d, d_p)$ are Polish.

(b) If (X, d) is Polish, then any closed subset is still Polish with the same metric.

   (i) Indeed, closedness ensures completeness.

   (ii) Any subspace of second countable spaces is second countable (hence separable).

Remember from MATH 255 (Honours Analysis 2) that in a complete metric space (X, d), a subset $A \subseteq X$ being closed implies that all Cauchy sequences in A converge.

Proof (of (b)(ii)). Take the countable basis for X and restrict it to (intersect each set with) A, then this new collection of sets is a countable basis for A. □

What about open subsets (say $(0,1) \subseteq \mathbb{R}$)? The same metric wouldn't work because it wouldn't be be complete, but maybe we can take a different equivalent metric that is complete. Indeed, d_∞ on $\mathbb{R}$ is complete and $\mathbb{R}$ "looks like" $(0,1)$, in other words, they are homeomorphic. We can "copy" the complete metric to (0,1) via homeomorphism. More concretely we may define $d_{(0,1)}(x,y) = d_\infty(x,y) + \left| \frac{1}{d_\infty(x,\{0,1\})} - \frac{1}{d_\infty(y,\{0,1\})} \right|$ where

$$d_\infty(x,A) = \inf_{t \in A} |x-t|$$

Two spaces are homeomorphic if there exists bijection that is continuous both ways between the two spaces.

Then $((0,1), d_{(0,1)})$ is Polish because:

(i) This is a complete metric on (0,1).

(ii) This metric is equivalent to d_∞, hence (0,1) is separable.

Proof of (i). Verify at home. □

Proof of (ii). Verify at home. □

d_∞ is equivalent to $d_{(0,1)} \implies$ both metrics induce the same open sets $\implies$ the basis that is the witness to second countability for d_∞ also shows second countability for $d_{(0,1)} \implies$ (0,1) is separable.

Such sets are called Polishable. In fact it is a theorem in descriptive set theory that a subset of a Polish set is Polishable $\iff$ it is G_δ (countable intersection of open sets).

(c) The space $C([0,1])$ of continuous functions on $[0,1]$ with uniform metric

$$d_u(f,g) = \sup_{x \in [0,1]} |f(x) - g(x)| \quad \left(= \max_{x \in [0,1]} |f(x) - g(x)| \right)$$

is Polish. Indeed, we know that:

(i) A Cauchy sequence of continuous functions converge to a continuous function, so d_u is complete.

(ii) Polynomials with rational coefficients form a countable dense set (by Weierstrass's theorem), or more easily, piecewise linear functions with finitely many pieces with rational breakpoints form a dense set.

Proof of (i). Result from MATH 255. □

See the Stone-Weierstrass Theorem

Proof of (ii). Verify at home. □

(d) The tree spaces: Cantor space $2^{\mathbb{N}} := \{0,1\}^{\mathbb{N}}$ and Baire space $\mathbb{N}^{\mathbb{N}}$ are Polish. More generally, let A be a nonempty countable set. Let $X := A^{\mathbb{N}}$ the space of infinite sequences of A. We can depict $A^{\mathbb{N}}$ as the infinite branches through the $A^{<\mathbb{N}}$:= the set of finite sequences in A.

We equip $A^{\mathbb{N}}$ with the metric:

$$d(x,y) := \begin{cases} 2^{-\Delta(x,y)} & \text{where } \Delta(x,y) = \min_{i \in \mathbb{N}} x_i \neq y_i \quad x \neq y \\ 0 & x = y \end{cases}$$

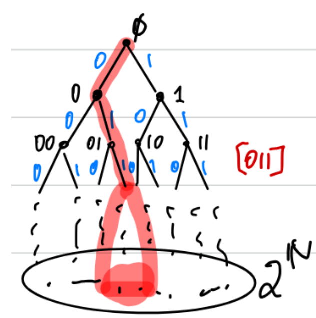


Figure 1.1: Visualization of $2^{\mathbb{N}}$.

Then d is indeed a metric on $A^{\mathbb{N}}$ and in fact in ultrametric (HW). Then:

(i) d is a complete metric (HW)

(ii) For a fixed $a_0 \in A$, the set of sequences which are eventually a_0 form a countable dense set .

Hence in general $(A^{\mathbb{N}}, d)$ is Polish.

An ultrametric is a metric with a stronger version of the triangle inequality: for $x,y,z \in X$, $d(x,z) \leq \max\{d(x,y), d(y,z)\}$.

Proof sketch (ii). Take a fixed $a_0 \in A$, Then the set of all the sequences that are eventually a_0 is a countable dense set (verify at home). □

1.1.1 *The topology of $A^{\mathbb{N}}$*

For $2^{-n} < r \leq 2^{-(n-1)}$ consider an open ball $B_r(x)$ in $(A^{\mathbb{N}}, d)$ with d defined above:

$$\begin{aligned} B_r(x) :=& \{y \in A^{\mathbb{N}} : d(x,y) < r\} \\ =& \{y \in A^{\mathbb{N}} : d(x,y) \leq 2^{-n}\} \\ =& \{y \in A^{\mathbb{N}} : x|_n = y|_n\} \\ =:& [x|_n] \end{aligned}$$

We call $[x|_n]$ the cylinder with base $x|_n \in A^n$. More generally, for a finite word $w \in A^{<\mathbb{N}}$, we define the cylinder $[w]$ with base w as:

In general, all open balls are cylinders and all cylinders are open balls.

$$[w] := \{y \in A^{\mathbb{N}} : y \geq w\} = \{y \in A^{\mathbb{N}} : y|_{\ln(w)} = w\}$$

Note that every ball is an open ball as well as a closed ball. Moreover, the centres of these balls are not unique: you could pick any other element in the ball, which would give you an equivalent ball. Thus every open set is a union of cylinders, hence the cylinders form a countable basis for $A^{\mathbb{N}}$. When working with $A^{\mathbb{N}}$, we work with this basis. Cylinders are clopen, which means $A^{\mathbb{N}}$ is totally disconnected, in fact 0-dimensional.

Every ball is closed, since the complement of the ball (cylinder) will be a union of cylinders (which are all open), hence open.

Proof that cylinders form a countable basis of $2^{\mathbb{N}}$. Every open set can be written a union of cylinders. The set of all cylinders in countable. Hence the set of cylinders form a countable basis □

Proposition 1.1.2. $A^{\mathbb{N}}$ is compact $\iff$ A is finite.

Proof. HW. Uses König's lemma (look up) □

Caution 1.1.1. The statement (compact $\iff$ closed and bounded) does not hold in this space. In fact, the concept of bounded is meaningless, since the largest value d can take is 1.

$A^{\mathbb{N}}$ is a nice space because it is so disconnected that it behaves like a discrete space so that you can do combinations, but you can also do limits.

1.2 *Algebras and sigma algebras*

Definition 1.2.1 ((σ-)algebra)**.** Let X be a nonempty set. A collection $\mathcal{A} \subseteq \mathcal{P}(X)$ of subsets of X is called an algebra (resp. σ-algebra) if $\emptyset \in A$ and $\mathcal{A}$ is closed under complements and finite (resp. countable) unions (hence also under finite (resp. countable) intersections).

A is co-finite $\iff$ A^C is finite.

Example 1.2.1 (algebras and σ-algebras)**.**

(a) Let X be a set. The power set $\mathcal{P}(X)$ is a σ-algebra.

(b) The collection $\mathcal{A}$ of finite and co-finite sets is an algebra.

(c) In a metric/topological space, the collection of clopen sets is an algebra and we call it the algebra of clopen sets.

Proof ((b) is an algebra). The empty set is finite. Finite unions of finite sets are finite. Complements of finite sets are co-finite. □

Proof ((b) is not a σ-algebra). Consider the collection $\bigcup_{i\in\mathbb{N}}\{i\}$ which is neither finite nor cofinite as a counterexample, hence $\mathcal{A}$ is not closed under countable unions. □

Proof sketch ((c) is an algebra). Finite unions of closed sets are closed and finite intersections of open sets are open. □

(d) For a finite nonempty set A, the clopen sets in $A^{\mathbb{N}}$ are exactly the finite disjoint unions of cylinders (HW), where the finiteness comes from the compactness $A^{\mathbb{N}}$.

(e) A box in $\mathbb{R}^d$ is a set of the form $B = I_1 \times \ldots \times I_d$, where each I_i is a (potentially unbounded) interval. Also, a complement of a box B is a finite disjoint union of boxes, so the collection of finite disjoint unions of boxes is an algebra.

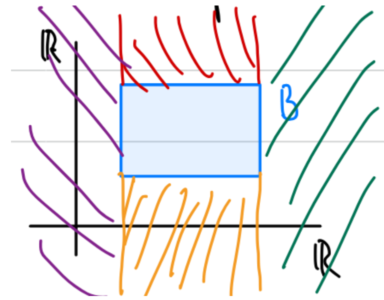


Figure 1.2: **(d)** The complement of B can be expressed as a finite disjoint union of boxes..

Interval means any type of interval, e.g. $(-7, \pi)$, $(0, 1]$, $[9, \infty)$, $(-\infty, \infty)$. Also, note than a finite union of boxes can be written as a finite disjoint union of some other boxes

Observation 1.2.2. *An arbitrary intersection of (σ-)algebras is a (σ-)algebra, i.e. if A_i, $(i \in I)$, are (σ-)algebras then $\bigcap_{i \in I} A_i$ is a (σ-)algebra.*

This observation allows us to define:

Definition 1.2.3 (Generated (σ-)algebra)**.** Let X be a set and $\mathcal{C} \subseteq \mathcal{P}(X)$. The ($\sigma$-)algebra generated by $\mathcal{C}$ is the smallest (σ-)algebra containing $\mathcal{C}$ ($\mathcal{P}(X)$ is a (σ-)algebra containing $\mathcal{C}$). Namely,

$$\text{algebra } \langle C \rangle := \bigcap_{\mathcal{A} \subseteq \mathcal{C}} \{\mathcal{A} : \mathcal{A} \subseteq \mathcal{P}(X) \text{ is an algebra and } \mathcal{A} \supseteq \mathcal{C}\}$$

$$\sigma\text{-algebra } \langle C \rangle_\sigma := \bigcap_{\mathcal{A} \subseteq \mathcal{C}} \{\mathcal{S} : \mathcal{S} \subseteq \mathcal{P}(X) \text{ is an } \sigma \text{ -algebra and } \mathcal{S} \supseteq \mathcal{C}\}$$

Definition 1.2.4 (Borel σ-algebra)**.** For a metric/topological space X, the σ-algebra $\mathcal{B}(X)$ generated by open sets is called the Borel σ-algebra and the sets in it are called Borel sets.

Proposition 1.2.1 (Algorithmic construction of generated (σ-)algebra)**.** Let X be a set, $\mathcal{C} \subseteq \mathcal{P}(X)$. Then:

(a) $\langle C \rangle = \bigcup_{n \in \mathbb{N}} \mathcal{C}_n$ where $\mathcal{C}_0 := \mathcal{C}$ and $\mathcal{C}_{n+1} := \{B^C : B \in \mathcal{C}_n\} \cup \{\bigcup_{i<k} B_i : B_i \in \mathcal{C}_n \text{ and } k \in \mathbb{N}\}$

(b) $\langle C \rangle_\sigma = \bigcup_{\alpha \in \omega_1} \mathcal{C}_\alpha$ where $\mathcal{C}_0 = \mathcal{C}$ and $\mathcal{C}_{\alpha+1} := \{B^C : B \in \mathcal{C}_\beta, \beta < \alpha\} \cup \{\bigcup_{i \in \mathbb{N}} : B_i \in \mathcal{C}_n\}$

ω_1 is the smallest uncountable cardinal.

Proof. (a) HW. (b) HW (optional). □

Observation 1.2.5. *In a metric/topological space X, for any countable basis $\mathcal{U}$, the σ-algebra generated by $\mathcal{U}$ is all Borel sets, i.e. $\mathcal{B}(X) = \langle \mathcal{U} \rangle_\sigma$*

Proof. Every open set O is a union of sets in $\mathcal{U}$, hence a countable (by assumption since $\mathcal{U}$ is countable) union of sets in $\mathcal{U}$, hence $O \in \langle \mathcal{U} \rangle_\sigma$, so $\langle \mathcal{U} \rangle_\sigma$ is a σ-algebra containing all open sets, hence $\mathcal{B}(X) \subseteq \langle \mathcal{U} \rangle_\sigma$. For the reverse inclusion, since $\mathcal{U}$ is a collection of open sets , $\langle \mathcal{U} \rangle_\sigma \subseteq \mathcal{B}(X)$ □

Basically, what we have showed is: (1) $\{$ open sets in $X\} \subseteq \mathcal{U} \implies \mathcal{B}(X) \subseteq \langle U \rangle_\sigma$ and (2) $\mathcal{U} \subseteq \{$ open sets in $X\} \implies \langle \mathcal{U} \rangle_\sigma \subseteq \mathcal{B}(X)$

1.3 Measures

Definition 1.3.1 (finitely/countably additive function)**.** Let X be a set and $\mathcal{C} \subseteq \mathcal{P}(X)$. A function $\mu : \mathcal{C} \to [0,\infty]$ is finitely (resp. countably) additive if:

$$\mu\left(\bigsqcup_{i<k} A_i\right) = \sum_{i<k} \mu\left(A_i\right) \quad \text{whenever } k \in \mathbb{N}, A_i \in \mathcal{C} \text{ and } \bigsqcup_{i<k} A_i \in \mathcal{C}$$

$$\mu\left(\bigsqcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu\left(A_i\right) \quad \text{whenever } k \in \mathbb{N}, A_i \in \mathcal{C} \text{ and } \bigsqcup_{i\in\mathbb{N}} A_i \in \mathcal{C}$$

Definition 1.3.2 (measure)**.** Let $(X, \mathcal{S})$ be a measurable space. A function $\mu : \mathcal{S} \to [0,\infty]$ is a measure on $(X, \mathcal{S})$ if:

(i) $\mu(\emptyset) = 0$

(ii) μ is countably additive

The triple $(X, \mathcal{S}, \mu)$ is called a measure space.

Caution 1.3.1. People also deal with finitely additive measures on algebras, but a finitely additive measure, even when defined on a σ-algebra, is in general not a measure since it may not by countable additive.

Definition 1.3.3 (probability/finite/σ-finite measure)**.** A measure μ on a measurable space $(X, \mathcal{S})$ is called:

(i) a probability measure when $\mu\left(X\right) = 1$

(ii) finite if $\mu(X) < \infty$

(iii) σ-finite if $\exists$ a finite collection $\{B_n\}$ if $X = \bigcup_n B_n, B_n \in \mathcal{S}$ and $\mu\left(B_n\right) < \infty$

Observation 1.3.4. *Let $(X, \mathcal{S})$ be a measurable space.*

(a) Any countable non-negative linear combination of measures on $(X, \mathcal{S})$ is a measure. Concretely, if μ_n are measures and $a_n \geq 0$, then $\sum_{n\in\mathbb{N}}^{a_n\mu_n}$ is a measure.

(b) Any convex combination of probability measures on $(X, \mathcal{S})$ is a probability measure. Concretely, if μ_n are probability measures and $a_n \geq 0$, $\sum a_n = 1$. then $\sum_{i=1}^{n} a_n\mu_n$ is a probability measure.

Example 1.3.1 (basic measures)**.**

(a) In any set X, the zero measure $\zeta : \mathcal{P}(X) \to \{0\}$ is a measure

(b) Let X be a nonempty set and fix $x_0 \in X$. Then the Dirac (delta) measure $\delta_{x_0} : \mathcal{P}(X) \to \{0,1\}$ defined:

$$\delta_{x_0}(B) := \begin{cases} 1 & x_0 \in A \\ 0 \text{ o.w.} & \end{cases}$$

is a measure

(c) Let X be a set. Then the counting measure $\chi : \mathcal{P}(X) \to [0, \infty]$ defined:

$$\chi(B) := \begin{cases} |B| \text{ when } B \text{ is finite} \\ \infty \text{ o.w.} \end{cases}$$

is a measure. Note that when X is countable, then $\chi = \sum_{x \in X} \delta_x$. We can also say that χ is finite when B is finite and χ is σ-finite when B is countable.

Take the countable collection of sets $\{\{x\} : x \in B\}$. Then $B = \bigcup_{x \in B}\{x\}, \chi(x) = 1 < \infty \forall x \in B$ (and trivially $\{x\} \in \mathcal{P}(X)$).

(d) Given a set X, define a measure μ on the σ-algebra of countable and co-countable subsets of X as:

$$\mu(B) = \begin{cases} 0 & \text{if } B \text{ is countable} \\ 1 & \text{if } B \text{ is co-countable} \end{cases}$$

Note that when X is countable, this is equal to the zero measure.

Definition 1.3.5 (atomic and atomless measures)**.** Let $(X, \mathcal{S}, \mu)$ be a measure space. A set $B \in \mathcal{S}$ is called an atom (or μ-atom) if $\mu(B) > 0$ and $\forall A \subset B, A \in \mathcal{S}$, either $\mu(A) = 0$ or $\mu(A) = \mu(B)$. The measure space is called

- atomic (also purely atomic) if any set in $\mathcal{S}$ with positive measure contains an atom
- atomless if there are no atoms.

Caution 1.3.2. The zero measure is both atomic and atomless. In general atomic and atomless are not mutually exclusive.

All measure (a) - (d) are atomic. To define interesting atomless measures, we need to first define them on an algebra $\mathscr{A}$ and then learn how to extend them to the σ-algebra generated by $\mathscr{A}$.

Definition 1.3.6 (finitely additive measure, premeasure)**.** Let $\mathscr{A}$ be an algebra on a set X. A function $\mu : \mathscr{A} \to [0, \infty)$ is called a finitely additive measure (resp. a countable additive measure or premeasure) if $\mu(\emptyset) = 0$ and μ is finitely (resp. countably) additive.

Lemma 1.3.1 (Disjointification trick)**.** For an algebra $\mathscr{A}$, any countable union $\bigcup_{n \in \mathbb{N}}$ of sets $A_n \in \mathscr{A}$ is equal to a countable disjoint union $\bigsqcup_{i \in \mathbb{N}} A'_n$ of sets $A'_n \in \mathscr{A}$ defined: $A'_0 := A_0, A'_n := A_n \setminus_{i<n} A_i$.

Proposition 1.3.1 (Properties of finitely additive measures)**.** Let μ be a finitely additive measure on an algebra $\mathscr{A}$ on a set X. Then:

(a) μ is monotone, i.e. if $A \subseteq B$ then $\mu(A) \leq \mu(B) \quad \forall A, B \in \mathscr{A}$.

(b) μ is countably superadditive, i.e.

$$\mu\left(\bigsqcup_{i\in\mathbb{N}} A_n\right) \geq \sum_{n\in\mathbb{N}} \mu(A_n)$$

for all $A_n \in \mathcal{A}$ such that $\bigsqcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$

(c) μ is finitely subadditive: $\mu\left(\bigsqcup_{n<N} A_n\right) \leq \sum_{n<N} \mu\left(A_n\right)$ for all $A_n \in \mathcal{A}$ such that $\bigsqcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$

(d) Moreover, if μ is a countably additive measure (i.e. a premeasure, then it is also countably subadditive)

When building measures, we show that it is subadditive, then argue that is countably subadditive, which will prove the countable additivity of the measure (since it is automatically superadditive by (b)).

If μ is a premeasure (i.e. countably additive, then from the same argument it holds that it is countably subadditive).

Proof.

(a) $\mu(B) = \mu(B \setminus A) + \mu(A) \geq \mu(A)$

(b) $\forall N \in \mathbb{N}$,

$$\begin{aligned}\mu\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) &= \mu\left(\bigsqcup_{n=1}^{N} A_n \sqcup \bigsqcup_{n=N+1}^{\infty} A_n\right) \\ &= \sum_{n\leq N} \mu(A_n) + \sum_{n=N+1}^{\infty} A_n \\ &\geq \sum_{n\geq N} \mu(A_n)\end{aligned}$$

Then in the limit as $N \to \infty$,

$$\mu\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) \geq \sum_{n\in\mathbb{N}} \mu(A_n)$$

(c) By disjointification , let A'_n be the disjointified A_n. Then by :

$$\mu\left(\bigcup_{n<N} A_n\right) = \mu\left(\bigsqcup_{n<N} A'_n\right) = \sum_{n<N} \mu(A'_n) \leq \sum_{n<N} \mu(A_n)$$

□

1.4 Construction of premeasures

1.4.1 Bernoulli premeasures

LET $X = 2^{\mathbb{N}}$ and $p \in (0,1)$. Any probability measure on 2 is of the form $\nu_p(1) = p, \nu_p(0) = 1 - p$. We will define a measure on the algebra of clopen sets of $2^{\mathbb{N}}$, which is the algebra of finite disjoint

$2 := \{0,1\}$

unions of cylinders (see Example 1.2.1 (d)). This measure μ_p will satisfy the property that $\mu_p([w]) = p^{\text{\# of 1s in w}}(1-p)^{\text{\# of 0s in w}}$. We first define μ_p on a cylinder $[w]$, $w \in 2^{<\mathbb{N}}$, by

$$\tilde{\mu}_p([w]) = p^{\text{\# of 1s in w}}(1-p)^{\text{\# of 0s in w}}$$

Then for each $B \in \mathcal{A}$, we "define"

"define" is in quotation marks because we still have to show that μ_p is well-defined. Specifically, we must show that is the same regardless of choice of representation as a finite disjoint unions of cylinders.

$$\mu_p(B) := \sum_{n<N} \tilde{\mu}([w_n])$$

where $B = \bigsqcup_{n<N}[w_n]$.

We need to show that this is well defined, i.e. doesn't depend on how B is written as a disjoint union of cylinders.

Claim (a)**.** $\tilde{\mu}_p$ is finitely additive on equal-length cylinders, i.e. for any cylinder $[w]$ and $n \in \mathbb{N}$,

$$\tilde{\mu}_p([w]) = \sum_{u \in 2^{\mathbb{N}}} \tilde{\mu}_p([wu])$$

Proof. By induction on n, it is enough to verify for $n = 1$.

$$\tilde{\mu}_p([w0]) + \tilde{\mu}_p([w1]) = \tilde{\mu}_p([w])(1-p) + \tilde{\mu}_p([w])p = \tilde{\mu}_p([w])$$

□

Claim (b)**.** Let $A \in \mathcal{A}$ and $\mathcal{P}_1, \mathcal{P}_2$ be two finite partition of A into cylinders. Then:

$$\sum_{P_1 \in \mathcal{P}_1} \tilde{\mu}_p(P_1) = \sum_{P_2 \in \mathcal{P}_2} \tilde{\mu}_p(P_2)$$

Definition 1.4.1 (common refinement)**.** If $\mathcal{P}$, $\mathcal{Q}$ are partitions of a set, then a common refinement $\mathcal{R}$ of $\mathcal{P}$ and $\mathcal{Q}$ is a collection of sets such that $\forall R \in \mathbb{R}$, $\exists P, Q \in \mathcal{P}, \mathcal{Q}$ such that $R \subseteq P \cap Q$.
Note that a natural common refinement can be obtained by constructing the set $\{P \cap Q : P \in \mathcal{P}, Q \in \mathcal{Q}\}$ (still finite if $\mathcal{P}, \mathcal{Q}$ are both finite).

Proof. Let $\mathcal{Q}$ be a common refinement of $\mathcal{P}_1, \mathcal{P}_2$ and take $\mathcal{Q}$ such that all cylinders in $\mathcal{Q}$ have the same length. Then:

Note that the intersection of two cylinders is either empty or another cylinder.

$$\begin{aligned}\sum_{P_1 \in \mathcal{P}_1} \tilde{\mu}_p(P_1) &= \sum_{P_1 \in \mathcal{P}_1} \sum_{Q \in \mathcal{Q}, Q \in P_1} \tilde{\mu}_p(Q) \\ &= \sum_{Q \in \mathcal{Q}} \tilde{\mu}_p(Q) \\ &= \sum_{P_2 \in \mathcal{P}_2} \sum_{Q \in \mathcal{Q}, Q \in P_2} \tilde{\mu}_p(Q) \\ &= \sum_{P_2 \in \mathcal{P}_2} \tilde{\mu}_p(P_2)\end{aligned}$$

□

Claim (b) shows the well-definedness of μ_p on $\mathcal{A}$. It also implies:

Corollary 1.4.1. μ_p is finitely additive

Proof. Almost immediate; HW. □

Claim (c)**.** μ_p is countably additive, i.e. a premeasure.

Proof. This is automatic by compactness. Let $\{U_n\} \subseteq \mathcal{A}$ and $U = \bigcup_{n\in\mathbb{N}} U_n \in \mathcal{A}$ (this is true by assumption, not by the fact that we are taking a countable union of sets that are closed in $\mathcal{A}$ since in general a countable union of closed sets is not necessarily closed.). By compactness (U is a closed subset of a compact set, hence compact), U can be written as a disjoint finite union of cylinders ($\{U_n\}$ is an open cover of U, hence there exists a finite subcover), so all but finitely many of these clopen sets have to be empty. Hence $\mu_p(A) = \sum_{n\in\mathbb{N}} \mu_p(U_n)$. □

This construction of the Bernoulli premeasure works for any $A^{\mathbb{N}}$ for A finite and nonempty and every probability measure on A.

1.4.2 *Lebesgue premeasure on* $\mathbb{R}^d$

ANALOGOUSLY TO Bernoulli measures on $A^{\mathbb{N}}$, we define a premeasure on the algebra $\mathcal{A}$ generated by boxes in $\mathbb{R}^d$. Note that the elements of $\mathcal{A}$ are finite disjoint unions of boxes, just like how the clopen sets in $A^{\mathbb{N}}$ for finite A are finite disjoint unions of cylinders.

We first define a premeasure on boxes:

$$\tilde{\lambda}\,(I_d \times \ldots \times I_d) := \mathrm{Ln}(I_1) \cdot \ldots \cdot \mathrm{Ln}(I_d)$$

We set $0 \cdot \infty = 0$ by convention. We then "define" the potential premeasure by

$$\lambda(A) = \sum_{B\in\mathcal{P}} \tilde{\lambda}(B)$$

where $\mathcal{P}$ is any partition of A into disjoint boxes. As with Bernoulli, we need to show that this is well defined, i.e. doesn't depend on the choice of $\mathcal{P}$.

Claim (a)**.** If $\mathcal{P}$ is a grid-partition of a box B, then $\tilde{\lambda}(B) = \sum_{P\in\mathcal{P}} \tilde{\lambda}(P)$

If $B = I_1, \ldots, _d \in \mathbb{R}^d$ is a box, a grid-partition of B is the Cartesian product of partitions of each interval. Symbolically, if $I_k = \bigsqcup_{j\in J_k} I_{k,j}$ for $1 \leq k \leq d$ and J_k some index set, then the grid partition $\mathcal{P}$ is defined $\mathcal{P} := \{P_{j_1,\ldots,j_d} : j_k \in J_k \forall 1 \leq k \leq d\}$ where $P_{j_1,\ldots,j_d} = I_{1,j_1} \times \ldots \times I_{d,j_d}$

Proof. This is true by definition for $d = 1$. For $d > 1$ apply induction using the distributivity law: assume the statement holds for dimensions up to $d-1$, then write $B = A \times I_d$ where $A = I_1 \times \ldots \times I_{d-1}$. Let $\mathcal{P}_A$ be the grid partition of A and $\mathcal{P}_d$ be the grid partition of I_d.

Then $\mathcal{P} = \{P_A \times J : P_A \in \mathcal{P}_A, J \in \mathcal{P}_d\}$. Then:

$$\begin{aligned}
\sum_{P \in \mathcal{P}} \tilde{\lambda}(P) &= \sum_{J \in \mathcal{P}_d} \sum_{P_A \in \mathcal{P}_A} \tilde{\lambda}(P_A \times J) \\
&= \sum_{J \in \mathcal{P}_d} \sum_{P_A \in \mathcal{P}_A} \tilde{\lambda}(P_A)\tilde{\lambda}(J) \\
&= \sum_{J \in \mathcal{P}_d} \tilde{\lambda}(J) \sum_{P_A \in \mathcal{P}_A} \tilde{\lambda}(P_A) \\
&= \sum_{J \in \mathcal{P}_d} \tilde{\lambda}(J)\tilde{\lambda}(A) \\
&= \tilde{\lambda}(A) \sum_{J \in \mathcal{P}_d} \tilde{\lambda}(J) \\
&= \tilde{\lambda}(A)\tilde{\lambda}(I_d) \\
&= \tilde{\lambda}(A \times I_d) \\
&= \tilde{\lambda}(B)
\end{aligned}$$

□

Claim (b)**.** If $\mathcal{P}_1$ and $\mathcal{P}_2$ are two finite partitions of $A \in \mathcal{A}$ into boxes, then

$$\sum_{P_1 \in \mathcal{P}_1} \tilde{\lambda}(P_1) = \sum_{P_2 \in \mathcal{P}_2} \tilde{\lambda}(P_2)$$

Proof. Take a grid-partition of A that is a common refinement of $\mathcal{P}_1$ and $\mathcal{P}_2$. Details in HW. □

Corollary 1.4.2. λ is well-defined on $\mathcal{A}$ and finitely additive.

Claim (c)**.** λ is countably additive on $\mathcal{A}$.

Proof. Because a finitely additive measure is always countable superadditive , it suffices to prove countable subadditivity. We prove this in the case where a bounded box B is written as a countable disjoint union of boxes, i.e. $B = \bigsqcup_{n \in \mathbb{N}} B_n$. The general case for this follows easily and is left as HW.

In the case of cylinders and $2^{\mathbb{N}}$, we used that B is compact and the B_n are open, but for boxes neither is true in general. However, we can approximate B as closed and B_n by open boxes.

For $\varepsilon > 0$.

Notation. For $a, b \in \mathbb{R}$, we write $a \approx_\varepsilon b$ if $|a - b| \leq \varepsilon$.

Let $B' \subseteq B$ be a closed box s.t. $\lambda(B') \approx_{\varepsilon/2} \lambda(B)$. For each $n \in \mathbb{N}$, let $\tilde{B}_n \supseteq B_n$ be an open box s.t. $\lambda(\tilde{B}_n) \approx_{\varepsilon/2 \cdot 2^{-n}} \lambda(B_n)$. Then $\{\tilde{B}_n\}_{n \in \mathbb{N}}$ is an open cover of the compact set B', so there is a finite subcover $\{\tilde{B}_n\}_{n \leq N}$.

Then,

$$\lambda(B) \approx_{\varepsilon/2} \lambda(B') \leq \lambda\left(\bigcup_{n<N} \tilde{B}_n\right) \leq \sum_{n<N} \lambda(\tilde{B}_n) \leq \sum_{n \in \mathbb{N}} \lambda(\tilde{B}_n) \approx_{\varepsilon/2} \sum_{n \in \mathbb{N}} \lambda(B_n)$$

Hence,

$$\lambda(B) \leq \sum_{n \in \mathbb{N}} \lambda(B_n) + \varepsilon$$

which implies, since ε was arbitrary,

$$\lambda(B) \leq \sum_{n \in \mathbb{N}} \lambda(B_n)$$

□

Thus we call this premeasure λ on $\mathcal{A}$ the Lesbegue premeasure.

1.5 Carathéodory extension

To DEFINE measures, we always define a premeasure on some algebra and apply the following:

Theorem 1.5.1 (Carathéodory)**.** Let X a set, $\mathcal{A}$ an algebra on X and $\mu : \mathcal{A} \to [0, \infty]$ a premeasure. Then μ admits an extension to a measure on the σ-algebra $\langle \mathcal{A} \rangle_\sigma$. This measure is unique if the premeasure μ is σ-finite on $\mathcal{A}$, and in this case it is equal to the outer measure μ^* of μ, and we abuse the notation and denote this extension by μ.

To prove this, we need the following notion:

Definition 1.5.1 (Outer measure)**.** Let $\mathcal{A}$ be an algebra on X and μ a premeasure on $\mathcal{A}$. The outer measure of μ is the function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ defined by:

$$\mu^*(S) := \inf \left\{ \sum_{n \in \mathbb{N}} \mu(A_n) : \bigcup_{n \in \mathbb{N}} A_n \supseteq S \text{ and } \{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A} \right\}$$

Proposition 1.5.1 (Properties of outer measure)**.**

(a) (monotone): If $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$

(b) (countably subadditive): For $S, S_n \subseteq X$, $\mu^*\left(\bigcup_{n \in \mathbb{N}} S_n\right) \leq \sum_{n \in \mathbb{N}} \mu^*(S_n)$

Proof.

(a) Follows from the definition of μ^* because a cover of B is also a cover of A.

(b) Follows from the fact that a union of covers of the S_n is a cover of S.

□

Lemma 1.5.1. For any premeasure μ on an algebra $\mathcal{A}$, the outer measure μ^* on $\mathcal{A}$ is equal to μ, i.e.

$$\mu^*|_{\mathcal{A}} = \mu$$

Proof. Let $A \in \mathcal{A}$ and $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ be a cover of A. By disjointification , we may assume the A_n are disjoint. By replacing A_n with $A_n \cap A$ we may also assume $A = \bigsqcup_{n\in\mathbb{N}} A_n$. By countable additivity of μ we have $\mu(A) = \sum_{n\in\mathbb{N}} \mu(A_n)$, so even with the original A_n (which is larger than the new disjointified and intersected with A A_n's), we had that $\mu(A) \leq \sum_{n\in\mathbb{N}} \mu(A_n)$ by monotonicity. □

The direction $\mu^* \leq \mu$ is trivial: for any $A \in \mathcal{A}$, take the cover $\{A\}$ of A, then $\mu^*(A) \leq \mu(A)$.

Theorem 1.5.2 (Carathéodory's theorem: existence)**.** Let μ a premeasure on the algebra $\mathcal{A}$ on a set X. We want to extend μ to be a measure on $\langle \mathcal{A} \rangle_\sigma$. In fact, μ^* is such an extension.

Proof. To show that μ^* is countably additive on $\langle \mathcal{A} \rangle_\sigma$, it is enough to show that it is finitely additive because outer measures are countably additive and finite additivity implies countable superadditivity . See the following proofs for details. □

1.6 *Carathéodory's proof*

Carathéodory's proof. A set $B \subseteq X$ conserves a set $S \subseteq X$ if $\mu^*(S) = \mu^*(B \cap S) + \mu^*(B^C \cap S)$. Call B conservative if it conserves every set. Let $\mathcal{C}$ denote the collection of all conservative sets. Then we prove:

(i) $\mathcal{A} \subseteq \mathcal{C}$

(ii) $\mathcal{C}$ is a σ-algebra

(iii) μ^* is finitely additive on $\mathcal{C}$

Each step is left as homework. (i) follows from the proof that $\mu^* = \mu$. (ii) implies that $\langle \mathcal{A} \rangle_\sigma$. (iii) Follows almost by definition of $\mathcal{C}$ □

By subadditivity "$\mu^*(S) = \mu^*(B \cap S) + \mu^*(B^C \cap S)$" can only fail in the $<$ direction.
Why do we want the collection $\mathcal{C}$?
It seems counter intuitive to need to conserve **every** set since the sets we really care about are the Borel sets.
The reason behind it is similar to when doing an induction proof where you assume a stronger induction hypothesis than what you need prove so that you can "pull over" the induction step with a stronger induction hypothesis

1.7 *Terence Tao's proof*

Definition 1.7.1 (pseudometric)**.** Let X a set and d a function $d : X \times X \to [0, \infty)$. Then d is a pseudometric if $\forall x, y, z \in X$:

(i) $d(x, x) = 0$

(ii) $d(x, y) = d(y, x)$

(iii) $d(x, z) \leq d(x, y) + d(y, z)$

Pseudometrics are basically a metric but the axiom $d(x, y) = 0 \iff x = y$ might not hold. The intuition related to metrics still hold for pseudometrics: Cauchy, convergence, . . . , these notions still exist when working with pseudometrics. In fact, $d(\cdot, \cdot) = 0$ forms an equivalence relation and we can "mod out" by the equivalence relation to get a real metric space.

Tao's proof. We define a pseudo-metric $d_{\mu^*} : \mathcal{P}(X) \times \mathcal{P}(X) \to [0, \mu(X)]$ by

$$d_{\mu^*}(A, B) := \mu^*(A \triangle B)$$

Tao's proof only works for σ-finite measures. In the notes we proof the case for $\mathcal{M}$ finite. Then it is possible to deduce the σ-finite case from this, which is left as HW.
The secret of symmetric differences: $\mathcal{P}(X)$ with $\triangle$ is an abelian group with $\emptyset$ as the identity element and $A^{-1} = A$, hence it is order 2.

Claim (a)**.** d_{μ^*} is a pseudometric.

Proof.

(i) By definition $d_{\mu^*}(A, A) = \mu^*(A \triangle A) = \mu^*(\emptyset) = 0$

(ii) d_{μ^*} is symmetric by definition.

(iii) For the triangle inequality, let $A, B, C \in \mathcal{P}(X)$ and observe:

$$A \triangle C = (A \triangle B) \triangle (B \triangle C) \subseteq (A \triangle B) \cup (B \triangle C)$$

Hence,

$$\begin{aligned} \mu^* (A \triangle C) &\leq \mu^* ((A \triangle B) \cup (B \triangle C)) \\ &\leq \mu^* (A \triangle B) + \mu^* (B \triangle C) \\ &= d_{\mu^*}(A, B) + d_{\mu^*}(B, C) \end{aligned}$$

□

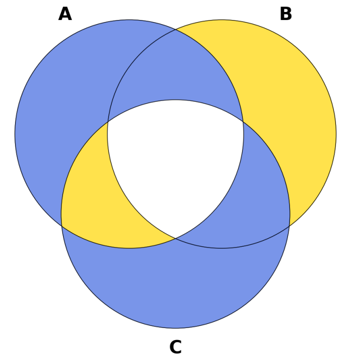


Figure 1.3: $A \triangle C$: blue region, and $(A \triangle B) \cup (B \triangle C)$ is the blue and yellow regions.

Let $\mathcal{M} := \overline{\mathcal{A}}^{d_{\mu^*}}$ be the closure of $\mathcal{A}$ inside $\mathcal{P}(X)$ with the pseudometric d_{μ^*}. We will show that $\mathcal{M}$ is a σ-algebra, (hence $\langle \mathcal{A} \rangle_\sigma \subseteq \mathcal{M}$ and μ^* is finitely additive on $\mathcal{M}$)

Thinking of sets as points might be weird, but get over it. Sets can be identified by the identity function $\mathbb{1}$ and we have already dealt with $C([0,1])$ which is weirder.

Claim (b)**.** The function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is continuous w.r.t d_{μ^*}. In fact, it is 1-Lipschitz

Proof. Observe that $\mu^*(A) = \mu^*(A \triangle \emptyset) = d_{\mu^*}(A, \emptyset)$, so by the $\triangle$ inequality,

$$|\mu^*(B) - \mu^*(A)| = \left| d_{\mu^*}(A, \emptyset) - d_{\mu^*}(B, \emptyset) \right| \leq d_{\mu^*}(A, B)$$

□

An isometry is a distance-preserving transformation (between metric spaces)

Claim (c)**.** The function $\mathcal{P}(X) \to \mathcal{P}(X)$, $A \mapsto A^C$ is continuous, in fact an isometry.

Proof. Observe that $A \triangle B = A^C \triangle B^C$, so

$$d_{\mu^*}(A, B) = \mu^*(A \triangle B) = \mu^*(A^C \triangle B^C) = d_{\mu^*}(A^C, B^C)$$

□

This implies that $\mathcal{M}$ is closed under complements: if $M \in \mathcal{M}$, then $\exists (A_n) \in \mathcal{A}$ with $\lim_{n\to\infty} A_n = M$. Moreover, by continuity, $\lim_{n\to\infty} A_n^C = M^C \in \mathcal{M}$.

Claim (d)**.** The function $\mathcal{P}(X) \times \mathcal{P}(X) \to \mathcal{P}(X)$, $(A, B) \mapsto A \cup B$ is continuous w.r.t the $d_{\mu^*} + d_{\mu^*}$ pseudometric on $\mathcal{P}(X) \times \mathcal{P}(X)$, in fact it is 1-Lipschitz

Proof.

$$\begin{aligned} d_{\mu^*}(A_1 \cup B_1, A_2 \cup B_2) &= \mu^* ((A_1 \cup B_1) \triangle (A_2 \cup B_2)) \\ &\leq \mu^* ((A_1 \triangle A_2) \cup (B_1 \triangle B_2)) \\ &\leq \mu^*(A_1 \triangle A_2) + \mu^*(B_1 \triangle B_2) \\ &= (d_{\mu^*} + d_{\mu^*})((A_1, A_2), (B_1, B_2)) \end{aligned}$$

□

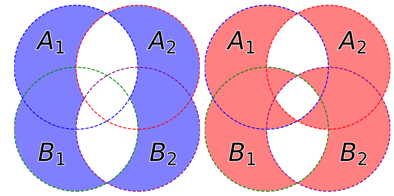


Figure 1.4: **Left**: $(A_1 \cup B_1) \triangle (A_2 \cup B_2)$ **Right**: $(A_1 \triangle A_2) \cup (B_1 \triangle B_2)$.

This implies that $\mathcal{M}$ is closed under finite unions, hence $\mathcal{M}$ is an algebra: If $A, B \in \mathcal{M}$, then $\exists \{A_n\}, \{B_n\} \subseteq \mathcal{M}$ such that $A_n \to A$ and $B_n \to B$ so by continuity, $\lim_{n\to\infty}(A_n \cup B_n) = A \cup B$.

Claim (e)**.** μ^* is finitely additive on $\mathcal{M}$

Proof. Let $A, B \in \mathcal{M}$ be disjoint. There exists sequences $(A_n), (B_n)$ in $\mathcal{M}$ such that $A_n \to A$ and $B_n \to B$. Recognize that

- $A_n \cup B_n \to A \sqcup B$ (by continuity of unions)
- $A_n \cap B_n \to A \cap B = \emptyset$ (by continuity of intersections)
- $\mu^*(A_n) \to \mu^*(A), \mu^*(B_n) \to \mu^*(B)$, and $\mu^*(A_n \cup B_n) \to \mu^*(A \sqcup B)$ (by the continuity of μ^*)

Also, we know $\mu^*|_{\mathcal{A}} = \mu$. Since μ is finitely additive, $\mu^*(A_n \cup B_n) = \mu^*(A_n) + \mu^*(B_n) - \mu^*(A_n \cap B_n)$. Hence,

$$\begin{aligned} \mu^*(A \sqcup B) &= \lim_{n\to\infty} \mu^*(A_n \cup B_n) \\ &= \lim_{n\to\infty} (\mu^*(A_n) + \mu^*(B_n) - \mu^*(A_n \cap B_n)) \\ &= \mu^*(A) + \mu^*(B) \end{aligned}$$

□

Now μ^* is a finitely additive measure on $\mathcal{M}$. We just need to show that $\mathcal{M}$ is a σ-algebra.

Claim (f)**.** $\mathcal{M}$ contains all countable unions of sets in $\mathcal{M}$

Proof. Let $(A_n) \subseteq \mathcal{A}$. By disjointification , we may assume the A_n are pairwise disjoint. It's enough to show that

Up until now, we haven't used that μ^* is finite. Now we will use it.

$$\lim_{n\to\infty} \left(\bigsqcup_{k \le n} A_k \right) = \bigsqcup_{k \in \mathbb{N}} A_k = A$$

Observe that:

$$\begin{aligned} d_{\mu^*}\left(\bigsqcup_{k\le n} A_n, A \right) &= \mu^* \left(\bigsqcup_{k \le n} A_k \triangle A \right) \\ &= \mu^* \left(\bigsqcup_{k>n} A_k \right) \\ &\le \sum_{k>n} \mu^*(A_k) \end{aligned}$$

Then $\sum_{k\in\mathbb{N}} \mu^*(A_k)$ converges: indeed, $\forall n \in \mathbb{N}, \sum_{k\le n} \mu^*(A_k) = \sum_{k\le n} \mu(A_k) = \mu\left(\bigsqcup_{k\le n} A_k\right) \le \mu(X) < \infty$

This shows that $\mathcal{M}$ is closed under countable unions, hence a σ-algebra. Indeed, if $(M_n) \subseteq \mathcal{M}$, let $\varepsilon > 0$, and take $A_n \in \mathcal{A}$ so that

$A_n \approx_{\varepsilon \cdot 2^{-n}} M_n$. Then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$ and

$$d_{\mu^*}\left(\bigcup_{n\in\mathbb{N}} A_n, \bigcup_{n\in\mathbb{N}} M_n\right) \leq \varepsilon$$

So $\bigcup_{n\in\mathbb{N}} M_n$ is ε-close to an element of $\mathbb{M}$. But ε is arbitrary and $\mathcal{M}$ is closed, hence $\bigcup_{n\in\mathbb{N}} M_n \in \mathcal{M}$. □

□

Theorem 1.7.1 (Carathéodory's extension: uniqueness)**.** Let $\mathcal{A}$ be an algebra on X and μ a premeasure on $\mathcal{A}$. Then for each extension ν of μ to a measure on $\langle\mathcal{A}\rangle_\sigma$, we have $\nu \leq \mu^*$.

Proof. Since μ^* is defined as an infimum over covers by sets in $\mathcal{A}$, we fix a set $S \in \langle\mathcal{A}\rangle_\sigma$ and a cover $\{A_n\} \subseteq \mathcal{A}$ of S . Then

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

Hence $\nu \leq \mu$. Now assume that μ is σ-finite. It's enough to prove $\nu = \mu^*$ assuming μ is finite because given a partition $\{X_n\}$ of X such that $X_n \in \mathcal{A}$ and $\mu(X_n) < \infty$, then $\forall S \in \langle\mathcal{A}\rangle_\sigma$,

$$\begin{aligned} \nu(S) &= \nu\left(\bigsqcup_{n\in\mathbb{N}} S \cap X_n\right) \\ &= \sum_{n\in\mathbb{N}} \nu(S\cap X_n) \\ &= \sum_{n\in\mathbb{N}} \mu^*(S\cap X_n) \\ &= \mu^*\left(\bigsqcup_{n\in\mathbb{N}} S \cap X_n\right) \\ &= \mu^*(S) \end{aligned}$$

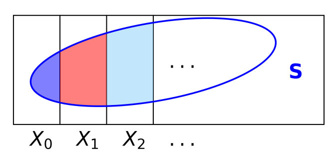


Figure 1.5: Each part $S \cap X_i$ has finite measure.s.

Hence suppose μ is finite. We show that the function $\nu|_{\langle\mathcal{A}\rangle}$ is continuous w.r.t the pseudometric d_{μ^*}. In fact, it is 1-Lipschitz:

$$\begin{aligned} |\nu(S_1) - \nu(S_2)| &\leq \nu(S_1 \setminus S_2) + \nu(S_1 \setminus S_2) \\ &= \mu(S_1 \triangle S_2) \\ &\leq \mu^*(S_1 \triangle S_2) \\ &= d_{\mu^*}(S_1, S_2) \end{aligned}$$

So ν and μ^* are continuous functions that coincide on $\langle\mathcal{A}\rangle$. Hence, they coincide on $\langle\mathcal{A}\rangle_\sigma$ since $\langle\mathcal{A}\rangle$ is dense in $\langle\mathcal{A}\rangle_\sigma$ w.r.t. d_{μ^*} since $\langle\mathcal{A}\rangle_\sigma = \overline{\langle\mathcal{A}\rangle}$. Thus, $\nu = \mu^*$ everywhere in $\langle\mathcal{A}\rangle_\sigma$.

If two continuous functions coincide on $\mathcal{Q}$, they coincide everywhere (MATH 255). Analogously for us, we have that $\mathcal{A}$ is $\mathbb{Q}$ and $\langle\mathcal{A}\rangle_\sigma$ is "everywhere".

Thus there are unique measures extending the Bernoulli and Lebesgue premeasures, and we call them the Bernoulli and Lebesgue measure. For Bernoulli measures, we mean the premeasure obtained on clopen sets for $A^{\mathbb{N}}$ for finite A and probability measure m on A. The corresponding Bernoulli measure is denoted by $m^{\mathbb{N}}$. □

Definition 1.7.2 (Borel measure). For a metric/topological space X, a Borel measure is any measure defined on the Borel σ-algebra $\mathbb{B}(X)$.

Example 1.7.1. Lebesgue and Bernoulli measures are Borel measures. Any Dirac measure is also Borel.

1.8 *Counterexamples of uniqueness of extensions*

Let $\mathcal{A}$ be the algebra generated by intervals of the form $[a, b)$ for $a \leq b$. Note that $\mathcal{A}$ consists of disjoint unions of intervals of this form. Define μ on $\mathcal{A}$ by

$$\mu(S) = \begin{cases} 0 & S = \emptyset \\ \infty & \text{o.w.} \end{cases}$$

Then $\langle \mathcal{A} \rangle_\sigma = \mathbb{B}(\mathbb{R})$ since countable unions of these intervals give any open interval. The outer measure μ^* on $\mathbb{B}(\mathbb{R})$ is

$$\mu^*(B) = \begin{cases} 0 & B = \emptyset \\ \infty & \text{o.w.} \end{cases}$$

The counting measure ν_C on $\mathbb{B}(\mathbb{R})$ is also an extension, but is is not equal to μ^*:

$$1 = \nu_C(0) < \mu^*(\{0\}) = \infty$$

extension in the sense that the domain of μ is expanded to more sets, and $\mu|_{\mathcal{A}} = \nu_C|_{\mathcal{A}}$

Another extension is the measure ν defined on $\mathbb{B}(\mathbb{R})$ defined by:

$$\nu(B) = \begin{cases} 0 & B \text{ countable} \\ \infty & \text{o.w.} \end{cases}$$

We have that $\nu \leq \nu_C \leq \mu^*$ and $0 = \nu(\{0\}) < 1 = \nu_C(\{0\}) < \infty = \mu^*(\{0\})$

1.9 *Null and measurable sets*

Definition 1.9.1 (μ-null). Let $(X, \mathcal{S}, \mu)$ be a measure space. A set $A \subseteq X$ is called μ-null (null) if there is $B \in \mathcal{S}$ such that $A \subseteq B$ and $\mu(B) = 0$. Denote the family of all μ-null sets by Null_μ.

Note that null sets are subsets of X, i.e. they don't necessarily need to be in $\mathcal{S}$. The intuition is that we want to extend μ to other sets not in $\mathcal{S}$ such that it has unique values (0).

Observation 1.9.2. *Null sets are closed downward under subsets, i.e. if A is μ-null then $\mathcal{P}(A)$ consists of μ-null sets.*

Observation 1.9.3. *μ-null sets form a σ-ideal, i,e, they are closed downward and under countable unions.*

In words, $=_\mu$ is the equivalence relation of being equal up to null sets.

Notation 1.9.1. *We write $A =_\mu B$ if $\mu(A \triangle B) = 0$.*

Definition 1.9.4 (μ-measurable)**.** For sets $A, B \subseteq X$, write $A =_\mu B$ if $A \triangle B$ is μ-null. Call a set $X \subseteq X$ μ-measurable (measurable) if $A =_\mu B$ for some $B \in \mathcal{B}$. Denote the family of all μ-measurable sets Meas_μ.

Observation 1.9.5. Meas_μ *is a σ-algebra. In fact,* $\text{Meas}_\mu = \langle \mathcal{B} \cup \text{Null}_\mu \rangle_\sigma$.

Proof. It is clear that $\emptyset \in \text{Meas}_\mu$. For complements, note that if $A \in \text{Meas}_\mu$, let $A =_\mu B$ for some $B \in \mathcal{B}$. Then $A \triangle B = A^C \triangle B^C$, hence $A \triangle B$ is null $\iff$ $A^C \triangle B^C$ is null. Hence since $B^C \in \mathcal{B}$, $A^C \in \text{Meas}_\mu$.

For countable unions, if $\{A_n\} \subseteq \text{Meas}_\mu$, then there exist sets $\{B_n\} \subseteq \mathcal{B}$ such that $A_n \triangle B_n$ is null for each $n \in \mathbb{N}$. Then:

$$\left(\bigcup_{n \in \mathbb{N}} A_n \right) \triangle \left(\bigcup_{n \in \mathbb{N}} B_n \right) \subseteq \bigcup_{n \in \mathbb{N}} A_n \triangle B_n$$

hence $\left(\bigcup_{n \in \mathbb{N}} A_n\right) \triangle \left(\bigcup_{n \in \mathbb{N}} B_n\right)$ is null so $\bigcup_{n \in \mathbb{N}} A_n \in \text{Meas}_\mu$.

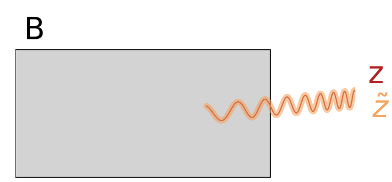
Hence Meas_μ is a σ-algebra. Moreover, $\text{Meas}_\mu \supseteq (\mathcal{B} \cup \text{Null}_\mu)$ so $\text{Meas}_\mu \supseteq \langle \mathcal{B} \cup \text{Null}_\mu \rangle_\sigma$. The other inclusion follows by the definition of μ-measurable sets.

$\square$

Remark 1.9.6. *It is in HW to show that* Meas_μ *is what we've obtained in both Carathéodory's and Tao's proofs of Carathéodory extension.*

Proposition 1.9.1. Let $(X, \mathcal{B}, \mu)$ be a measure space. Then:

$$\begin{aligned} \text{Meas}_\mu &= \{B \setminus Z : B \in \mathcal{B} \text{ and } Z \text{ is } \mu\text{-null}\} \\ &= \{B \sqcup Z : B \in \mathcal{B} \text{ and } Z \text{ is } \mu\text{-null}\} \end{aligned}$$



Proof. Since $B \setminus Z$ and $B \cup Z$ are μ-measurable $\supseteq$ is clear and, it is enough to show $\subseteq$ i.e. that every μ-measurable set is of these two forms.

Let M be a μ-measurable set. Then $\exists B \in \mathcal{B}$ such that $Z := M \triangle B$ is μ-null. Thus, $M = B \triangle Z$. Let $\tilde{Z} \supseteq Z$ be in $\mathcal{B}$ such that $\mu(\tilde{Z}) = 0$. Let $B' := B \setminus \tilde{Z}$ and $\tilde{B} := B \cup \tilde{Z}$. Then:

$$\underbrace{B'}_{\in \mathcal{B}} \sqcup \underbrace{(B \cap (\tilde{Z} \setminus Z)) \sqcup (B^C \cap Z)}_{= M \setminus B' \in \text{Null}_\mu} = M = \underbrace{\tilde{B}}_{\in \mathcal{B}} \setminus \underbrace{(B \cap Z) \setminus (B^C \cap (\tilde{Z} \setminus Z))}_{= \tilde{B} \setminus B \in \text{Null}_\mu}$$

$\square$

Corollary 1.9.1. For any μ-measurable set M, there are $B_0, B_1 \in \mathcal{B}$ such that $B_0 \subseteq M \subseteq B_1$ and $\mu(B_0) = \mu(M) = \mu(B_1)$.

This is useful because sometimes in proofs, you want a larger or smaller Borel set with the same measure, since they may be nicer to deal with.

Most textbooks will define particular implements, (e.g. with interval coverings) of this, but they are just examples of this abstract idea.

Definition 1.9.7 (Completeness (of measure space))**.** A measure space $(X, \mathcal{B}, \mu)$ is called complete iff $\mathcal{B} = \text{Meas}_\mu$.

Proposition 1.9.2 (Completion)**.** Every measure μ on a measurable space $(X, \mathcal{B})$ admits a unique completion, i.e. a unique extension to a measure on Meas_μ.

Proof. Existence: Let M be a μ-measurable set, so $M = B \sqcup Z$ for $B \in \mathcal{B}$ and Z μ-null . Define $\bar{\mu}(M) := \mu(B)$. To show this is well defined, let $M = \tilde{B} \sqcup \tilde{Z}$ for $\tilde{B} \in \mathcal{B}$ and $\tilde{Z}$ μ-null. Then $B \triangle \tilde{B} \subseteq Z \cup \tilde{Z}$ which is μ-null, hence $\mu(B) = \mu(\tilde{B})$.

Uniqueness: If ν is another extension and M is μ-measurable, then $M = B \sqcup Z$ where $B \in \mathcal{B}$, Z μ-null, so by :

$$\nu(M) = \nu(B) + \nu(Z) = \nu(B) = \mu(B) = \bar{\mu}(B \sqcup Z) = \bar{\mu}(M)$$

□

Remark 1.9.8. *There are typically many more sets in* Meas_μ *than in* $\mathcal{B}$*. For example, if* X *is a second countable metric/topological space, then it has at most* $|\mathbb{R}| = |2^{\mathbb{N}}|$ *(a.k.a continuum-many) Borel sets. Meanwhile, there are* $|2^{\mathbb{R}}|$ *many measurable sets.*

1.10 Nonmeasurable sets

We will give an example of a non-Lebesgue measurable subset of $\mathbb{R}$.

Definition 1.10.1 (transversal, selector)**.** Let E be an equivalence relation on a set X. A transversal (a.k.a a set of representatives) for E is a set $Y \subseteq X$ which meets each E class in exactly one point. A selector for E is a map $s : X \to X$ such that $s(x) = [x]_E$ and $xEy \iff s(x) = s(y) \quad \forall x, y \in X$. For a selector s, we can get a transversal $Y := s(X)$ and vice-versa, from a transversal Y we can get a selector s by $s(x) =$ the unique $y \in Y \cap [x]_E$. Selectors and transversals exist by the axiom of choice, but this results (typically) in ill-behaved(i.e. nonmeasurable) functions and sets.

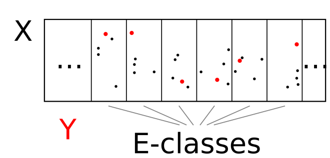


Figure 1.6: The set Y is a transversal of E..

$E_{\mathbb{Q}}$ is also called the Vitali equivalence relation on $\mathbb{R}$.

Example 1.10.1. Let $E_{\mathbb{Q}}$ be the coset equivalence relation of $\mathbb{Q}$ as a subgroup of $\mathbb{R}$ under addition, i.e. $xE_{\mathbb{Q}}y \iff y - x \in \mathbb{Q}$. This is also the orbit equivalence relation of the action of $\mathbb{Q}$ on $\mathbb{R}$ by translation. For each $x \in \mathbb{R}$, the equivalence class $[x]_{E_{\mathbb{Q}}} = x + \mathbb{Q}$. In particular, it intersects $[0, 1]$

Claim. Any transversal $Y \subseteq [0, 1]$ of $E_{\mathbb{Q}}$ is non-measurable w.r.t the Lebesgue measure λ.

Proof. Observe that:

$$[0, 1] \subseteq \bigsqcup_{q \in \mathbb{Q} \cap [-1,1]} q + Y \subseteq [-1, 2]$$

Hence, if Y was measurable, then it's translation $q + Y$ would also be measurable. Hence,

$$\begin{aligned} 1 = \lambda([0,1]) &\leq \lambda\left(\bigsqcup_{q\in\mathbb{Q}\cap[-1,1]} q+Y\right) \\ &= \sum_{q\in\mathbb{Q}\cap[-1,1]} \lambda(q+Y) = \infty\cdot\lambda(Y) \\ &\leq \lambda([-1,2]) = 3 \end{aligned}$$

which is a contradiction. □

Remark 1.10.2. *It is tempting to think that nonmeasurable sets can only arise from Axiom of Choice. This has some truth to it but not entirely. Indeed, in Solovay's model of ZF (Zermelo-Fraenkl set theory without Choice) where Axiom of Choice fails badly, all subsets of $\mathbb{R}$ are Lebesgue measurable. On the other hand, there are simple constructions of subsets of $\mathbb{R}$ without Choice, which yield non-measurable sets in some other models of ZF. More concretely, there is a G_δ (countable intersection of open sets) subset B of $\mathbb{R}^3$ such that whether or not the set $\operatorname{proj}_{\mathbb{R}}(\mathbb{R}^3 \setminus \operatorname{proj}_{\mathbb{R}^2}(B))$ is measurable is independent of ZFC!!! This is to say measurability is a subtle property, and even the measurability of projections of Borel sets (called analytic sets) is a difficult theorem.*

1.11 Pocket tools for working with measures

Proposition 1.11.1 (monotone convergence of measure)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space. Then:

(a) Let (A_n) be an increasing sequence of μ-measurable sets, i.e. $\forall n \in \mathbb{N}$, $A_n \subseteq A_{n+1}$. Then:

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

(b) Let (B_n) be a decreasing sequence μ-measurable sets, i.e. $\forall n \in \mathbb{N}$, $B_n \supseteq B_{n+1}$, where $\mu(B_0) < \infty$. Then:

$$\mu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \lim_{n\to\infty} \mu(B_n)$$

Caution 1.11.1. If $\forall n \in \mathbb{N}, \mu(B_n) = \infty$, then B may not hold. For example, take $B_n := (n, \infty)$. Then B_n is decreasing. Then $\bigcap_{n\in\mathbb{N}} B_n = \emptyset$ so it is null, but $\lim_{n\to\infty} \lambda(B_n) = \infty$

Proof. (a) Let A_n' be the disjointified A_n's, i.e. $A_1' = A_1$ and $A_n' =$

$A_n \setminus \bigcup_{k<n} A_k$. Then $\bigcup_{n\in\mathbb{N}} A_n = \bigsqcup_{n\in\mathbb{N}} A'_n$. Hence:

$$\begin{aligned}\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \mu\left(\bigsqcup_{n\in\mathbb{N}} A'_n\right)\\ &= \sum_{n\in\mathbb{N}} \mu(A'_n)\\ &= \lim_{N\to\infty} \sum_{n\le N} \mu(A'_n)\\ &= \lim_{N\to\infty} \mu\left(\bigsqcup_{n\le N} \mu(A'_n)\right)\\ &= \lim_{N\to\infty} \mu(A_N)\end{aligned}$$

(b) Then sets $A_n := B_0 \setminus B_n$ are increasing, so,

$$\begin{aligned}\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \lim_{n\to\infty} \mu(A_n)\\ &= \lim_{n\to\infty} \mu(B_0 \setminus B_n)\\ &= \mu\left(\lim_{n\to\infty}(B_0 \setminus B_n)\right) && \text{[(a)]}\\ &= \mu(B_0) - \lim_{n\to\infty} \mu(B_n)\end{aligned}$$

On the other hand,

$$\begin{aligned}\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \mu\left(B_0 \setminus \bigcap_{n\in\mathbb{N}} B_n\right)\\ &= \mu(B_0) - \mu\left(\bigcap_{n\in\mathbb{N}} B_n\right)\end{aligned}$$

Hence,

$$\mu(B_n) = \mu\left(\bigcap_{n\in\mathbb{N}} B_n\right)$$

□

Borel-Cantelli lemmas are just important pigeonhole principles.

The symbol $\exists_n^\infty$ means $\forall m \in \mathbb{N} \exists n \geq m$. This is useful because it is said that the brain can only handle 3 quantifiers at once - by using this symbol you can already remove one.

Theorem 1.11.1 (Borel-Cantelli)**.** Let$(X, \mathcal{B}, \mu)$ be a measure space. Let (A_n) be a sequence of μ-measurable sets.

(a) If $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$, then a.e. $x \in X$ is eventually not in A_n, i.e. the set
$$\limsup_{n\to\infty} A_n = \{x \in X : \exists_n^\infty x \in A_n\}$$
is μ-null.

(b) (Measure compactness) Suppose $\mu(X) < \infty$. If $\exists \delta > 0$ such that $\mu(A_n) \geq \delta$ for all $n \in \mathbb{N}$, then:
$$\mu\left(\limsup_{n\to\infty} A_n\right) \geq \delta$$

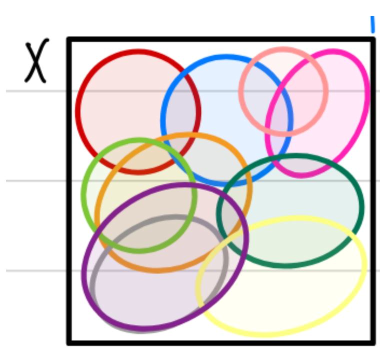


Figure 1.7: **(b):** Each set is represented by a coloured circle. If each set has measure $\geq \delta$, then $\limsup_{n\to\infty} A_n$ has measure $\geq \delta$..

Proof. Note that $\limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k$.

(a) Note that $\forall m \in \mathbb{N}$,

$$\limsup_{n\to\infty} A_n = \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n \subseteq \bigcup_{n\geq m} A_n$$

so by

$$\mu\left(\limsup_{n\to\infty} A_n\right) \leq \mu\left(\bigcup_{n\geq m} A_n\right) \leq \sum_{n\geq m} \mu(A_n)$$

Then $\sum_{n\geq m} \mu(A_n) \to 0$ as $m \to \infty$ since $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$ by assumption. Hence in the limit,

$$\mu\left(\limsup_{n\to\infty} A_n\right) \leq \lim_{m\to\infty} \sum_{n\geq m} \mu(A_n) = 0$$

(b) Since $\mu(X) < \infty$,

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} A_n\right) = \lim_{m\to\infty} \mu\left(\bigcup_{n\geq m} A_n\right) \geq \delta$$

Note that the last equality follows from [Proposition 1.11.1] using the fact that $\left(\bigcup_{n\geq m} A_n\right)$ is a decreasing sequence and that $\mu(X) < \infty$.

□

Definition 1.11.1 (vanishing, almost vanishing)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space. A sequence (V_n) of μ-measurable sets is called vanishing (resp. almost vanishing) if (V_n) is decreasing and

$$\bigcap_{n\in\mathbb{N}} V_n \textit{ is empty (resp. null)}$$

Proposition 1.11.2. Let $\mathcal{F}$ be a collection of μ-measurable sets that is closed under countable unions. If $\mathcal{F}$ contains sets of arbitrarily small positive measure, then it contains an almost vanishing sequence of positive measure sets.

Proof. For each $n \in \mathbb{N}$, let $A_n \in \mathcal{F}$ be a positive measure set with $\mu(A_n) \leq 2^{-n}$. Define $V_n := \bigcup_{m\geq n}$. Then $\bigcap_{n\in\mathbb{N}} V_n = \limsup_{n\to\infty} A_n$ is null by Borel-Cantelli , since $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$. □

Definition 1.11.2 (almost everywhere)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space and P a property of points in X. Then if $\{x : X \text{ satisfies } P\}$ is co-null we say:

P holds almost everywhere (a.e.) in X

or almost every (a.e.) $x \in X$ satisfies P

or P holds almost surely (a.s.)

1.12 *Measure exhaustion*

THINK OF greedy algorithms.

Definition 1.12.1 (almost disjoint). In a measure space, a collection $\mathcal{C}$ of sets is almost disjoint if the pairwise intersection of sets in $\mathcal{C}$ are null.

This theorem is a specific case of the countable chain condition (ccc) which appears in other areas of math.

Theorem 1.12.1 (Countable pigeonhole principle, for σ-finite measures). Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space. Then any almost disjoint collection $\mathcal{C}$ of μ-measurable positive measure sets is countable.

Proof. First we prove this in the case that $\mu(X) < \infty$. Then for each $n \in \mathbb{N}^+$, the set

$$\mathcal{C}_n := \left\{ C \in \mathcal{C} : \mu(C) \geq \frac{1}{n} \right\}$$

has $|\mathcal{C}_n| \leq n \cdot \mu(X)$, hence is finite. Then $\mathcal{C} = \bigcup_{n \in \mathbb{N}^+} \mathcal{C}_n$, so $\mathcal{C}$ is countable.

For the general σ-finite case, let $X = \bigsqcup_{n \in \mathbb{N}} X_n$ where $X_n \in \mathcal{B}$ and $\mu(X_n) < \infty$ for each X_n. Define

$$\mathcal{D}_n := \{ C \in \mathcal{C} : \mu(C \cap X_n) > 0 \}$$

By the finite case, each $\mathcal{D}_n$ is countable and $\mathcal{C} = \bigcup_{n \in \mathbb{N}} \mathcal{D}_n$, so $\mathcal{C}$ is countable. □

w_1 denotes the first uncountable ordinal.

Theorem 1.12.2 (Transfinite measure exhaustion). Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space and let $(A_\alpha)_{\alpha < \omega_1}$ be an increasing sequence of μ-measurable sets. Then the sequence almost stabilizes at some countable ordinal γ, i.e. $\forall \alpha \geq \gamma$, $A_\alpha =_\mu A_\gamma$.

Proof. We disjointify: define $A'_\alpha := A_\alpha \setminus \bigcup_{\beta < \alpha} A_\beta$, so $\{A'_\alpha\}_{\alpha < \omega_1}$ is disjoint (hence also almost-disjoint) collection. By the countable pigeonhole principle , all but countably many of the A'_α are null, hence $\exists$ a countable ordinal γ such that for all $\alpha > \gamma$, A_α is null. Hence $\forall \alpha > \gamma$, $A_\alpha =_\mu A_\gamma$ since $A_\alpha \setminus A_\gamma = \bigcup_{\gamma < \beta \leq \alpha} A'_\beta$ which is a countable union of null sets, hence null . □

Remark 1.12.2. *This allows us to run transfinite algorithms which at each step handle a positive measure set. Then it is guaranteed that the algorithm will stop at a countable stage, having handled a conull set.*

transfinite algorithms are algorithms that run for a greater than countable number of steps

We now discuss an important application. In a measure space with atoms, we can't achieve every value of measure between 0 and $\mu(X)$, but this is the only obstruction (i.e. if the measure space is atomless than we can achieve every value, which is what the next theorem states).

Theorem 1.12.3 (Sierpinski's Theorem). Let $(X, \mathcal{B}, \mu)$ be an atomless measure space. Then $\forall 0 < r \leq \mu(X)$, $\exists B \in \mathcal{B}$ such that $\mu(B) = r$.

Proof. We first prove a more humble statement.

Claim. Every positive measure set Y contains positive measure sets of arbitrarily small measure.

Proof. Y is not an atom so $\exists X_0 \subseteq Y$ with $\mu(X_0) < \mu(Y)$. We build a sequence $(X_s)_{s\in 2^{<\mathbb{N}}}$ of positive measure sets such that $X_s = X_{s0} \sqcup X_{s1}$ as follows: if X_s is defined, it's not an atom, so there is $X_{s0} \subseteq X_s$ in $\mathcal{B}$ with $0 \leq \mu(X_{s0}) \leq \mu(X_s)$. Let $X_{s1} := X_s \setminus X_{s0}$. For each $s \in 2^{<\mathbb{N}}$, one of X_{s0} and X_{s1} has measure $\frac{1}{2}\mu(X_s)$, which gives an infinite branch $(X_{s_n})_n \in \mathbb{N}$ in the tree of positive measure sets with $\mu(X_{s_n}) \leq \frac{1}{2^n}\mu(X_\emptyset)$.

□

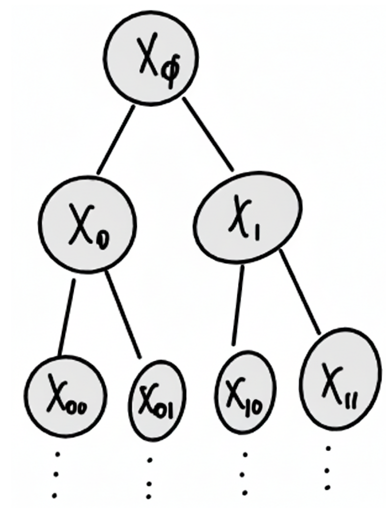


Figure 1.8: The tree of positive measure sets..

Proof via transfinite exhausion. Define a sequence $(A_\alpha)_{\alpha<\omega_1} \subseteq \mathcal{B}$ of pairwise disjoint sets such that $\mu\left(\bigsqcup_{\alpha<\beta} A_\alpha\right) \leq r$ for each $\beta \leq \omega_1 \leq r$ by induction as follows: if $(A_\alpha)_{\alpha<\beta}$ is already defined, let A_β be defined: $A_\beta := \emptyset$ if $r - \mu\left(\bigsqcup_{\alpha<\beta} A_\alpha\right) = 0$, otherwise (i.e. $r - \mu\left(\bigsqcup_{\alpha<\beta} A_\alpha\right) > 0$), $A_\beta :=$ a subset of $X \setminus \bigcup_{\alpha<\beta} A_\alpha$ of measure $0 < \cdot \leq r - \mu\left(\bigsqcup_{\alpha<\beta} A_\alpha\right)$

Now the proof of the countable pigeonhole for measures (using the condition that $\forall \beta < \omega_1$, $\mu\left(\bigsqcup_{\alpha<\beta} A_\alpha\right) \leq r$ instead of the finiteness of μ) gives that all but countably many of the A_α are null, i.e. $\exists \beta < \omega_1$ with A_α null for all $\alpha \geq \beta$. Thus $\mu\left(\bigsqcup_{\alpha<\beta} A_\alpha\right) = r$ □

Proof via $\frac{1}{2}$-greedy algorithm. We inductively build a sequence $(B_n)_{n\in\mathbb{N}} \subseteq \mathcal{B}$ of pairwise disjoint sets such that $\mu\left(\bigsqcup_{i\leq n}\right) \leq r$ as follows: suppose $(B_i)_{i<n}$ is defined and take $B_n \in \mathcal{B}$ to be any set with

$$\mu(B_n) \geq \frac{1}{2} \sup \left\{ B \in \mathcal{B} : B \subseteq X \setminus \bigsqcup_{i<n} B_i \text{ and } \mu(B) \leq r - \mu\left(\bigsqcup_{i<n} B_i\right)\right\}$$

Now that $(B_n)_{n\in\mathbb{N}}$, monotone convergence implies that $\sum_{n\in\mathbb{N}} \mu(B_n) = \mu\left(\bigsqcup_{n\in\mathbb{N}} B_n\right) \leq r$. In particular, $\lim_{n\to\infty} B_n = 0$. We now check that the set

$$B_\infty := \bigsqcup_{n\in\mathbb{N}} B_n$$

has measure r. Indeed, otherwise, $\mu(B_\infty) < r$ so by Claim 1, there is $B' \subseteq X \setminus B_\infty$ in $\mathcal{B}$ such that $0 < \mu(B') \leq r - \mu(B_\infty)$. But taking a large enough $n \in \mathbb{N}$ so that $\mu(B_n) < \frac{1}{2}\mu(B')$ we get a contradiction with the choice of B_n. □

1.13 Approximating measurable sets

1.13.1 The 99% lemma

Observation 1.13.1 (Carrots in soup)**.** *In a measure space $(X, \mathcal{B}, \mu)$, for sets A, B such that $0 < \mu(A), \mu(B) < \infty$, if $\frac{\mu(A \cap B)}{\mu(B)} \geq 0.99$ then for any partition $\{B_n\}$ of B into measurable sets, then for at least one $n \in \mathbb{N}$, $\frac{\mu(A \cap B_n)}{\mu(B_n)} \geq 0.99$.*

Colloquially, if a soup is 60% carrots, and then is partitioned into bowls, at least one of the bowls is 60% carrots.

This theorem is used like a replacement for a nice set - instead of a measurable set containing a nice set, it can contain most of a nice set which is sufficient for some arguments.

Theorem 1.13.1 (99% lemma)**.** Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space and $\mathcal{C} \subseteq \mathcal{B}$ be a generating collection of sets whose finite disjoint union forms an algebra. Then each μ-measurable set $M \subseteq X$ admits a set $C \in \mathcal{C}$ that is 99% M, i.e.

$$\forall \varepsilon > 0 \exists C \in \mathcal{C} \text{ such that } \frac{\mu(M \cap C)}{\mu C} = 1 - \varepsilon$$

Proof. By the uniqueness of Carathéodory's extension , $\mu = (\mu|_{\langle \mathcal{C} \rangle})^*$ where the notation is abused so that $\mu|_{\langle \mathcal{C} \rangle}$ is the premeasure, and $(\cdot)^*$ is defining the outer measure based on a premeasure . Thus,

$$\mu(M) = \inf \left\{ \sum_{n \in \mathbb{N}} \mu(A_n) : \{A_n\} \subseteq \langle \mathcal{C} \rangle \text{ and } \bigcup_{n \in \mathbb{N}} A_n \supseteq M \right\}$$

By disjointification we may assume that the A_k are pairwise disjoint; furthermore, since each A_n is a finite disjoint union of sets in $\mathcal{C}$ (by assumptions of the theorem), we get

$$\mu(M) = \inf \left\{ \sum_{k \in \mathbb{N}} \mu(C_k) : \{C_k\} \subseteq \mathcal{C} \text{ and } \bigsqcup_{k \in \mathbb{N}} C_k \supseteq M \right\}$$

Using σ-finiteness, M has a μ-measurable subset of positive finite measure , so by shrinking M, we may assume $\mu(M) < \infty$. Then $\exists \{C_k\} \subseteq \mathcal{C}$ such that $\bigcup_{k \in \mathbb{N}} C_k \supseteq M$ and $\frac{\mu(M)}{\mu(\bigsqcup_{k \in \mathbb{N}} C_k)} \geq 1 - \varepsilon$ for some $k \in \mathbb{N}$ by the carrots and soup observation . □

If $X = \bigsqcup_{n \in \mathbb{N}} X_n$ for $X_i \in \mathcal{B}$, $\mu(X_i) < \infty$ then $M = \bigsqcup_{n \in \mathbb{N}} M \cap X$. In particular, $0 \leq \mu(M \cap X_k) < \infty$ for some $k \in \mathbb{N}$.

If $M' \subseteq M$ and C is 99% M', then $C \geq 99\% M$. We can replace M with $X_k \cap M$ as defined above.

Example 1.13.1 (Use of 99% lemma)**.**

(a) For the measure space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, take $\mathcal{C} :=$ boxes, hence we get that every positive measure set contains 99% of a box.

(b) For the measure space $(A^{\mathbb{N}}, \mathcal{B}(A^{\mathbb{N}}), \mu)$, where A is finite and μ is Bernoulli, we take $\mathcal{C} :=$ cylinders, so every positive measure set contains 99% of a cylinder.

Remark 1.13.2. *In both of the examples, we can take the box or cylinder to be arbitrarily small, hence the carrot/soup observation applies.*

1.14 *Application: Ergodicity*

Definition 1.14.1 (ergodic)**.** Let (X, μ) be a measure space and let E be an equivalence relation on X. E is called ergodic w.r.t. μ or μ-ergodic if every E-invariant (i.e. union of E-classes) μ-measurable set is null or conull. In other words, X is not decomposable into two E-invariant positive measure sets.

The σ algebra is omitted. In other mathematical texts it will often be omitted, and the implied σ-algebra is the Borel σ-algebra.

This definition is technically "wrong by a null set", but equivalent in all the following examples. The correct definition replaces E-invariant with almost E-invariant, which is correct when there are uncountably-many equivalence classes.

Example 1.14.1 (Example of equivalence relations)**.**

(a) Let Γ be a countable group acting on a measure space $(X, \mathcal{B}, \mu)$, so that $\gamma \cdot \mathcal{B}$ for all $\gamma \in \Gamma$ and $B \in \mathcal{B}$. For example, translation actions $\mathbb{R} \curvearrowright \mathbb{R}$ or $\mathbb{Q} \curvearrowright \mathbb{R}$, or dialations $(\mathbb{Q}_{>0}, \cdot) \curvearrowright \mathbb{R}$. Then the orbit equivalence relation on X of this action, denoted E_Γ, and defined by:

$$\begin{aligned} x\ E_\Gamma\ y :\iff &\quad x \text{ and } y \text{ are in the same } \Gamma\text{-orbit} \\ :\iff &\quad y = \gamma \cdot x \text{ for some } \gamma \in \Gamma \end{aligned}$$

(b) Let $(X, \mathcal{B}, \mu)$ be a measure space and $T : X \to X$ be not necessarily a bijection. Typically, we assume T is μ-measurable. The orbit equivalence relation, denoted E_T, is defined by:

$$x\ E_T\ y := \iff T^n(x) = T^n(y) \qquad \text{for some } n, m \in \mathbb{N}$$

We draw an edge $x \to T(x)$. Then the T-orbits are exactly the connected components of this graph, which is the graph of T as a subset of $X \times X$.

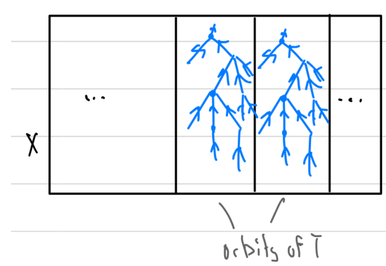


Figure 1.9: T-orbits correspond to connected components of the graph of T..

Example 1.14.2 (Examples of ergodic/nonergodic equivalence relations)**.**

(a) Nonergodic. Let $\mathbb{Z} \curvearrowright \mathbb{R}$ by translation: $z \cdot r = z + r$ for $z \in \mathbb{Z}, r \in \mathbb{R}$. The orbit equivalence relation of just the coset equivalence relation of $\mathbb{Z} \leq \mathbb{R}$. The orbit of $x \in \mathbb{R}$ is $x + \mathbb{Z}$. Then $A := \left(0, \frac{1}{2}\right) + \mathbb{Z}$ is $E_\mathbb{Z}$-invariant, but both it and it's complement have positive measure, so $E_\mathbb{Z}$ is not λ-ergodic, where λ is the Lebesgue measure. Also note that $E_\mathbb{Z}$ admits a measurable transversal, e.g. $[0, 1)$.

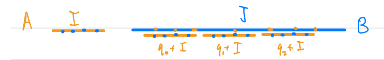
a transversal for the equivalence classes that compose A is $\left(0, \frac{1}{2}\right)$

(b) Ergodic. Let $\mathbb{Q} \curvearrowright \mathbb{R}$ by translation, so it's orbits' equivalence relation $E_\mathbb{Q}$ is the coset equivalence relation of $\mathbb{Q} \leq \mathbb{R}$. Recall that $E_\mathbb{Q}$ doesn't admit a measurable transversal, and the reason for this is that $E_\mathbb{Q}$ is ergodic.

Claim. $E_\mathbb{Q}$ is ergodic.

Proof. $E_\mathbb{Q}$ is ergodic. Suppose otherwise, i.e. there is a positive measure $A \subseteq \mathbb{R}$ with $B := \mathbb{R} \setminus A$ of positive measure. By the 99% lemma, there is a positive measure interval J whose 99% is

B. Again by the 99% lemma, there is a positive measure interval I whose 99% is A and moreover, $\mathrm{Ln}(I) < \mathrm{Ln}(J)$.



Using that rationals are dense, we can cover $\geq$ half of J by finitely many pairwise disjoint rational translations of I, i.e.

$$\bigsqcup_{i<K} (q_i + I) \subseteq J \quad \text{and} \quad \lambda\left(\bigsqcup_{i<K} q_i + I\right) \geq \frac{1}{2}\lambda(J)$$

Since $q_i + A = A$ for all i, we have that 99% of each $q_i + I$ is still A. So $\geq 0.5 \cdot 99\%$ of J is A, contradiction that only $\leq 1\%$ of J is A. □

1.15 *Regularity of measures (approximating with open/closed sets)*

Definition 1.15.1 (regular, strongly regular). Let $(X, \mathcal{B}, \mu)$ be a measure space and X a metric space. The μ is called regular if each μ-measurable set M satisfies:

$$\begin{aligned}\mu(M) &= \inf\{\mu(U) : U \supseteq M \text{ open}\} \\ &= \sup\{\mu(C) : C \subseteq M \text{ closed}\}\end{aligned}$$

μ is called strongly regular if:

$$\begin{aligned}0 &= \inf\{\mu(U \setminus M) : U \supseteq M \text{ open}\} \\ &= \sup\{\mu(M \setminus C) : C \subseteq M \text{ closed}\}\end{aligned}$$

Observation 1.15.2. *All finite regular measures are strongly regular*

Proposition 1.15.1. If μ is strongly regular, then for every measurable set M, $\exists$ a G_δ set G and F_σ set F such that $F \subseteq M \subseteq G$ and $\mu(F) = \mu(M) = \mu(G)$.

Proof. By strong regularity , for each set $n \in \mathbb{N}^+$ there are sets U_n open and C_n closed such that $C_n \subseteq M \subseteq U_n$ and $\mu(M \setminus C_n), \mu(U_n \setminus M) \leq \frac{1}{n}$. Let $G := \bigcap_{n\in\mathbb{N}} U_n$ and $F := \bigcup_{n\in\mathbb{N}} C_n$. Then $F \subseteq M \subseteq G$ and $\mu(M \setminus F) \leq \mu(M \setminus C_n) \leq \frac{1}{n} \to 0$ and $\mu(G \setminus M) \leq \mu(U_n \setminus M) \leq \frac{1}{n} \to 0$ as $n \to \infty$, so $F =_\mu M =_\mu G$. □

Theorem 1.15.1. Every finite Borel measure μ on a metric space X is strongly regular.

Proof. Let $\mathcal{S}$ be the collection of all μ-measurable sets $M \subseteq X$ which are strongly regular.

Claim. $\mathcal{S}$ contains all open sets

Proof. Recall that in metric spaces, open sets are F_σ (i.e. countable union of closed sets, HW), so for an open set $U \subseteq X$, we have that

$U = \bigcup_{n\in\mathbb{N}} C_n$ which we may turn into an increasing union by replacing C_n with $\bigcup_{i\leq n} C_i$, so assume $U = \bigcup_{n\in\mathbb{N}} C_n$, an increasing union of closed sets. By monotone convergence of measures , we have that $\mu(U) = \lim_{n\to\infty} \mu(C_n)$ □

Claim. $\mathcal{S}$ is an algebra.

Proof sketch. Follows from the fact that the complement of open/closed sets are closed/open and the finite unions of open/closed sets are closed/open. □

More proof details. Let $S \in \mathcal{S}$. Let $\varepsilon > 0$. Then there exists an open set $U \supseteq S$ such that $\mu(U \setminus S) < \varepsilon$. Note that U^C is a closed set and $U^C \subseteq S^C$. Then $U \setminus S = S^C \setminus U^C$, so $\mu(S^C \setminus U^C) < \varepsilon$. Hence $\sup\{\mu(S^C \setminus C) : C \subseteq S \text{ closed }\} < \varepsilon$. Since ε was arbitrary, $\sup\{\mu(S^C \setminus C) : C \subseteq S \text{ closed }\} = 0$. This shows inner strong regularity. The other arguments for outer regularity and finite unions are similar. □

Claim. $\mathcal{S}$ is closed under countable unions, hence it is a σ-algebra.

Proof. Let $M := \bigcup_{n\in\mathbb{N}} M_n$ where $M_n \in \mathcal{S}$, and by replacing M_n with $\bigcup_{i\leq n} M_n$, we can assume M_n is an increasing union. Let $U_n \supseteq M_n$ be an open set such that $\mu(M_n) \approx_{\varepsilon\cdot 2^{-(n+1)}} \mu(U_n)$. Hence $U := \bigcup_{n\in\mathbb{N}} U_n \supseteq M$ is open and:

$$\begin{aligned}\mu(U \setminus M) &= \mu(U \setminus \bigcup_{n\in\mathbb{N}} M_n) \\ &\leq \mu\left(\bigcup_{n\in\mathbb{N}} (U_n \setminus M_n)\right) \\ &\leq \sum_{n\in\mathbb{N}} \mu(U_n \setminus M_n) \\ &\leq \sum_{n\in\mathbb{N}} \varepsilon \cdot 2^{-(n+1)} = \varepsilon\end{aligned}$$

□

This $\mathcal{S}$ contains all Borel sets since it contains all open sets and is a σ-algebra . For a μ-measurable set M, $\exists B_0, B_1$ Borel sets such that $B_0 \subseteq M \subseteq B_1$ and $\mu(B_0) = \mu(M) = \mu(B_1)$. Let $U \supseteq B_1, C \subseteq B_0$ such that $\mu(U) \approx_\varepsilon \mu(B_1)$ and $\mu(C) \approx_\varepsilon \mu(B_0)$. Hence $\exists U$ open, C closed such that $C \subseteq M \subseteq U$ such that $\mu(C) \approx_\varepsilon \mu(M)$ and $\mu(U) \approx_\varepsilon \mu(M)$. Thus $\mathcal{S}$ contains all measurable sets. □

Caution 1.15.1. It is not true that σ-finite Borel measures on metric spaces are regular, let along strongly regular.

Example 1.15.1 (Non regular σ-finite Borel measure on a metric space)**.** Let X := one-point compactification of $\mathbb{R}$, i.e. $X = \mathbb{R} \cup \{\infty\}$ with the metric of the circle S^1 what we identify with X via stereographic projection $p : S^1 \to X$ by $p(N) = \infty$ and $p : S^1 \setminus N \sim\to \mathbb{R}$ as in the picture.

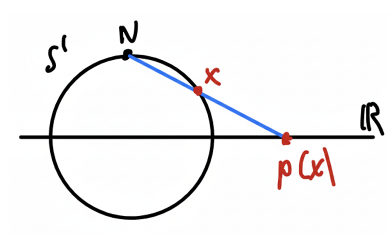


Figure 1.10: One-point compactification of $\mathbb{R}$. .

Thus the open sets of X are the open sets of $\mathbb{R}$ together with all sets of the form $(-\infty, a) \cup (b, \infty) \cup \{\infty\}$ (and unions of the above). Let $\bar{\lambda}$ be the Lebesgue measure on X defined by setting $\bar{\lambda}|_{\mathbb{R}}$ = the

Lebesgue measure and $\bar{\lambda}(\{\infty\}) = 0$. Thus $\bar{\lambda}$ is a σ-finite Borel measure since λ is σ-finite: $X = \{\infty\} \cup \bigcup_{n\in\mathbb{N}}(-n, n)$ where each set is a finite $\bar{\lambda}$ measure. Moreover, $X = \{\infty\} \cup \bigcup_{n\in\mathbb{N}}(-n, n)$ where each set has finite $\bar{\lambda}$-measure. And $X \equiv S'$ is obviously a metric space, in fact it is compact and Polish. However, $\bar{\lambda}(\infty) = 0 \neq \nu(U)$ where $U \supseteq \{\infty\}$ is open because $U \supseteq (-\infty, a) \cap (b, +\infty)$.

Note that in the previous example, X cannot be written as a countable union of finite measurable open sets. It turns out that this is the only obstruction.

Definition 1.15.3 (finite on compact sets, locally finite)**.** Let X be a Hausdorff topological space (e.g. a metric space) and μ a Borel measure on X. We say that μ is:

- σ-finite by open sets if $X = \bigcup_{n\in\mathbb{N}} U_n$ where each U_n is open and has that $\mu(U_n) < \infty$
- finite on compact sets if each compact set has finite measure
- locally finite if every point $x \in X$ admits a neighbourhood V of finite μ-measure, in particular, an open neighbourhood of finite measure.

Corollary 1.15.1. For a metric space X, every Borel measure that's σ-finite by open sets is strongly regular.

Proof. Since μ is σ finite by open sets, let $X = \bigcup_{n\in\mathbb{N}} U_n$ where each U_n is open and $\mu(U_n) < \infty$. Let $M \subseteq X$ be μ-measurable. For each $n \in \mathbb{N}$, viewing each U_n as a metric space and $\mu|_{U_n}$ as a finite Borel measure, we get that $\mu|_{U_n}$ is strongly regular so there exists a set $V_n \subseteq U_n$ open relative to U_n (hence open in X) such than $V_n \supseteq U_n \cap M$ and $\mu(V_n) \approx_{\varepsilon \cdot 2^{-(n+1)}} \mu(M \cap U_n)$. Thus $V = \bigcup_{n\in\mathbb{N}} V_n$ is open in X and $\mu(V \setminus M) = \mu\left(\bigcup_{n\in\mathbb{N}} V_n \setminus M\right) \leq \sum_{n\in\mathbb{N}} \varepsilon 2^{-(n+1)} = \varepsilon$. This handles strong outer regularity.

For strong inner regularity, let $U \supseteq M^C$ be an open set with $\mu\left(U \setminus M^C\right) \leq \varepsilon$. But U^C is closed and $U \setminus M^C = M \setminus U^C$ hence $\mu(M \setminus M^C) \leq \varepsilon$. □

Proposition 1.15.2. Let X be a Hausdorff space (e.g. metric space). Consider the properties:

(1) μ is finite on open sets

(2) μ is locally finite

(3) μ is σ finite by open sets

For a Borel measure μ, the following implications hold:

- $(3) \implies (2) \implies (1)$ always
- $(1) \implies (2)$ when X is locally compact
- $(2) \implies (3)$ when X is second-countable

In particular, the three properties are equivalent for $\mathbb{R}^d$, $A^{\mathbb{N}}$ (A finite), and hence both are strongly regular.

Lemma 1.15.1. For a metric space X, and a set $K \subseteq X$, the following are equivalent:

(1) K is compact

(2) K is sequentially compact

(3) K is complete and totally bounded

Proof.

$(3) \implies (2)$ If $X = \bigcup_{n\in\mathbb{N}} U_n$ where each U_n is open and $\mu(U_n) < \infty$ then $\forall x \in X \exists n \in \mathbb{N}$ such than $x \in U_n$ with finite measure.

$(2) \implies (1)$ Let K be compact and $\forall x \in K$ let U_x be an open neighbourhood of finite measure such that $x \in U_x$. Then the open cover $\{U_x\}_{x\in K}$ has a finite subcover $\{U_{x_1}, \dots, U_{x_n}\}$ so $K \subseteq \bigcup_{i\le n} U_{x_i}$ has finite measure.

$(1) \implies (2)$ Assume that X is locally compact. Then every point $x \in X$ has a compact neighbourhood and then by (1), compact sets have finite measure.

$(2) \implies (3)$ Assume X is second countable. Let $\{U_n\}$ be a countable basis for X. Then by (2), for each $x \in X$ there is an open neighbourhood U of finite measure such that $x \in U$, hence $\exists n_x$ with $x \in U_{n_x} \subseteq U$ so U_{n_x} has finite measure. But then $X = \bigcup_{x\in X} U_{n_x}$ but this unions is actually countable.

□

1.16 *Tightness*

Definition 1.16.1 (tight measure)**.** A Borel measure μ on a Hausdorff topological space (e.g. metric space) is called tight if for every μ-measurable set $M \subseteq X$, we have that

$$\mu(M) = \sup\{\mu(K) : K \subseteq M \text{ compact}\}$$

Theorem 1.16.1. Finite Borel measures on Polish spaces are tight.

Proof. Since we know a finite Borel measure μ on a Polish space is strongly regular , every μ-measurable set can be approximated from below by closed sets, so it is enough to show that closed sets can be approximated from below by compact sets. Let C be a closed set. Since C is Polish with the same metric, we might as well assume $X = C$. Let $\varepsilon > 0, \varepsilon_n = \frac{1}{n}$ and for each $n \in \mathbb{N}$, let $\{B_l^{\varepsilon_n}\}_{l\in\mathbb{N}}$ be a countable cover of X with closed balls of radius $\leq \varepsilon$ (such a cover exists by separability). Because $X = \bigcup_{L\in\mathbb{N}} \left(\bigcup_{l\leq L} B_l^{\varepsilon_n}\right)$ we have that $\mu(X) \approx_{\varepsilon\cdot 2^{-(n+1)}} \mu\left(\bigcup_{l<L_n} B_l^{\varepsilon_n}\right)$ for L_n large enough by monotone convergence . Let $C_n := \bigcup_{l\leq L_n} B_l^{\varepsilon_n}$ so C_n is closed and $K := \bigcap_{n\in\mathbb{N}} C_n$ is still closed but also totally bounded by definition , hence compact, and finally, $\mu(X \setminus K) \leq \mu\left(\bigcup_{n\in\mathbb{N}} X \setminus C_n\right) \leq \sum_{n\in\mathbb{N}} \varepsilon \cdot 2^{-(n+1)} = \varepsilon.$ ☐

Corollary 1.16.1 (strong regularity and tightness)**.** Let X be a Polish space. Then every locally finite Borel measure on X is strongly regular and tight.

Proof. Polish spaces are second countable , so local finiteness is equivalent to σ-finiteness by open sets , i.e. $X = \bigcup_{n\in\mathbb{N}}$ where each U_n is open and $\mu(U_n) < \infty$ Definition 1.15.3. Thus μ is strongly regular and we only need to show tightness. From descriptive set theory, we know that open sets of Polish spaces are Polish (with a different equivalent metric), so on each U_n, we know that μ is tight . We leave the rest as HW (similar to the proof of inner regularity). ☐

Definition 2.0.1. Let $(X, \mathcal{I}), (Y, \mathcal{J})$ be measurable spaces. A function $f : X \to Y$ is said to be:

(a) $(\mathcal{I}\text{-}\mathcal{J})$-measurable if $f^{-1}(J) \in \mathcal{I}$ for all $J \in \mathcal{J}$.

(b) $\mathcal{I}$-measurable (or measurable if $\mathcal{I}$ is clear) if Y is a metric/topological space and $\mathcal{J} = \mathcal{B}(Y)$ is the Borel σ-algebra of Y and f is $(\mathcal{I}\text{-}\mathcal{B}(Y))$-measurable.

(c) Borel if X, Y are metric/topological spaces, $\mathcal{I} = \mathcal{B}(X), \mathcal{J} = \mathcal{B}(Y)$, and f is $\mathcal{B}(X)\text{-}\mathcal{B})Y)$-measurable.

(d) μ-measurable if μ is a measure on $(X, \mathcal{I})$, Y is a metric spaces, $\mathcal{J} = \mathcal{B}(Y)$, and f is Meas_μ-measurable, i.e. $f^{-1}(B)$ is μ-measurable for each Borel set $B \subseteq Y$

Remark 2.0.2. *For functions $\mathbb{R} \to \mathbb{R}$, we view the left $\mathbb{R}$ as a measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ and the right $\mathbb{R}$ as a metric space, so the definition of λ-measurable is asymmetric. This is done to get more function to be called measurable since the theory works for them.*

Proposition 2.0.1. Let $(X, \mathcal{I})$, $(Y, \mathcal{J})$ be measurable spaces and $f : X \to Y$. If for some generating collection $\mathcal{J}_0$ of $\mathcal{J}$, $f^{-1}(J_0) \in \mathcal{I}$ for all $J_0 \in \mathcal{J}_0$, then f is $(\mathcal{I}\text{-}\mathcal{J})$-measurable.

It is important to learn this proof technique where if we want to prove a property about all the sets in a sigma algebra, we consider the set on which the property holds and argue that it is a σ-algebra.

Proof. Let $\mathcal{S} := \{J \in \mathcal{J} : f^{-1}(J) \in \mathcal{I}\}$ and observe that $\mathcal{S} \supseteq \mathcal{J}_0$. Moreover, $\mathcal{S}$ is a σ-algebra since preimages respect unions and complements. Hence $\mathcal{S}$ is a σ-algebra containing $\mathcal{J}$, i.e. $\mathcal{S} \supseteq \mathcal{J}$. □

Corollary 2.0.1. Let $(X, \mathcal{I})$ be a measurable, space, Y a metric space, and $f : X \to Y$. If $f^{-1}(V) \in \mathcal{I}$ for each $V \subseteq Y$, V open, then f is $\mathcal{I}$-measurable. In particular, continuous functions are Borel, since $f^{-1}(\text{open})$ is open.

The following is one of the reasons for building measure theory:

Theorem 2.0.1. Let $(X, \mathcal{I})$ be measurable space, Y a separable metric space, (f_n) a sequence of measurable functions $f_n : X \to Y$ such that $f_n \to f$ pointwise, then f is $\mathcal{I}$-measurable.

$\forall_n^\infty$ denotes infinitely often, i.e. $\forall n \in \mathbb{N} \exists m \geq n$ such that $f_n(x) \in U$.

Proof. By the last corollary, it is enough to show that $f^{-1}(U) \in \mathcal{I}$ for each open $U \subseteq Y$. Note that the openness of U gives the following: for $x \in X$,

$$f(x) \in U \implies \forall_n^\infty f_n(x) \in U$$

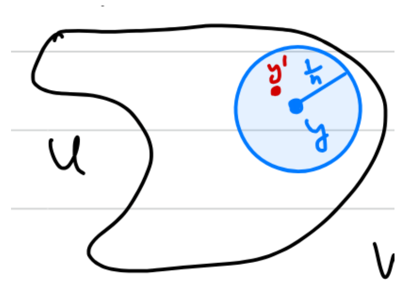
If the converse were true, we would be done since

This is given as motivation for the below proof

$$f^{-1}(U) = \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} f_n^{-1}(U)$$

so then we would have $f^{-1}(U) \in \mathcal{I}$.

But the converse isn't true: for example, let $U = (0,1) \subseteq \mathbb{R}$, $f_n := \frac{1}{n}$, so $f_n(x) \in U$ for all n, but the limit of f_n is $0 \notin U$. The converse holds for closed sets, but U is open. However, using separability, we can present U so it enjoys both properties (behaves both as a closed and open set).



Claim. $\bigcup_{k \in \mathbb{N}} V_k = U = \bigcup_{k \in \mathbb{N}} \overline{V}_n$ for some collection of open sets $V_k \subseteq Y$.

Proof. Let $D \subseteq Y$ be a countable dense set and $\mathcal{V} := \{B_{1/n}(y) : y \in D, n \in \mathbb{N}^+, \overline{B_{1/n}(y)} \subseteq U\}$. Note that if $V \in \mathcal{V}$, then $\overline{V} \subseteq U$ by definition, so it is enough to show that $U = \bigcup_{V \in \mathcal{V}} U$. Fix $y \in U$, hence $\overline{B_{1/n}(y)} \subseteq U$ for large enough $n \in \mathbb{N}$. Let $y' \in D$ such that $y' \in B_{\frac{1}{2n}}(y)$. This is equivalent to $y \in B_{\frac{1}{2n}}(y')$. But: $\overline{B_{\frac{1}{2n}}(y')} \subseteq \overline{B_{\frac{1}{n}}(y)} \subseteq U$ so $\overline{B_{\frac{1}{2n}}(y')} \in \mathcal{V}$. Hence $y \in \bigcup_{V \in \mathcal{V}} V$. □

□

Proposition 2.0.2. Let X, Y be metric spaces where Y is second countable. Let μ be a Borel measure on X. Let $f : X \to Y$ be a μ-measurable functions. Then:

(a) f is Borel on a conull Borel sets, i.e. $f|_{X'} : X' \to Y$ is a Borel function for some conull Borel set X'.

(b) Luzin's Theorem: $\forall \varepsilon > 0$, f is continuous on a closed set C with $\mu(X \setminus C) < \varepsilon$.

$\mathcal{B}(X') = \{B \in \mathcal{B}(X) : B \subseteq X'\}$

Proof. (a) f^{-1} is μ-measurable, hence $f^{-1}(V_n) =_\mu B_n$ for some Borel $B_n \subseteq X$. Let $Z := \bigcup_{n \in \mathbb{N}} \left(f^{-1}(V_n) \triangle B_n\right)$, so Z is null, hence $Z \subseteq \tilde{Z}$ where $\tilde{Z}$ is Borel and still null. Put $X' := X \setminus Z$, so X' is Borel and conull. Then $(f|_{X'})^{-1}(V_n) = f^{-1}(V_n) \cap X' = B_n \cap X'$ which is Borel. So $f|_{X'}$ is Borel.

(b) $f^{-1}(V_n)$ is μ-measurable, hence by strong outer regularity, $\exists U_n \subseteq X$ open such that $d_\mu\left(U_n, f^{-1}(V_n)\right) := \mu\left(U_n \triangle f^{-1}(V_n)\right) \leq \varepsilon \cdot 2^{-(n+2)}$ so $\mu(Z) \leq \sum_{n \in \mathbb{N}} \varepsilon \cdot 2^{-(n+1)} = \frac{\varepsilon}{2}$. Again by outer regularity, there is an open set $\tilde{Z} \supseteq Z$ with $\mu(\tilde{Z} \setminus Z) \leq \frac{\varepsilon}{2}$ so $\mu(\tilde{Z}) \leq \varepsilon$. Take $C := X \setminus \tilde{Z}$, so it's closed and $\mu\left(X \setminus C\right) \leq \varepsilon$. Furthermore:

$$(f|_C)^{-1}(V_n) = f^{-1}(V_n) \cap C = U_n \cap C$$

so $(f|_C)^{-1}(V_n)$ is open relative to C, hence $f|_C$ is continuous.

□

2.1 Pushforward Measures

Definition 2.1.1 (pushforward measure)**.** Let $(X, \mathcal{I}), (Y, \mathcal{J})$ be measurable spaces and let μ be a measure on $\mathcal{I}$. Let $f : X \to Y$ be an $(\mathcal{I}$-$\mathcal{J})$-measurable function. Then the f-pushward of μ is the measure $f_*\mu$ (also written μf^{-1}) on $\mathcal{J}$ defined by: for $J \in \mathcal{J}$:

$$f_*\mu(J) = \mu(f^{-1}(J))$$

Example 2.1.1 (Examples of pushforward measures)**.**

(a) Let $S^1 \subseteq \mathbb{C}$ denote the unit circle, which is usually considered a group under complex multiplication. This is identified with the group $(\mathbb{R} \setminus \mathbb{Z}, +)$ as follows: $\mathbb{R} \setminus \mathbb{Z} \simeq [0, 1)$ since $[0, 1)$ is a transversal for the coset equivalence relation of $\mathbb{Z} \leq \mathbb{R}$. Define $f : [0, 1) \to S^1$ by $x \mapsto e^{2\pi i x}$ which is a group isomorphism $(\mathbb{R} \setminus \mathbb{Z}, +) \xrightarrow{\sim} (S^1, \cdot)$. Then $f_*\lambda$ is a Borel measure on S^1.

(b) Let $\mathbb{R}_{>0}$ be the group of positive reals under $\cdot$. Consider $f : \mathbb{R} \to \mathbb{R}_{>0}$ by $x \mapsto e^x$ and take the pushforward measure $f_*\lambda$. In particular, for $(a, b) \in \mathbb{R}$,

$$f_*\lambda((a, b)) = \lambda((\ln a, \ln b)) = \ln b - \ln a$$

Definition 2.1.2 (measure preserving/invariant function/group action)**.** Let $(X, \mathcal{I}, \mu)$ be a measure space.

- For an $(\mathcal{I}$-$\mathcal{I})$-measurable function $f : X \to X$, we say μ is f-invariant or that f preserves μ if $f_*\mu = \mu$, i.e. $\mu(B) = \mu(f^{-1}(B))$ for each $B \in \mathcal{I}$.
- For a group action $\Gamma \curvearrowright X$ such that each group element acts as an $(\mathcal{I}$-$\mathcal{I})$-measurable function, we say μ is Γ-invariant or Γ preserves μ if for each $\gamma \in \Gamma$, $\gamma_*\mu = \mu$.

A topological group is a group equipped with topology making multiplication and inverse continuous. A Borel measure μ on G is called left-invariant (resp right-invariant) if it is invariant under the left-translation (resp. right-translation) action $G \curvearrowright G$ i.e. $\mu(g \cdot B) = \mu(B)$ (resp. $\mu(B \cdot G) = \mu(B)$).

Theorem 2.1.1 (Haar)**.** Every locally compact (Hausdorff) group admits a unique (up to rescaling) locally-finite Borel measure (also a right-invariant Borel measure) that is locally finite (i.e. finite on compact sets). This measure is called the left Haar measure (resp. right Haar measure).

Proof. Note given. Too technical, need a bit more knowledge about topological groups. □

Example 2.1.2 (Haar measures)**.**

(a) For $(\mathbb{R}^d, +)$, the Lebesgue measure is the Haar measure.

(b) For $(\mathbb{R}_{>0}, \cdot)$, the pushforward of the Lebesgue measure by $x \mapsto e^x$, $\mathbb{R} \to \mathbb{R}_{>0}$ is a Haar measure because this function is a topological group isomorphism and λ is Haar for $(\mathbb{R}, +)$.

(c) For $(S^1, \cdot)$, the pushforward of λ by $x \mapsto e^{2\pi i x}$, $[0,1] \to S^1$ is Haar for $[0,1) \simeq \mathbb{R} \setminus \mathbb{Z}$. Note that S^1 is compact and this measure is a probability measure.

(d) Consider the group $(\mathbb{Z} \setminus 2\mathbb{Z})^{\mathbb{N}} \simeq 2^{\mathbb{N}}$ as a compact group with the same topology as $2^{\mathbb{N}}$ under coordinate-wise addition mod 2. The Bernoulli(1/2) measure is the Haar probability measure on $(\mathbb{Z} \setminus 2\mathbb{Z})^{\mathbb{N}}$. Incidentally, this is also the pushforward of the Lebesgue measure on $[0,1) \simeq \mathbb{R} \setminus \mathbb{Z}$ by $f : [0,1) \to 2^{\mathbb{N}} \simeq (\mathbb{Z} \setminus 2\mathbb{Z})^{\mathbb{N}}$ by $x \mapsto$ the binary representation of x.

Technically rationals can have multiple representations (e.g. $0.5 = 0.1000\ldots = 0.0111\ldots$) but we sweep this detail under the rug since this only fails on a null set (the dyadic rationals, i.e. rationals with a power of two in the denominator). For this map to be well-defined, we pick the first form (ending in zeros) by convention and hence this map is injective into $2^{\mathbb{N}}$.

Example 2.1.3 (Nonexample)**.** The point of this example is to illustrate that not all Haar measures are pushforwards measures of the Lebesgue measure. The more correct ideas is that if two topological groups G, H are isomorphic via a measurable function f and μ is a Haar measure on G, then $f_*\mu$ is a Haar measure on H, but the reverse does not hold:

The group $\mathrm{GL}_n(\mathbb{R})$ of all invertible real matrices under multiplication is locally compact when viewed as a subset of $\mathbb{R}^{n^2}$. Indeed, if $M \in \mathrm{GL}_n(\mathbb{R}) \iff \det M \neq 0$, and the latter is an open condition, so $\mathrm{GL}_n(\mathbb{R})$ is an open subset of $\mathbb{R}^{n^2}$. Furthermore, it's complement, $\{M \in \mathbb{R}^{n \times n} : \det M = 0\}$, is a "lower dimensional" closed set, and one can show that it's null, so $\mathrm{GL}_n(\mathbb{R})$ is a Lebesgue conull open subset of $\mathbb{R}^{n^2}$. However, the Lebesgue measure on $\mathrm{GL}_n(\mathbb{R})$ is not a Haar measure (neither left nor right) because for example multiplication by $2I_n$ scales the Lebesgue measure by 2. The Haar measure on $\mathrm{GL}_n(\mathbb{R})$ is defined using the Jacobian, i.e. integral with $\frac{1}{\det}$ in it.

The determinant map $\det : \mathbb{R}^{n^2} \to \mathbb{R}$ is continuous (composition of finitely many addition and multiplications). Hence it maps the preimage of open sets to open sets. $\mathbb{R} \setminus \{0\}$ is open, hence $\det^{-1}(\mathbb{R} \setminus \{0\} = \mathrm{GL}_n(\mathbb{R}))$ is open.

We showed that translation actions $\mathbb{Q} \curvearrowright \mathbb{R}$ and $\bigoplus_{n\in\mathbb{N}} (\mathbb{Z} \setminus 2\mathbb{Z}) \curvearrowright \prod_{n\in\mathbb{N}} \mathbb{Z} \setminus 2\mathbb{Z}$ are ergodic, and one can also similarly show that for an irrational $\lambda \in \mathbb{R} \setminus \mathbb{Q}$, the rotation by $2\pi\lambda$ on S^1 is ergodic, which is the same as the translation action on the (dense) subgroup $\langle e^{2\pi i\lambda} \rangle \leq S^1$ on S^1. The following show that this is a general phenomenon.

2.2 *Borel/measure isomorphism theorems*

The following is one of the basic theorems in descriptive set theory which is used by mathematicians (e.g. ergodic theorists, probability theorists) all the time without mention.

Theorem 2.2.1 (Borel isomorphism theorem)**.** Any two uncountable Polish spaces are Borel-isomorphic, i.e. $\exists f : X \to Y$ a bijection such that f and f^{-1} are Borel.

$\hookrightarrow$ denotes an injective mapping.

An embedding is an injective continuous map such that the map to it's image is a homeomorphism.

Proof Sketch. For an uncountable Polish space X, it's enough to show that X is Borel-isomorphic to $2^{\mathbb{N}}$. By the Borel version of the Cantor-Shroder-Berstein theorem, it is enough to show that $\exists$ injective Borel maps $2^{\mathbb{N}} \hookrightarrow X$ and $X \hookrightarrow 2^{\mathbb{N}}$.

The first is called the Cantor-Bendixson theorem: for each uncountable Polish X, there is a continuous embedding $2^{\mathbb{N}} \hookrightarrow X$. □

Lemma 2.2.1 (Binary representation)**.** Any second countable metric space X admits a Borel

$$\ell_0 : X \hookrightarrow 2^{\mathbb{N}}$$

which we call a binary representation map.

Proof. Let (U_n) be a countable basis for X so it separates points. Then define $b : X \hookrightarrow 2^{\mathbb{N}}$ by $x \mapsto (\mathbb{1}_{U_n}(x))_{n\in\mathbb{N}}$. To check that b is Borel, it suffices to check that $b^{-1}(V_n) = U_n$ is Borel where $V_n := \{x \in 2^{\mathbb{N}} : x(n) = 1\}$, because $\{V_n\}$ generates $\mathcal{B}(2^{\mathbb{N}})$ (each cylinder in $2^{\mathbb{N}}$ is a finite intersection of these V_n and their complements). □

We proved this in HW

Definition 2.2.1. A measurable space $(X, \mathcal{I})$, is called a standard Borel space if there is a Polish metric on X such that $\mathcal{I} = \mathcal{B}(X)$.

In other words, X was a Polish space, but we forgot which sets are open and only kept the Borel sets.

This the Borel isomorphism theorem says taht there is only one (up to isomorphism) standard Borel space.

Definition 2.2.2 (measure isomorphism)**.** Let $(X, \mathcal{I}, \mu), (Y, \mathcal{J}, \nu)$ by measure spaces. A function $f : X \to Y$ is called a measure isomorphism if there are conull sets $X' \subseteq X, Y' \subseteq Y$ such that:

- $(f|_{X'} : X' \to Y')$ is a bijection,
- $(f|_{X'})$ is ($\mathcal{I}$-$\mathcal{J}$)-measurable and $(f^{-1}|_{Y'})$ is ($\mathcal{J}$-$\mathcal{I}$)-measurable,
- $f_*\mu = \nu$

Theorem 2.2.2 (Measure isomorphism theorem)**.** Every atomless Borel measure on a Polish space X is isomorphic to $([0,1], \lambda)$. In fact, there is a Borel isomorphism $f : X \to [0,1]$ with $f_*\mu = \lambda$.

Proof. By the Borel isomorphism theorem, there is a Borel isomorphism $g : X \to [0,1]$, so by replacing X with $[0,1]$ and μ with $g_*\mu$ we may assume $X = [0,1]$ and μ is an atomless Borel probability measure on $[0,1]$. [1] Let $f : [0,1] \to [0,1]$ by defined $x \mapsto \mu([0,x])$. This is a nondecreasing function and continuous. It is right-continuous from downward monotone convergence, $\mu([0,x]) = \lim_{x_n \to x^+} \mu([0,y])$ and left-continuous because of upward monotone convergence and

[1] e.g. maybe $\mu|_{[0,\frac{1}{2}]} = 0$ and $\mu|_{(\frac{1}{2},1]} = 2\lambda$.

atomlessness, $\mu([0,x]) = \mu([0,x)) = \lim_{x_n \to x^-} \mu([0,x_n])$. Furthermore, by definition, $f^{-1}([0,y]) = [0,x]$ where $\mu([0,x]) = y$ and this x is maximal, so $\lambda([0,y]) = y = \mu(f^{-1}([0,y])) = f_*\mu([0,y])$.

Since the sets generate the Borel σ-algebra, the measures λ and $f_*\mu$ coincide. It remains to show that f is a bijection on a conull set. f might not be bijective because there could be intervals $[a.b]$ on which f is constant, but then $\mu((a,b]) = 0$ and there are only countably many such maximal intervals because maximal ones are pairwise disjoint and $\mathbb{R}$ is separable. So the union Z of all these maximal intervals $(a,b]$ is still μ-null and $f|_{X \setminus Z} : X \setminus Z \to [0,1]$ is still a bijection. □

Corollary 2.2.1. Every σ-finite Borel measure μ on a Polish space X is isomorphic to $(\mathbb{R}, \lambda)$.

Proof. Write X is a countable disjoint union of finite positive measure pieces and use that each is isomorphic to an interval in $\mathbb{R}$. Details left as an exercise. □

Definition 2.2.3 (standard measure space)**.** A measure space $(X, \mathcal{B}, \mu)$ is called <u>standard</u> if it is σ-finite and $(X, \mathcal{B})$ is standard Borel.

Remark 2.2.4. *In dynamics and probability theory, one mostly works with standard probability spaces (since we know that atomless ones are all isomorphic)*

Theorem 2.2.3.

- A standard atomless infinite measure space is isomorphic to $(\mathbb{R}, \lambda)$.
- A standard atomless probability measure space is isomorphic to $([0,1], \lambda)$.

2.3 *Integration*

Given a measurable space $(X, \mathcal{B})$, we denote:

- $L(X, \mathcal{B})$ the set of all $\mathcal{B}$-measurable functions $X \to \overline{\mathbb{R}} := [-\infty, \infty]$
- $L^+(X, \mathcal{B}) := \{f : L(X, \mathcal{B}) : f \geq 0\}$

Note that $L(X, \mathcal{B})$ is a vector space and $L^+(X, \mathcal{B})$ is closed under non-negative linear combinations. Both are closed under products and limits.

Definition 2.3.1 (integral)**.** An <u>integral on $L^+(X, \mathcal{B})$</u> is a countably additive linear functional $I : L^+(X, \mathcal{B}) \to [0, \infty]$, i.e.

(i) $I(\alpha \cdot f + \beta \cdot g) = \alpha \cdot I(f) + \beta \cdot I(g) \quad \forall \alpha, \beta \geq 0, f, g \in L^+(X, \mathcal{B})$

(ii) $I(\sum_{n \in \mathbb{N}} f_n) = \sum_{n \in \mathbb{N}} I(f_n)$

Observation 2.3.2. *Every integral I on $L^+(X,\mathcal{B})$ defines a measure μ_I on $(X,\mathcal{B})$ by $\mu_I(B) = I(\mathbb{1}_B)$*

Proof. Indeed, observe that:

(i) $\mu_I(\emptyset) = I(\mathbb{1}_\emptyset) = I(0) = 0$

(ii) If $B = \bigcup_{n\in\mathbb{N}} B_n$, then $\mathbb{1}_B = \sum_{n\in\mathbb{N}} \mathbb{1}_{B_n}$, so

$$\mu_I(B) = I(\mathbb{1}_B) = I\left(\sum_{n\in\mathbb{N}} \mathbb{1}_{B_n}\right) = \sum_{n\in\mathbb{N}} I(\mathbb{1}_{B_n}) = \sum_{n\in\mathbb{N}} \mu_I(B_n)$$

□

We would like to solve the inverse problem, and indeed we will prove the following:

Theorem 2.3.1. Every measure μ on $(X,\mathcal{B})$ admits a unique integral $I : L^+(X,\mathcal{B}) \to [0,\infty]$ so that $\mu_I = \mu$. This integral is called the integral over μ, denoted:

$$\int f(x)\, d\mu(x) = I(f) = \int f\, d\mu$$

Remark 2.3.3. *Once I is defined on L^+, we will also be able to define it on a subspace of $L(X,\mathcal{B})$, using the fact that for each function, $f = f_- + f_+$ where:*

$$f_-(x) = \begin{cases} -f(x) & \text{if} f(x) < 0 \\ 0 & \text{o.w.} \end{cases} \quad \text{and} \quad f_+(x) = \begin{cases} f(x) & \text{if} f(x) \geq 0 \\ 0 & \text{o.w.} \end{cases}$$

Definition 2.3.4 (simple function)**.** A function $f \in L(X,\mathcal{B})$ is called simple if it is a finite linear combination of indicator functions of sets from $\mathcal{B}$, so nonnegative simple functions are nonnegative linear combinations of indicator functions. Denote by $S(X,\mathcal{B})$ and $S^+(X,\mathcal{B})$ the subspaces of simple and nonnegative simple functions.

Note that a nonnegative simple function f is of the form:

$$f = \sum_{i<n} \alpha_i \mathbb{1}_{A_i}$$

for some $n \in \mathbb{N}, \alpha_i \geq 0, A_i \in \mathcal{B}$. It has infinitely many representatives but there is a standard one.

Observation 2.3.5. *A function $f \in L(X,\mathcal{B})$ is simple $\iff$ range of $f(x)$ is finite.*

Definition 2.3.6 (standard representation of simple function)**.** For a simple function $f \in L(X, \mathcal{B})$, the representation:

$$f = \sum_{i<n} \alpha_i \mathbb{1}_{f^{-1}(\alpha_i)}$$

is <u>standard</u> where $f(X) = \{\alpha_0, \ldots, \alpha_{n-1}\}$. In particular, $X = \bigsqcup_{i<n} f^{-1}(\alpha_i)$

Definition 2.3.7 (integral of simple function)**.** Given a measure μ on $(X, \mathcal{B})$, we define by setting for each $f \in S^+(X, \mathcal{B})$,

$$\int f \, d\mu = \sum_{i<n} \alpha_i \mu(A_i) \quad (*)$$

where $\sum_{i<n} \alpha_i \mu(A_i)$ is some representation of f.

It is an exercise to show that this is well define, i.e. does not depend on choice of representation.

Remark 2.3.8. *We could alternatively define the integral for the standard representation and prove that* $(*)$ *holds for any representation.*

Proposition 2.3.1. The integral on S^+ satisfies the following:

(i) Linearity: $\int \alpha f + \beta g \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu$

(ii) Monotonicity: if $f > 0$, then $\int f \, d\mu \geq 0$.

Equivalently, $f \leq g \implies \int f \, d\mu \leq \int g \, d\mu$

(iii) $\int f \, d\mu = 0 \implies f = 0$ a.e.

(iv) For each $f \in S^+$, we define a measure μ_f on $(X, \mathcal{B})$ by $\mu_f(B) = \int \mathbb{1}_B f \, d\mu = \int_B f \, d\mu$

Proof. (i)-(iii) follow by definition once well-definedness is established.

(iv):

(i) $\mu_f(\varnothing) = 0$ is clear.

(ii) Let $B = \bigsqcup_{n\in\mathbb{N}} B_n$, so $\mathbb{1}_B = \sum_{n\in\mathbb{N}} \mathbb{1}_{B_n}$, hence,

$$\begin{aligned}
\mu_f(B) &= \int \mathbb{1}_B f \, d\mu \\
&= \int \sum_{i<m} \alpha_i \mathbb{1}_B \mathbb{1}_{A_i} \, d\mu \\
&= \int \sum_{i<m} \alpha_i \mathbb{1}_{A_i \cap B} \, d\mu \\
&= \sum_{i<m} \alpha_i \mu(A_i \cap B) \\
&= \sum_{i<m} \alpha_i \sum_{n\in\mathbb{N}} \mu(A_i \cap B_n) \\
&= \sum_{n\in\mathbb{N}} \sum_{i<m} \alpha_i \mu(A_i \cap B_n) \\
&= \sum_{n\in\mathbb{N}} \int \mathbb{1}_{B_n \cdot f} \, d\mu \\
&= \sum_{n\in\mathbb{N}} \mu_f(B_n)
\end{aligned}$$

□

Proposition 2.3.2 (approximation by simple functions)**.**

(a) For each $f \in L^+(X, \mathcal{B})$, there is an increasing sequence of non-negative simple functions $f_n \nearrow f$ such that the convergence is uniform on every set $B \subseteq X$on which f is bounded, i.e.

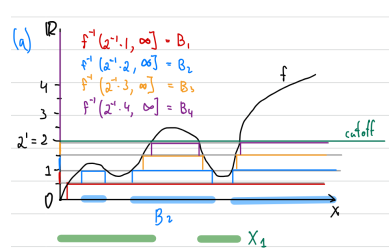
$$\|f_n|_B - f|_B\|_u \to 0 \text{ as } n \to \infty$$

(b) $\forall f \in L(X, \mathcal{B})$, there is a sequence of simple functions $f_n \to f$ such that $|f_n| \nearrow |f|$ and the convergence is uniform on every set $B \subseteq X$ on which f is bounded.

Proof. (b) follows from (a) applied to f^+ and f^-, thus obtaining non-negative simple functions $f_n^+ \nearrow f^+$ and $f_n^- \nearrow f^-$ with the uniform convergence statement satisfied, so the functions $f_n := f_n^+ - f_n^-$ are as desired.

To define f_n, we cut off at $\leq 2^n$ and split $[0, 2^n]$ into piece of size 2^{-n}, so 2^{2n} pieces. Let:

$$\begin{aligned}
A_k &:= f^{-1}((2^{-n} \cdot k, 2^{-n} \cdot (k+1)]) \\
B_k &:= f^{-1}((2^{-n} \cdot k, \infty])
\end{aligned}$$



so $B_0 \supseteq B_1 \supseteq \ldots$ and $A_k = B_{k+1} \setminus B_k$. Then we set:

$$f_n = \sum_{k=0}^{2^{2n-1}} (2^{-n} \cdot k) \mathbb{1}_{A_k} = \sum_{i=1}^{2^{2n}} 2^{-k} \mathbb{1}_{B_k}$$

Note that $f_n \nearrow f$ since the sets $X_n := f^{-1}([0, 2^{-n}]) \nearrow X \setminus \{x \in X : f(x) = \infty\}$ so $f_n \to f$ on $\bigcup_{n\in\mathbb{N}} X_n$ and also, on $X_\infty := \{x \in X : f(x) = \infty\}$ then $f_n|_{X_\infty} = 2^n$, so $f_n|_{X_\infty} \nearrow \infty = f|_{X_\infty}$.

If f is bounded on $B \subseteq X$, then $B \subseteq X_n$ for large enough $n \in \mathbb{N}$ and

$$\|f_n|_{X_n} - f|_{X_n}\|_u \leq 2^{-n} \to 0$$

□

Definition 2.3.9. For $f \in L^+(X, \mathcal{B})$, define:

$$\int f\, d\mu := \sup\left\{\int s\, d\mu : 0 \leq s \leq f, s \in S(X, \mathcal{B})\right\}$$

Proposition 2.3.3. Let $f, g \in L^+(X, \mathcal{B})$. Then :

(a) $f \leq g \implies \int f\, d\mu \leq \int g\, d\mu$

(b) $\int a \cdot f\, d\mu = a \cdot \int f\, d\mu\ \forall a \geq 0$

(c) $\int f\, d\mu = 0 \iff f = 0$ a.e.

Caution 2.3.1. $\int f + g\, d\mu = \int f\, d\mu + \int g\, d\mu$ because we don't know whether every simple function $s \leq f + g$ splits $s = \hat{f} + \hat{g}$ into such a simple function $\hat{f} \leq f, \hat{g} \leq g$. To prove this, we need to replace sup with limits of sequences.

Theorem 2.3.2 (Monotone convergence theorem (MCT))**.** Let $f_n, f \in L^+(X, \mathcal{B})$. If $f_n \nearrow f$ then $\int f_n\, d\mu \nearrow \int f\, d\mu$.

Proof. WLOG we may assume that $f > 0$ by restricting to $\{x \in X : f(x) > 0\}$.

Because $f_n \leq f$, we have $\int f_n\, d\mu \leq \int f\, d\mu$ and $\int f_n\, d\mu$ is increasing.

Hence the limit $\lim_{n\to\infty} \int f_n\, d\mu$ exists, and we need to show that:

$$\lim_{n\to\infty} \int f_n\, d\mu \geq \int f\, d\mu$$

Fix $0 \leq s \leq f$. We aim to show that $\lim_{n\to\infty} \int f_n\, d\mu \geq \int s\, d\mu$. For this, it suffices to fix $\varepsilon > 0$ and show that:

$$\lim_{n\to\infty} \int f_n\, d\mu \geq \int (1-\varepsilon) s\, d\mu$$

since

$$\int (1-\varepsilon) s\, d\mu = (1-\varepsilon) \int s\, d\mu \nearrow \int s\, d\mu \text{ as } \varepsilon \to 0$$

Note that $(1-\varepsilon)s < f$. Then define $X = \bigcup_{n\in\mathbb{N}} X_n$ where $X_n := \{x \in X : f_n(x) > (1-\varepsilon)s\}$. By the monotonicity of the integral,

$$\int f_n\, d\mu \geq \int_{X_n} f_n\, d\mu \geq \int_{X_n} (1-\varepsilon) s\, d\mu = \mu_{(1-\varepsilon)s}(X_n)$$

Note that since $X_n \nearrow X$, $\mu_{(1-\varepsilon)s} \nearrow \mu_{(1-\varepsilon)s}(X)$ by monotone convergence for measures. Hence,

$$\lim_{n\to\infty} \int f_n \, d\mu \geq \mu_{(1-\varepsilon)s}(X) = \int (1-\varepsilon)s \, d\mu$$

□

Corollary 2.3.1 (countably additivity of integrals)**.** The integral on $L^+(X, \mathcal{B})$ is countably additive, i.e.

$$\int \sum_{n\in\mathbb{N}} f_n \, d\mu = \sum_{n\in\mathbb{N}} \int f_n \, d\mu$$

Proof. First we show finitely aditivity. Let $f, g \in L^+(X, \mathcal{B})$, and $(f_n), (g_n)$ sequences of simple functions such that $f_n \nearrow f, g_n \nearrow g$. By MCT on $f_n + g_n \nearrow f + g$,

$$\begin{aligned}\int f + g \, d\mu &= \lim_{n\to\infty} \int (f_n + g_n) \, d\mu \\ &= \lim_{n\to\infty} \left(\int f_n \, d\mu + \int g_n \, d\mu \right) \\ &= \lim_{n\to\infty} \int f_n \, d\mu + \lim_{n\to\infty} \int g_n \, d\mu\end{aligned}$$

Again by MCT on $f_n \nearrow f, g_n \nearrow g$,

$$\int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu$$

For an infinite sum $f := \sum_{n\in\mathbb{N}} f_n$, observe that

$$\sum_{n<N} f_n \nearrow_N \sum_{n\in\mathbb{N}} f_n$$

So by MCT,

$$\lim_{N\to\infty} \int \sum_{n<N} f_n \, d\mu = \int \sum_{n\in\mathbb{N}} f_n$$

By finite additivity,

$$\lim_{N\to\infty} \int \sum_{n<N} f_n \, d\mu = \lim_{N\to\infty} \sum_{n<N} \int f_n \, d\mu = \sum_{n\in\mathbb{N}} \int f_n \, d\mu$$

□

Corollary 2.3.2 (functions defining measures)**.** Each $f \in L^+(X, \mathcal{B})$ defines a measure μ_f on $(X, \mathcal{B})$ by setting:

$$\mu_f(B) = \int \mathbb{1}_B f \, d\mu = \int_B f \, d\mu$$

Proof. Clearly, $\mu_f(\emptyset) = 0$. Then let $B = \bigsqcup_{n\in\mathbb{N}} B_n$ be a disjoint union of sets $B_n \in \mathcal{B}$. Then $\mathbb{1}_B = \sum_{n\in\mathbb{N}} \mathbb{1}_{B_n}$, so $\mathbb{1}_B \cdot f = \sum_{n\in\mathbb{N}} \mathbb{1}_{B_n} \cdot f$ so by countably additivity,

$$\mu_f(B) = \int \mathbb{1}_B f \, d\mu = \int \sum_{n\in\mathbb{N}} \mathbb{1}_{B_n} \cdot f = \sum_{n\in\mathbb{N}} \int \mathbb{1}_{B_n} \cdot f = \sum_{n\in\mathbb{N}} \mu_f(B_n)$$

□

Theorem 2.3.3 (Fatou's Lemma)**.** Let $\{f_n\} \subseteq L^+(X, \mathcal{B})$. Then:

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu$$

Proof. By definition:

$$\liminf_{n\to\infty} f_n = \lim_{n\to\infty} \inf_{i\ge n} f_i$$

then observe that

$$\inf_{i\ge n} f_i \nearrow \liminf_{n\to\infty} f_n$$

Hence by MCT,

$$\int \liminf_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int \inf_{i\ge n} d\mu$$

Also observe that by monotonicity of the integral, since $\forall m \ge n$, $\inf_{i\ge n} f_i \le f_m$:

$$\int \inf_{i\ge n} f_i \, d\mu \le \int f_m \, d\mu \quad \forall m \ge n$$

hence,

$$\int \inf_{i\ge n} f_i \, d\mu \le \inf_{m\ge n} \int f_m \, d\mu$$

so

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \lim_{n\to\infty} \inf_{m\ge n} \int f_m \, d\mu = \liminf_{n\to\infty} \int f_n \, d\mu$$

□

Example 2.3.1 (strict inquality in Fatou)**.** For $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, let $f_n := \mathbb{1}_{[n,n+1)}$, so $f_n \to 0$ but $\int f_n \, d\mu = 1$. Hence:

$$\int \liminf_{n\to\infty} f_n \, d\mu = \int \lim_{n\to\infty} f_n \, d\mu = \int 0 \, d\mu = 0 < \liminf_{n\to\infty} \int f_n \, d\mu = \lim_{n\to\infty} 1 = 1$$

Definition 2.3.10. Let $(X, \mathcal{B}, \mu)$ be a complete measure space (i.e. $\mathcal{B} = \text{Meas}_\mu$). A μ-measurable function is said to be integrable w.r.t. μ if $\int |f| \, d\mu < \infty$. In this case,

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$$

Remark 2.3.11. *If f is complex-valued, then $\int f \, d\mu := \int \text{Re} f \, d\mu + i \int \text{Im} f \, d\mu$*

We denote the space of μ-integrable functions by $L^1(X, \mathcal{B}, \mu) =: L^1(X, \mu)$. Commonly, $L^1(X, \mu)$ denotes the quotient of the set of all integrable functions by this equivalence relation $f = g$ a.e. We will typically use $L^1(X, \mu)$ as literally the space of all μ-integrable functions with the understand that we could have used the quotient. Define a (pseudo) norm on $L^1(X, \mu)$ by:

$$\|f\|_1 := \int |f| \, d\mu$$

Observation 2.3.12. $(L^1(X,\mu), \|\cdot\|_1)$ *is a (pseudo) normed vector space, i.e.* $\forall f, g \in L^1(X,\mu)$*,*

(1) $\|f\|_1 \geq 0$ *and* $\|f\|_1 = 0 \iff f = 0$ *a.e.*

(2) $\|\alpha f\|_1 = |\alpha| \cdot \|f\|_1 \quad \forall \alpha \in \mathbb{R}$

(3) *Triangle inequality:* $\|f+g\|_1 \leq \|f\|_1 + \|g\|_1$

Works for α complex as well.

It is a vector space regardless if whether or not you take the quotient or literal interpretation.

Proof. Only the proof for (3) is necessary. Note that $\|f+g\|_1 = \int |f+g|\,d\mu \leq \int |f| + |g|\,d\mu = \int |f|\,d\mu + \int |g|\,d\mu = \|f\|_1 + \|g\|_1$ □

As always, a (pseudo) normed vector space is also a (pseudo) metric space by $d_1(f,g) = \|f-g\|_1$. Thus,

Definition 2.3.13. Let $(f_n) \subseteq L^1(X,\mu)$ be a sequence of functions. We say that the sequence converges in the L^1-norm to $f \in L^1(X,\mu)$ and write $f_n \to_{L_1} f$ if it converges in the (pseudo) metric d_1, i.e. $f_n \to f$ if $d_1(f_n, f) := \|f - f_n\| \to 0$ as $n \to \infty$.

Observation 2.3.14. *Note that convergence* $f_n \to f$ *in* L_1 *also implies that:*

$$\int f_n\,d\mu \to \int f\,d\mu$$

Proof.

$$\left|\int d\mu - \int f_n\,d\mu\right| = \left|\int f - f_n\,d\mu\right| \leq \int |f - f_n|\,d\mu = \|f - f_n\|$$

□

Notation 2.3.1. *When* μ *is the counting measure on* X*, then* $\ell^1(X) = L^1(X, \text{counting measure})$

Example 2.3.2.

(a) $\ell^1(\mathbb{N})$:= absolutely summable sequences. Indeed, for $f \in \ell^1(\mathbb{N})$, and μ the counting measure on $\mathbb{N}$,

$$\int f\,d\mu = \sum_{n\in\mathbb{N}} f(n)$$

(b) Let $d \in \mathbb{N}$. Then $\ell^1(d)$ is $\mathbb{R}^d$ with the 1-norm.

$$\|f\| := \sum_{i<d} |f(i)|$$

2.4 *Pointwise vs. L^1 convergence*

Examples of disagreement: All examples below are in $L^1(\mathbb{R}, \lambda)$.

Example 2.4.1.

(a) Let $f_n := \mathbb{1}_{[n,n+1)}$ and $f := 0$. Then $f_n \to 0$ pointwise but $\forall n \in \mathbb{N}, \int f_n \, d\mu = 1$, hence $\|f_n - f\|_1 = \int |f_n| \, d\mu \not\to 0$ so $f_n \not\to 0$ in L_1.

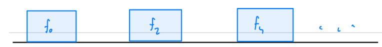


Figure 2.1: Example (a).

(b) Let $f_n := n \cdot \mathbb{1}_{(0,1/n]}$ for $n \in \mathbb{N}^+$. Then $f_n \to 0$ pointwise but $\int f \, d\mu = 1$ for each $n \in \mathbb{N}$ so $f_n \not\to_{L_1} 0$

(c) Example of $f_n \to_{L_1} 0$ but not at any point. In general, we have

$$f_n := \mathbb{1}_{[\frac{j}{2^k}, \frac{j+1}{2^k}]}$$

with $n = 2^k + j$ and $0 \le j < 2^k$, $k = \lceil \log_2(n+1) \rceil$
$\|f_n - 0\|_1 = \int f \, d\lambda = \frac{1}{2^k} \to 0$, so $f_n \to_{L_1} 0$. But $f_n(x)$ diverges for each $x \in [0,1]$ since $f_n(x)$ has infinitely many 0s and 1s.

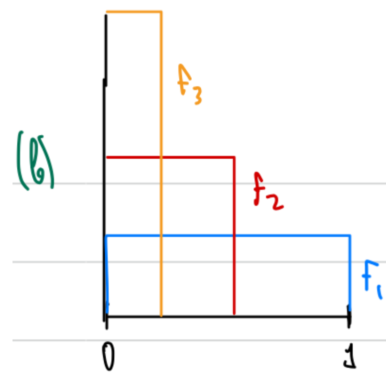


Figure 2.2: Example (b).

We will discus example (c) and it's fix later after introducing convergence in measure, but we can fix examples (a)-(b) now. In these examples, there is no integrable $g \ge 0$ dominating all of the $|f_n|$. Indeed, in (a) such a g has to be $\mathbb{1}_{[0,\infty)}$ and in (b), $\ge \max\{f_n\} \sim \frac{1}{x}$.

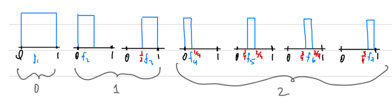


Figure 2.3: Example (c).

Theorem 2.4.1 (Dominated convergence theorem (DCT))**.** Let f_n, f be μ-measurable functions, with $f_n \to f$ a.e. If there is a dominating function $g \in L^1(X, \mu)$ i.e. $g \ge 0$ and $|f_n| \le g$ for all $n \in \mathbb{N}$, then $f_n, f \in L^1(X, \mu)$ and $\int f_n \, d\mu \to \int f \, d\mu$ as $n \to \infty$. In fact, $f_n \to_{L_1} f$ as $n \to \infty$.

Proof. This condition $|f_n| \le g$ implies $|f| \le g$ a.e., hence $f_n, f \in L^1(X, \mu)$ First, apply Fatou's lemma to $g + f_n$ to obtain that,

$$\int \liminf_{n\to\infty} (g + f_n) \, d\mu \le \liminf_{n\to\infty} \int (g + f_n) \, d\mu$$

Since g is independent of n, we can take it out of the lim inf. Moreover, since $f_n \to f$ a.e., if $E := \{x \in X : f_n(x) \to f(x)\}$, $\int \liminf f_n \, d\mu = \int_E \liminf f_n \, d\mu = \int_E f \, d\mu = \int f \, d\mu$. Hence,

$$\int g + \int f \, d\mu \le \int g + \liminf_{n\to\infty} \int f_n \, d\mu$$

Applying Fatou to $g - f_n = g + (-f_n)$ and doing the same yields

$$\int g - \int f \, d\mu \le \int g + \liminf_{n\to\infty} \int -f_n \, d\mu$$

Then since $\liminf_{n\to\infty} -a_n = -\limsup_{n\to\infty} a_n$,

$$\int g - \int f \, d\mu \le \int g - \limsup_{n\to\infty} \int f_n \, d\mu$$

Cancelling g from both sides and combining, we get,

$$\limsup_{n\to\infty} f_n \, d\mu \le \int f \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu$$

Hence the limit exists and $\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu$.

To get L_1 convergence, apply the same process to $f - f_n$ which is dominated by $2g$. Hence,

$$\lim_{n\to\infty} \int |f_n - f| \, d\mu = \int \lim_{n\to\infty} |f_n - f|$$

But $\int |f_n - f| \, d\mu = \|f_n - f\|$ and $f_n \to f$ a.e. implies $|f_n - f| = 0$ a.e, hence

$$\|f_n - f\|_1 \to_{L_1} 0 \text{ as } n \to \infty$$

□

2.5 L^1 as a (pseudo) metric space

We analyze dense functions in L^1 as well as whether the L^1 metric is complete

Proposition 2.5.1. Simple functions are dense in L^1 (in the L^1 metric).

Proof. If $f \in L^1(X, \mu)$, then $\exists$ a sequence of simple functions (s_n) such that $s_n \to f$ (pointwise, hence also a.e.). Hence by DCT, $s_n \to_{L_1} f$ as $n \to \infty$. □

Proposition 2.5.2. If a measure space $(X, \mathcal{B}, \mu)$ is countably generated (i.e. Meas_μ is separable), there is a countable collection of simple functions dense in $L^1(X, \mu)$ in the L^1-metric. In particular $L^1(X, \mu)$ is separable.

Proof. HW □

Example 2.5.1.

(a) $L^1(\mathbb{R}^d, \lambda)$ is separable since $\mathbb{R}^d$ is second-countable and $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$

(b) $L^1(A^{\mathbb{N}}, \mu)$ where A is finite and μ is a Bernoulli measure is separable since $A^{\mathbb{N}}$ is second countable and $\mathcal{B} = \mathcal{B}(A^{\mathbb{N}})$.

(c) $\ell^1(X)$ for X a set is separable $\iff$ X is countable. Indeed, if X is countable then $\mathcal{B} = \mathcal{P}(X)$ is generated by the singletons. If X is uncountable, then the family $\{\mathbb{1}_{\{x\}}\}_x \in X$ is discrete and uncountable, hence $\ell^1(X)$ is not separable.

(d) Any σ-finite Borel measure μ on a second countable measure space X has that $L^1(X, \mu)$ is separable.

We now discuss completeness of $L^1(X, \mu)$ for which we first give a criterion of completeness for normed vector spaces.

Definition 2.5.1. Let $(V, \|\cdot\|)$ be a normed vector space with metric $d(f,g) = \|f - g\|$. For a series $\sum_{n\in\mathbb{N}} f_n$ of elements of V, we say it

- converges in norm if $\exists f \in V$ s.t. $\sum_{n<N} f_n \to f$ as $N \to \infty$ in norm, i.e. $\left\|f - \sum_{n\le N} f_n\right\| \to 0$ as $N \to \infty$. In this case, we write $\sum_{n\in\mathbb{N}} f_n = f$
- converges absolutely if $\sum_{n\in\mathbb{N}} \|f_n\| < \infty$

Lemma 2.5.1 (Criteria for completeness)**.** Let $(V, \|\cdot\|)$ be a normed vector space. Then V is complete (every Cauchy sequence converges) $\iff$ every absolutely convergent series converges in norm.

Proof. MATH 255. □

Theorem 2.5.1. For any measure space $(X, \mathcal{B}, \mu)$, the space $L^1(X, \mathcal{B}, \mu)$ is complete.

Proof. Let $\sum_{n\in\mathbb{N}} f_n$ be an absolutely convergent series of functions, i.e. $\sum_{n\in\mathbb{N}} \|f_n\| < \infty$.

Then $g = \sum_{n\in\mathbb{N}} |f_n|$ dominates the sequences of partial sums $\sum_{n<N} f_n$ and $g \in L^1(X,\mu)$ since $\|g\| = \int \sum_{n\in\mathbb{N}} |f_n|\, d\mu = \sum_{n\in\mathbb{N}} \int |f_n|\, d\mu = \sum_{n\in\mathbb{N}} \|f_n\| < \infty$.

In particular, the function $\sum_{n\in\mathbb{N}} |f_n|$ is finite a.e, hence converges a.e. Let $f(x) = \sum_{n\in\mathbb{N}} f_n(x)$. Then: (1) f is measurable because it is the limit of partial sums (which are measurable) and (2) $f \in L^1(X,\mu)$ since $|f| = |\sum_{n\in\mathbb{N}} f_n| \le \sum_{n\in\mathbb{N}} |f_n| = g$. Thus we can apply DCT to the sequence $(\sum_{n<N} f_n)$ and ge that these partial sums converge to f in the L_1-metric. □

2.6 *Properties of integrable functions*

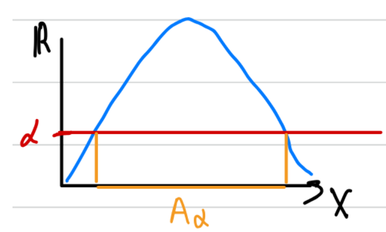
Theorem 2.6.1 (Chebyshev's/Markov's inequality)**.** For each $f \in L^1(X,\mu)$ and $\alpha \in [0,\infty]$, we have:

$$\mu\left(\{x \in X : |f(x)| > \alpha\}\right) \le \frac{1}{\alpha} \|f\|_1$$

Proof. Let $A_\alpha = \{x \in X : |f(x)| > \alpha\}$. Then

$$\|f\|_1 = \int |f|\, d\mu \ge \int_{A_\alpha} |f|\, d\mu \ge \int_{A_\alpha} \alpha\, d\mu = \alpha\mu(A_\alpha)$$

□



Corollary 2.6.1. For each $f \in L^1(X,\mu)$, it's support $\text{supp}(f) = \{x \in X : f(x) \ne 0\}$ is σ-finite for μ, i.e.

$$\text{supp}(f) = \bigcup_{n\in\mathbb{N}} B_n$$

where B_n is measurable and $\mu(B_n) < \infty$

Proof. Let $B_n = \{x \in X : |f(x)| > \frac{1}{n}\}$ for all $n \in \mathbb{N}^+$. Then $\text{supp}(f) = \bigcup_{n\in\mathbb{N}^+} B_n$ and $\mu(B_n) \leq n \|f\|_1 < \infty$ by Chebyshev. □

This has nothing to do with the 99% lemma.

Theorem 2.6.2 (99% boundedness)**.** For any $f \in L^1(X, \mu)$, and $\varepsilon > 0$, there is a measurable set $X' \subseteq X$ such that $f|_{X'}$ is bounded and

$$\int_{X\setminus X'} |f|\, d\mu \leq \varepsilon$$

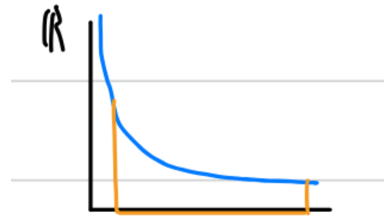


Figure 2.4: Illustration of 99% boundedness: If the blue curve $f \in L^1(X, \mu)$, we can find a yellow set X' such that f is bounded on X' and $\int_{X\setminus X'} |f|\, d\mu$ is arbitrary small..

recall that for any $f \in L^1(X, \mu)$, we get a measure defined $\mu_{|f|}(A) = \int_A |f|\, d\mu$

Proof. Let $X_n = \{x \in X : |f(x)| \leq n\}$. WLOG, change f so that $|f| < \infty$ everywhere. Then $X = \bigcup_{n\in\mathbb{N}} X_n$ so by increasing monotone convergence for $\mu_{|f|}$ we have that :

$$\lim_{n\to\infty} \mu_{|f|}(X_n) = \mu_{|f|}(X) < \infty$$

so

$$\mu_{|f|}(X \setminus X_n) \to 0 \text{ as } n \to \infty$$

so for large enough $n \in \mathbb{N}$, $\mu(X \setminus X_n) < \varepsilon$, so for this large enough n we can set $X' = X_n$. □

Definition 2.6.1. Let $(X, \mathcal{B})$ be a measurable space, μ, ν be measures on $(X, \mathcal{B})$. We way ν is absolutely continuous w.r.t. μ and write $\nu \ll \mu$ if every μ-null set is also ν-null.

Example 2.6.1. For any $f \in L^1(X, \mu)$, the measure $\mu_{|f|}$ is finite and $\mu_{|f|} \ll \mu$ because if $B \subseteq X$ is μ-null, then $\mu_{|f|}(B) = \int_B |f|\, d\mu = 0$.

The following justifies the terminology of absolutely continuous.

Proposition 2.6.1. Let ν, μ be measures on a measurable space $(X, \mathcal{B})$. If ν is finite, then:

$$\nu \ll \mu \iff \forall\varepsilon > 0 \exists\delta > 0 \text{ such that } \mu(B) < \delta \implies \nu(B) < \varepsilon \forall B \in \mathcal{B}$$

Proof.

$\impliedby$ Is trivial since $0 < \varepsilon$.

Let $B \in \mathcal{B}$. Assume $\mu(B) = 0$. Let $\varepsilon > 0$ then $\nu(B) < \varepsilon$ since $\mu(B) < \delta$. Hence $\nu(B) = 0$

$\implies$ We prove the contrapositive. Assume that $\exists\varepsilon > 0$ such that $\forall\delta > 0 \exists B_\delta \in \mathcal{B}$ such that $\mu(B_\delta) < \delta$ but $\nu(B_\delta) \geq \varepsilon$. By the application of Borel-Cantelli applied to the collection

$$\mathcal{C} := \{B \in \mathcal{B} : \nu(B) \geq \varepsilon\}$$

and μ, this collection admits a μ-almost vanishing sequence, i.e. a decreasing sequence of sets $(B_n) \subseteq \mathcal{C}$ with $\mu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = 0$. But because ν is a finite measure, decreasing monotone convergence applies to (B_n) and ν yielding $\nu\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \lim_{n\to\infty} \nu(B_n) \geq \varepsilon > 0$ so $\nu \not\ll \mu$ □

Corollary 2.6.2 (absolute continuity of integrable functions)**.** For any $f \in L^1(X, \mu)$, we have that for each $\varepsilon > 0 \exists \delta > 0$ such that whenever $\mu(B) < \delta$ we have $\int_B |f| \, d\mu < \varepsilon$.

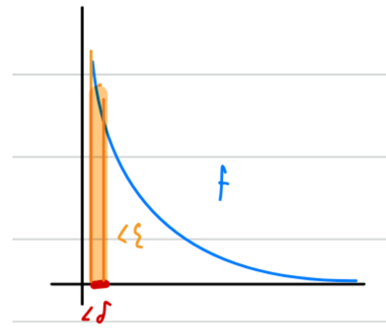


Figure 2.5: Illustration of Corollary 2.6.2.

Proof. This property also follows directly from 99% boundedness (HW) □

2.7 *Convergence in measure*

RECALL THE example of L^1 convergence but not pointwise. There are subsequences of (f_n) that converge to 0 a.e., e.g. (f_{2^n}). It turns out that this is a general phenomenon: every L_1-convergent sequence admits a subsequence converging a.e. To prove this, we need an "intermediate" notion of convergence.

Definition 2.7.1. Let (X, μ) be a measure space, $f, g : X \to \overline{\mathbb{R}}$ be μ-measurable functions, and $\alpha > 0$. Define:

$$\Delta_\alpha(f, g) = \{x \in X : |f(x) - f(g)| \geq \alpha\}$$
$$\delta_\alpha(f, g) = \mu(\Delta_\alpha(f, g))$$

Remark 2.7.2. *For μ-measurable sets $A, B \subseteq X$, $\Delta_\alpha(\mathbb{1}_A, \mathbb{1}_B) = A \triangle B$ and $d\mu(A, B) = \delta_\alpha(f, g)$ for all $0 < \alpha \leq 1$.*

While δ_α doesn't satisfy the triangle inequality, it is "kind of a pseudometric." through the following relation:

Proposition 2.7.1 (additive triangle inequality for δ)**.** For all $\alpha, \beta > 0$ and $f, g, h : X \to \overline{\mathbb{R}}$ μ-measurable,

- $\Delta_{\alpha+\beta}(f, h) \subseteq \Delta_\alpha(f, g) \cup \Delta_\beta(g, h)$
- $\delta_{\alpha+\beta}(f, h) \leq \delta_\alpha(f, g) + \delta_\beta(g, h)$

Proof. For each $x \in X$,

$$\begin{aligned} x \in \Delta_{\alpha+\beta}(f, h) &\iff |f(x) - h(x)| \geq \alpha + \beta \\ &\implies |f(x) - g(x)| + |g(x) - h(x)| \geq \alpha + \beta \\ &\implies |f(x) - g(x)| \geq \alpha \text{ or } |g(x) - h(x)| \geq \beta \\ &\implies x \in \Delta_\alpha(f, g) \cup \Delta_\beta(g, h) \end{aligned}$$

□

Definition 2.7.3. For a measure space (X, μ), μ-measurable functions (f_n) and f, we say that (f_n) converges in measure to f, denoted $f_n \to_\mu f$ if $\delta_\alpha(f_n, f) \to_n 0 \; \forall \alpha > 0$, i.e. that

$$\lim_{n \to \infty} \delta_\alpha(f_n, f) \quad \forall \alpha > 0$$

Example 2.7.1.

(a) Let $f_n := \mathbb{1}_{[n,n+1)}$. Then $f_n \to 0$ pointwise but not in measure and not in L_1

(a) Let $f_n := n\mathbb{1}_{(0,1/n)}$. Then $f_n \to 0$ pointwise but not in L^1, since $\int f_n \, d\lambda = n$. However, $f_n \to_\lambda 0$ because for each α, $\delta_\alpha(f_n, 0) = 1/n$ for all large enough n.

(a) $f_n \to_\lambda 0$ because $\delta_\alpha(f_n, 0) = 2^{-k}$ if n is in the k-th group

The moral is that convergence in measure doesn't detect how badly the functions f_n differ from the limit, but instead, just detect how large this place they differ is.

Proposition 2.7.2. For a measure space (X, μ) $f_n \to_{L_1} f \implies f_n \to_\mu f$

Proof. Assume $f_n \to_{L_1} f$. Fix $\alpha > 0$. By Chebyshev's inequality,

$$\delta_\alpha(f_n, f) \leq \frac{1}{\alpha} \|f - f_n\| \to 0 \text{ as } n \to \infty$$

□

Lemma 2.7.1 (switch of quantifiers trick)**.** Let (X, μ) be a finite measure space. Let $P_n \in X$ be an increasing sequence of μ-measurable sets. For every $\varepsilon > 0$,

$$\forall x \in X \exists n \in \mathbb{N} x \in P_n \implies \exists n \in \mathbb{N} \forall^\mu_{-\varepsilon} x \in X, x \in P_n$$

where $\forall^\mu_{-\varepsilon}$ means $\forall x \in X \setminus Z$ for some μ-measurable $Z, \mu(Z) < \varepsilon$.

Proof. Let $\varepsilon > 0$. By LHS, $\bigcup_{n \in \mathbb{N}} P_n = X$, so $\mu(X) = \lim_{n\to\infty} \mu(P_n)$, hence for large enough $n \in \mathbb{N}$, $\mu(P_n) \geq \mu(X) - \varepsilon$. □

Proposition 2.7.3. In a finite measure space (X, μ), $f_n \to f$ a.e. $\implies f_n \to_\mu f$.

Proof. Discarding a null set, assume that $f_n \to f$ everywhere. So $\forall \alpha > 0$, we have $\forall x \in X \exists N \in \mathbb{N} : |f_n(x) - f(x)| < \alpha$. For any $\varepsilon > 0$, we switch the quantifiers and get:

$$\exists N \in \mathbb{N}\, \forall^\mu_{-\varepsilon} x \in X, \forall n \geq N, |f_n(x) - f(x)| < \alpha \text{ i.e.}$$
$$\exists N \in \mathbb{N}\, \forall n \geq N \forall^\mu_{-\varepsilon} x \in X, , |f_n(x) - f(x)| < \alpha$$

so $\exists N \in \mathbb{N} \forall n \geq N \ \delta_\alpha(f_n, f) < \varepsilon$ which means $\lim_{n\to\infty} \delta_\alpha(f_n, f) = 0$. □

By a switch of quantifiers trick, one can also prove Egorov's theorem about almost uniform convergence (HW)

Now let's study convergence in measure.

Proposition 2.7.4 (almost uniqueness of limit)**.** In a measure space (X, μ), if $f_n \to_\mu f$ and $f_n \to_\mu g$ then $f = g$ a.e.

Proof. For each $\alpha > 0$, we have $\delta_\alpha(f, g) \leq \delta_{\alpha/2}(f, f_n) + \delta_{\alpha/2}(f_n, g)$. Taking the limit $n \to \infty$, we obtain that, $\delta_\alpha(f, g) = 0$. Since this holds $\forall \alpha > 0$, we have that $\{x \in X : |f(x) - g(x)| > 0\} = \bigcup_{n \in \mathbb{N}^+} \Delta_{1/n}(f, g)$ and the latter set is μ-null. Hence $f = g$ a.e. □

Definition 2.7.4. Call a sequence (f_n) Cauchy in measure if $\forall \alpha > 0$, $\delta_\alpha(f_n, f_m) \to 0$ as $\min\{n, m\} \to \infty$

$\min\{n, m\} \to \infty$ is equivalent to the perhaps more familiar $\exists N \in \mathbb{N} \forall m, n \geq N \ldots$

Proposition 2.7.5.

(a) If $f_n \to_\mu f$ then (f_n) is Cauchy

(b) If (f_n) is Cauchy and admits a subsequence $f_{n_k} \to_\mu f$ as $k \to \infty$ then $f_n \to_\mu f$ as $n \to \infty$

Proof. HW □

Theorem 2.7.1 (completeness of convergence in measure)**.** Every sequence (f_n) which is Cauchy in measure converges in measure, i.e. $\exists \mu$-measurable f s.t. $f_n \to_\mu f$. Moreover, there is a subsequence (f_{n_k}) such that $f_{n_k} \to f$ a.e.

Proof. Note that by part (b) of the previous proposition we may restrict to any subsequence and show that it converges to get the convergence of f_n.

Claim ((a))**.** We may assume WLOG that $\delta_{2^{-n}}(f_n, f_{n+1}) \leq 2^{-n}$ by restricting to a subsequence

Proof. Define (n_k) recursively: let $n_1 = 0$. Once n_{k-1} is chosen, choose $n_k > n_{k-1}$ such that $\delta_{2^{-k}}(f_{n_k}, f_n) \leq 2^{-k}$ for all $n \geq n_k$. Such an n_k exists by the Cauchy assumption with $\alpha := 2^{-k}$. □

We now show that for a.e. $x \in X$, $(f_n(x))$ is Cauchy:

Claim ((b))**.** If $x \in B_N := \bigcup_{n \geq N} \Delta_{2^{-n}}(f_n, f_{n+1})$, then for all $m \geq n \geq N$, $|f_n(x) - f_m(x)| \leq 2^{-(n-1)}$. In particular, $|f_n(x) - f_m(x)| \to 0$ as $n \to \infty$

Proof.

$$|f_n(x) - f_m(x)| \leq \sum_{i=1}^{m-1} |f_i(x) - f_{i+1}(x)| \leq \sum_{i=1}^{m-1} 2^{-i} \leq \sum_{i=n}^{\infty} 2^{-i} = 2^{-(n-1)}$$

Note that

$$\mu(B_N) \leq \sum_{n \geq N} \delta_{2^{-n}}(f_n, f_{n+1}) \leq \sum_{n \geq N} 2^{-n} = 2^{-(N-1)} < \infty$$

so by Borel-Cantelli, a.e. $x \in X$ is eventually not in B_N, i.e. $\exists N \in \mathbb{N}$ such that $x \notin \bigcup_{n \geq N} B_N = B_N$ (where the inequality is because B_N is decreasing) Thus, by Claim (b), $((f_n))$ is Cauchy and let $f(x)$ denote the limit. The function $f \to \overline{\mathbb{R}}$ (defined a.e.) is μ-measurable being a a.e. pointwise limits of μ measurable functions f_n. It remains to show that $f_n \to_\mu f$. To this end fix $\alpha > 0$ and choose $N \in \mathbb{N}$ so that $2^{-N-2} \leq \alpha$. Then $\Delta_\alpha(f_N, f) \subseteq 2_{-(N-2)}(f_N, f)$ and by Claim 2, if $x \notin B_N$ then $|f_N(x) - f(x)| = \lim_{m\to\infty} |f_N(x) - f_m(x)| \leq 2^{-(N-1)} < 2^{-(N-2)}$, so $x \notin \Delta_{2^{-(N-2)}}(f_N, f)$. In other words, $\Delta_{2^{-(N-2)}}(f_N, f) \subseteq B_N$, so

$$\delta_\alpha(f_N, f) \leq \delta_{2^{-(N-2)}}(f_N, f) \leq \mu(B_N) \leq 2^{-(N-1)} \to 0 \text{ as } N \to \infty$$

Thus $f_N \to_\mu f$ as $N \to \infty$. □

□

Corollary 2.7.1. In any measure space (X, μ), if $f_n \to_\mu f$ then there exists a subsequence (f_{n_k}) such that $f_{n_k} \to f$ a.e. In particular, if $f_n \to_{L_1} f$, then $f_{n_k} \to f$ a.e. for some subsequence (f_{n_k}).

Proof. Since $f_{n_k} \to f$, (f_n) is Cauchy in measure, so for some μ-measurable function g, $f_n \to_\mu g$ and $f_{n_k} \to g$ a.e. for some subsequence $(f_{n_k}$. But limits are almost unique in convergence in measure, so $g = f$ a.e. □

2.8 *Product measures*

Let $(X, \mathcal{I}), (Y, \mathcal{I})$ be measurable spaces. Recall that $\mathcal{I} \otimes \mathcal{J}$ denotes the σ-algebra generated by rectangles, i.e. sets of the form $I \times J, I \in \mathcal{I}, J \in \mathcal{J}$.

Theorem 2.8.1. For any measure spaces $(X, \mathcal{I}, \mu), (Y, \mathcal{J}, \nu)$, there is a measure ρ on $(X \times Y, \mathcal{I} \otimes \mathcal{J})$ such that $\rho(I \times J) = \mu(I)\nu(J)$ for all measurable rectangles $I \times J$. If μ and ν are σ-finite, then such a measure ρ is unique and is called the product of μ and ν, denoted by $\mu \times \nu$.

Proof. Let $\mathcal{A}$ denote the algebra generated by all rectangles, hence $\mathcal{A}$ consists of finite disjoint unions of rectangles. To prove the theorem, it is sufficient to show that the formula

$$\rho(I \times J) := \mu(I) \cdot \nu(J)$$

defines a premeasure on $\mathcal{A}$ and then applying Carathéodory's theorem. Note that if μ, ν are σ finite by $X = \bigsqcup_{n\in\mathbb{N}} X_n, Y = \bigsqcup_{m\in\mathbb{N}} Y_m$, $\mu(X_i) < \infty, \nu(Y_i) < \infty$, then ρ is σ-finite by $X \times Y = \bigsqcup_{n\in\mathbb{N}} \bigsqcup_{m\in\mathbb{N}} X_n \times Y_m$, $\rho(X_n \times Y_m) = \mu(X_n)\nu(Y_m) < \infty$.

For this, it suffices to show that:

$$\mu(I) \times \nu(J) = \rho(I \times J) = \sum_{n \in \mathbb{N}} \rho(I_n \times J_n) = \sum_{n \in \mathbb{N}} \mu(I) \times \nu(J)$$

whenever

$$I \times J = \bigsqcup_{n \in \mathbb{N}} I_n \times J_n$$

Assuming the latter, note that $\mathbb{1}_{I \times J} = \sum_{n \in \mathbb{N}} \mathbb{1}_{I_n \times J_n}$ and $\mathbb{1}_{U \times V}(x, y) = \mathbb{1}_U(x)\mathbb{1}_V(y)$. For each $x \in X$, integrating over Y we get:

$$\begin{aligned}
\mathbb{1}_I(x)\nu(J) = \int_Y \mathbb{1}_I(x)\mathbb{1}_J(y)\, d\nu(y) &= \int_Y \sum_{n \in \mathbb{N}} \mathbb{1}_{I_n}(x)\mathbb{1}_{J_n}(y)\, d\nu(y) \\
&= \sum_{n \in \mathbb{N}} \int_Y \mathbb{1}_{I_n}(x)\mathbb{1}_{J_n}\, d\nu \qquad \text{(MCT)} \\
&= \sum_{n \in \mathbb{N}} \mathbb{1}_{I_n}\nu(J_n)
\end{aligned}$$

Now integrating over X,

$$\begin{aligned}
\mu(I)\nu(J) = \int_X \mathbb{1}_I(x)\nu(J)\, d\mu(x) &= \int_X \sum_{n \in \mathbb{N}} \mathbb{1}_{I_n}\nu(J_n)\, d\mu(x) \\
&= \sum_{n \in \mathbb{N}} \int_X \mathbb{1}_{I_n}\nu(J_n)\, d\mu(X) \\
&= \sum_{n \in \mathbb{N}} \mu(I_n)\nu(J_n)
\end{aligned}$$

□

Remark 2.8.1. *By the uniqueness part in Carathéodory's Theorem, the Lebesgue measure λ_d on $\mathbb{R}^d$ that we defined is equal to $\lambda_1 \times \ldots \times \lambda_1$ where λ_1 is the Lebesgue measure on $\mathcal{B}(\mathbb{R})$. Similarly, $\lambda_3 = \lambda_2 \times \lambda_1$.*

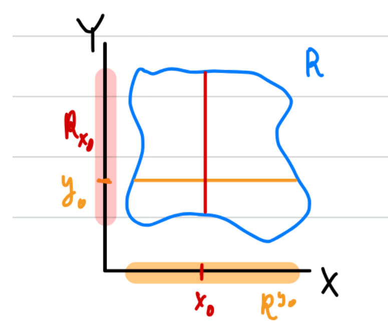


Figure 2.6: The fibres of R..

2.9 *Fubini-Tonelli Theorem*

Definition 2.9.1 (fibres of a set and function)**.** Let X, Y, Z be sets and $x_0 \in X, y_0 \in Y$. Then:

(a) For a set $R \subseteq X \times Y$, call the sets:

$$\begin{aligned}
R_{x_0} &= \{y \in Y : (x_0, y) \in R\} \\
R^{y_0} &= \{x \in X : (x_0, y) \in R\}
\end{aligned}$$

the vertical and horizontal fibres of $\mathcal{R}$ at x_0, y_0 respectively.

(b) For a function $f : X \times Y \to Z$, call the functions:

$$\begin{aligned}
f_{x_0} &: Y \to Z, y \mapsto f(x_0, y) \\
f^{y_0} &: X \to Z, x \mapsto f(x, y_0)
\end{aligned}$$

the vertical and horizontal fibres of f at x_0, y_0 respectively.

The following is one of the most useful theorems in math:

Theorem 2.9.1 (Fubini-Tonelli)**.** Let $(X, \mathcal{I}, \mu), (Y, \mathcal{J}, \nu)$ be σ-finite measure spaces. Let $f : X \times Y \to \overline{\mathbb{R}}$ be a $\mu \times \nu$-measurable function. Then:

(a) $f_x : Y \to \overline{\mathbb{R}}$ and $f^y : X \to \overline{\mathbb{R}}$ are ν-measurable and μ-measurable functions for μ-a.e. $x \in X$ and ν-a.e. $y \in Y$.

(b) Tonelli. If $f \geq 0$, then:

  (i) The functions $x \mapsto \int_Y f_x \, d\mu$ and $y \mapsto \int_X f^y \, d\mu$ are μ and ν-measurable.

  (ii) $\int_X \int_Y f_x(y) \, d\nu(y) \, d\mu(x) = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \int_X f^y(x) \, d\mu(x) \, d\nu(x)$

(c) Fubini. If f is $\mu \times \nu$-integrable, then:

  (i) The functions $x \mapsto \int_Y f_x \, d\mu$ and $y \mapsto \int_X f^y \, d\mu$ are μ and ν-measurable, and μ and ν-integrable.

  (ii) $\int_X \int_Y f_x(y) \, d\nu(y) \, d\mu(x) = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \int_X f^y(x) \, d\mu(x) \, d\nu(x)$

Remark 2.9.2. *All hypotheses are necessary, see examples in HW*

Remark 2.9.3. *Usually, we apply Tonelli to $|f|$ to verify $\int_{X \times Y} f \, d\mu \times \nu < \infty$, then apply Fubini to f*

Example 2.9.1. For sets N, M (e.g. $= \mathbb{N}$) and counting measure μ and ν on N and M respectively, the theorem says something we already know. Let $(a_{nm})_{n \in N, m \in M}$ be a matrix of reals.

(a) Tonelli: if $a_{nm} \geq 0$ for all $n \in N, m \in M$, then:

$$\sum_{n \in N} \sum_{m \in M} a_{nm} = \sum_{(n,m) \in N \times M} = \sum_{m \in M} \sum_{n \in N} a_{nm}$$

(b) Fubini: if a_{nm} is absolutely summable, i.e. $\sum_{(n,m) \in N \times M} |a_{nm}| < \infty$, then:

$$\sum_{n \in N} \sum_{m \in M} a_{nm} = \sum_{(n,m) \in N \times M} = \sum_{m \in M} \sum_{n \in N} a_{nm}$$

We prove a version of Fubini-Tonelli with $\mathcal{I} \otimes \mathcal{J}$-measurability and leave it as HW to deduce the $\mu \times \nu$-measurable version.

Proposition 2.9.1. Let $(X, \mathcal{I}), (Y, \mathcal{J})$ be measurable spaces.

(i) For all $\mathcal{I} \otimes \mathcal{J}$-measurable $R \subseteq X \times Y$, i.e. $R \in \mathcal{I} \otimes \mathcal{J}$, we have all fibres $R_x.R^y$ are $\mathcal{J}$- and $\mathcal{I}$-measurable, i.e. $R_x \in \mathcal{J}$ and $R^y \in \mathcal{I}$.

(ii) For all $\mathcal{I} \otimes \mathcal{J}$-measurable functions $f : X \times Y \to Z$ where $(Z, \mathcal{K})$ is some measurable space, all fibres f_x, f^y are $\mathcal{J}$- and $\mathcal{I}$-measurable

Proof.

(a) Let $\mathcal{C} := \{R \in \mathcal{I} \otimes \mathcal{J} : \forall x \in X, y \in Y, R_x \in \mathcal{J}, R^y \in \mathcal{I}\}$. By definition, $\mathcal{C}$ contains all rectangles $I \times J$ since $(I \times J)_x = \begin{cases} J & x \in I \\ 0 & \text{o.w.} \end{cases}$ and same for $y \in Y$. Also, $\mathcal{C}$ is closed under complement and countable unions because these operations commute with taking fibres, i.e. $(\bigcup_{n\in\mathbb{N}} R_n)_x = \bigcup (R_n)_x$ and $(R^C)_x = (R_x)^C$.

(b) Given $W \in \mathcal{K}'x$, we verify that $f_x^{-1}(W)$ and $f^{-1^y}(W)$ are $\mathcal{J}$- and $\mathcal{I}$-measurable because preimages commute with taking fibres: $f^{-1})_x(W) = (f^{-1}(W))_x$, same for y. Then $f_x^{-1}(W) = (f^{-1}(W))_x \in \mathcal{J}$ by (a).

□

Proposition 2.9.2 (Fubini-Tonelli for sets)**.** Let $(X, \mathcal{I}, \mu)$ and $(Y, \mathcal{J}, \nu)$ be σ-finite measure spaces. Let $R \subseteq \mathcal{I} \otimes \mathcal{J}$ i.e. $\mathbb{1}_R$ is $\mathcal{I} \otimes \mathcal{J}$-measurable. Then:

(i) The functions $x \mapsto \nu(R_x)$ and $y \mapsto \mu(R^y)$ are $\mathcal{I}$- and $\mathcal{J}$-measurable, respectively

(ii) $\int_X \nu(R_x) d\mu(x) = \mu \times \nu(R) = \int_Y \mu(R^y)\, d\nu(y)$

Proof. First, we may assume WLOG that μ and ν are finite by the usual argument of writing $X = \bigsqcup_{n\in\mathbb{N}} X_n$ and $Y = \bigsqcup_{m\in\mathbb{N}}$ for $X_n \in \mathcal{I}, Y_m \in \mathcal{J}$ of finite measure so $X \times Y = \bigsqcup_{n,m\in\mathbb{N}} X_n \times Y_m$ and $R = \bigsqcup_{n,m\in\mathbb{N}} R \cap (X_n \times Y_m)$, and using closure under limits for (i) and countable additivity for (ii). Let $\mathcal{C} := \{R \in \mathcal{I} \otimes \mathcal{J} :$ R satisfies (i) and (ii)$\}$. We aim to show that $\mathcal{C}$ is a σ-algebra containing the algebra $\mathcal{A}$ generated by rectangles $I \times J$ with $I \in \mathcal{I}$, $J \in \mathcal{J}$. □

Claim. $\mathcal{C}$ contains rectangles

Proof. Indeed, for $I \in \mathcal{I}, J \in \mathcal{J}$, the function $x \mapsto \nu((I \times J)_x) = \nu(J)\mathbb{1}_I(x)$ is clearly measurable, and the same for the y fibres. For (ii), observe that:

since it is constant times an indicator function

$$\int_X \nu((I \times J)_x)\, d\mu(X) = \int_X \nu(J)\mathbb{1}_I(x)\, d\mu(X) = \nu(J)\mu(I) = \mu \times \nu(I \times J)$$

and the same argument works for the y fibres. □

Claim. $\mathcal{C}$ is closed under (finite) disjoint unions

Proof. This is because the measures are finitely additive, measurable functions are closed under addition and the integral is linear. □

Let $A, B \in \mathcal{C}$, then for (i), note that $x \mapsto \nu((A \sqcup B)_x) = \nu(A_x \sqcup B_x) = \nu(A_x) + \nu(A_y)$ which is a sum of the measurable (by assumption) functions $x \mapsto \nu(A_x)$ and $x \mapsto \nu(A_y)$. The same argument works for the y fibres. For (ii), note that $\int_X (\nu((A \sqcup B)_x))\, d\mu(x) = \int_X \nu(A_x)\, d\mu(x) + \int_X \nu(B_x)\, d\mu(x) = \mu \times \nu(A) + \mu \times \nu(B) = \mu \times \nu(A \sqcup B)$ and the same for y-fibres.

Thus, $\mathcal{A} \subseteq \mathcal{C}$ because each element of $\mathcal{A}$ is a finite disjoint union of rectangles.

Using closedness of limits measurable functions and the monotone convergence theorem, we also get that $\mathcal{C}$ is closed under countable disjoint unions, and using finiteness of μ, ν and $\mu \times \nu$, we can also deduce that $\mathcal{C}$ is closed under complements. But to conclude that $\mathcal{C}$ is a σ-algebra, we still need to show closedness under finite intersections (to disjointify a countable union, this is needed), which is hard to show because measures of the intersection is not expressible by the measures of the sets. Instead, we appeal to the monotone class lemma, to be proved below, and see that this is enough to verify that $\mathcal{C}$ is closed under countable increasing unions and countable decreasing intersections, i.e. is a monotone class, because then, $\mathcal{C} \supseteq \langle \mathcal{A} \rangle_\sigma = \mathcal{I} \otimes \mathcal{J}$, hence $\mathcal{C} = \mathcal{I} \otimes \mathcal{J}$.

Claim. $\mathcal{C}$ is a monotone class.

Proof. Let $\{A_k\} \subseteq \mathcal{C}$ be a collection of sets in $\mathcal{C}$. For increasing unions, suppose $A_0 \subseteq A_1 \subseteq \ldots$ and let $A = \bigcup_{k \in \mathbb{N}} A_k$. Define $f_k : x \mapsto \nu((\bigcup_{j \leq k} A_k)_x) = \nu((A_k)_x)$ and $f : x \mapsto \nu((A)_x)$. Each f_k is measurable since $A_k \in \mathcal{C}$. By the monotone convergence theorem, $\lim_{k\to\infty} f_k = \lim_{k\to\infty} \nu((\bigcup_{j \leq k} A_k)_x) = \nu(A_x) = f$. Hence f is measurable since it is a pointwise limit of measurable functions. Moreover, $f_k \nearrow f$ by the monotonicity of measure ($f_k(x) = \nu((A_k)_x) \leq \nu((A_{k+1})_x) = f_{k+1}(x)$), hence by the monotone convergence theorem $\int f_k(x)\, d\mu(x) \to \int f\, d\mu(x)$. Hence,

$$\int_X \nu\left(\left(\bigcup_{k \in \mathbb{N}} A_k\right)_x\right) d\mu(x) = \lim_{k\to\infty} \int_X \nu((A_k)_x)\, d\mu(x) = \lim_{k\to\infty} \mu \times \nu(A_k)$$

Then by the monotone convergence theorem again,

$$\lim_{k\to\infty} \mu \times \nu(A_k) = \mu \times \nu(A)$$

completing the proof of (ii). The same arguments work for y fibres.

For decreasing unions, suppose $A_0 \supseteq A_1 \supseteq \ldots$ and let $A = \bigcap_{k \in \mathbb{N}} A_k$. Define $f_k : x \mapsto \nu((\bigcap_{j \leq k} A_k)_x) = \nu((A_k)_x)$. Each f_k is measurable since $A_k \in \mathcal{C}$. Moreover by the monotone convergence theorem, since ν is finite, $\lim_{k\to\infty} f_k = \lim_{k\to\infty} \nu((\bigcap_{j \leq k} A_k)_x) = \nu(A_x)$ where f is the function $x \mapsto \nu((A)_x)$. Hence f is measurable since it is a pointwise limit of measurable functions. Moreover, each f_k is dominated by f_0 and $\int |f_0(x)|\, d\mu(x) = \mu \times \nu(A_0) < \infty$, hence by the dominated convergence theorem,

$$\int_X \nu\left(\left(\bigcup_{k \in \mathbb{N}} A_k\right)_x\right) d\mu(x) = \lim_{k\to\infty} \int_X \nu((A_k)_x)\, d\mu(x) = \lim_{k\to\infty} \mu \times \nu(A_k)$$

Then by decreasing monotone convergence theorem again, noting that $\mu \times \nu$ is finite,

$$\lim_{k\to\infty} \mu \times \nu(A_k) = \mu \times \nu(A)$$

$\mathcal{C}$ is closed under countable disjoint unions. Let $\{A_k\}_{k \in \mathbb{N}} \subseteq \mathcal{C}$ be pair- Since $A \setminus B = A \cap B^c$ wise disjoint. Define f_k by $f_k(x) = \nu((\bigsqcup_{j \leq k} A_k)_x) = \sum_{j \leq k} \nu((A_k)_x)$. Each f_k is a sum of measurable functions (since each $A_k \in \mathcal{C}$), hence measurable. Moreover, $f_k \to f$ where f is the function $x \mapsto \nu((\bigsqcup_{k \in \mathbb{N}} A_k)_x) = \sum_{k \in \mathbb{N}} \nu((A_k)_x)$. Hence f is measurable. Moreover, $\int_X (\bigsqcup_{k \in \mathbb{N}} A_k)_x\, d\mu(x) = \sum_{k \in \mathbb{N}} \int \nu((A_k)_x)\, d\mu(x) = \sum_{k \in \mathbb{N}} \mu \times \nu(A_k) = \mu \times \nu(\bigsqcup_{k \in \mathbb{N}} A_k)$. The same arguments work for y fibres. □

completing the proof of (ii). The same arguments work for y fibres. □

Definition 2.9.4 (monotone class)**.** A collection $\mathcal{C}$ of subsets of a set X is called a monotone class if it is closed under countable increasing unions and countable decreasing intersections. For $\mathcal{A} \subseteq \mathcal{P}(X)$, the monotone class generated by $\mathcal{A}$ is the $\subseteq$-least monotone class containing $\mathcal{A}$, i.e. the intersection of all monotone classes containing $\mathcal{A}$.

Lemma 2.9.1 (Monotone Class Lemma)**.** Let $\mathcal{C} \subseteq \mathcal{P}(X)$ be a monotone class. If $\mathcal{C}$ contains an algebra $\mathcal{A}$, then $\mathcal{C} \supseteq \langle \mathcal{A} \rangle_\sigma$.

Proof. By shrinking $\mathcal{C}$, we may assume WLOG that $\mathcal{C}$ is the monotone class generated by $\mathcal{A}$. We then show that $\mathcal{C} = \langle \mathcal{A} \rangle_\sigma$. For this, it is enough to show that $\mathcal{C}$ is an algebra, since a countable union can be rewritten as an increasing countable union of finite unions which is in $\mathcal{C}$ because it is a monotone class and $\mathcal{C}$ is closed under finite unions (to be proved below). □

since if $\mathcal{C}' \subseteq \mathcal{C}$ is a monotone class and $\mathcal{C}' \supseteq \langle \mathcal{A} \rangle_\sigma$, then $\mathcal{C} \supseteq \mathcal{C}' \supseteq \langle A \rangle_\sigma$

$\bigcup_{n\in\mathbb{N}} = \bigcup_{n\in\mathbb{N}} \bigcup_{m\leq n} C_m$

Claim. $\mathcal{C}$ is closed under complements

Proof. Let $\mathcal{S} := \{S \in \mathcal{C} : S^C \in \mathcal{C}\}$. We show that $\mathcal{S} \supseteq \mathcal{A}$ and it is a monotone class. Let $A \in \mathcal{A}$. A^C is also in $\mathcal{S}$, hence $\mathcal{A} \subseteq \mathcal{S}$. To show it is a monotone class, consider for S_n an increasing sequence of sets, $(\bigcup_{n\in\mathbb{N}} S_n)^C = \bigcap_{n\in\mathbb{N}} S_n^C$. Since $\mathcal{C}$ is closed under countable decreasing intersections, we have that $\bigcap_{n\in\mathbb{N}} S_n^C \in \mathcal{C}$, hence $(\bigcup_{n\in\mathbb{N}} S_n) \in \mathcal{S}$. The same argument works for S_n a decreasing sequence of sets, since $(\bigcap_{n\in\mathbb{N}} S_n)^C = \bigcup_{n\in\mathbb{N}} S_n$ which is in $\mathcal{C}$. □

Then since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$, then $\mathcal{C} \supseteq \mathcal{S}$ so $\mathcal{C} = \mathcal{S}$.

Claim. $\mathcal{C}$ is closed under finite unions

Proof. For each $U \in \mathcal{C}$, let $\mathcal{S}(U) := \{V \in \mathcal{C} : U \cup V \in \mathcal{C}\}$. We show that $\mathcal{S}(U) \supseteq \mathcal{A}$ and is a monotone class, hence then $\mathcal{S}(U) = \mathcal{C}$. Monotone class: suppose $\{V_n\} \supseteq \mathcal{S}(U)$ is an increasing sequence of sets. Then $U \cup \bigcup_{n\in\mathbb{N}} V_n = \bigcup_{n\in\mathbb{N}} U \cup V_n$ hence it is in $\mathcal{S}(U)$ because $\mathcal{C}$ is closed under countable increasing unions. Now suppose $\{V_n\}$ is decreasing so then $U \cup \bigcap_{n\in\mathbb{N}} V_n = \bigcap_{n\in\mathbb{N}} U \cup V_n \in \mathcal{C}$ since $\mathcal{C}$ is closed under countable decreasing intersections. $\mathcal{A} \subseteq \mathcal{S}(U)$: Fix $A \in \mathcal{A}$. Note that $A \in \mathcal{S}(U) \iff U \in \mathcal{S}(A)$. We in fact show $\mathcal{S}(A) = \mathcal{C}$ which implies $U \in \mathcal{S}(A)$ since $U \in \mathcal{C}$ by assumption. We have already showed that $S(A)$ is a monotone class. Also, $\mathcal{A} \subseteq \mathcal{S}(A)$ because $\mathcal{A}$ is closed under finite unions and $\mathcal{C} \supseteq \mathcal{A}$. □

We can't immediately use the typical argument of defining a set that has the properties we want and showing that this set is everything since now we have to deal with pairs of sets. To get around this issue, we apply the argument to a set that is a union with a fixed set and do a little bit of extra work.

by showing that $\mathcal{S}(A)$ is a monotone class containing $\mathcal{A}$.

we already showed it for every $U \in \mathcal{C}$ above and $A \in \mathcal{A} \subseteq \mathcal{C}$ by assumption

$\mathcal{S}(A) = \{V \in \mathcal{C} : V \cup A \in \mathcal{C}\}$ so for every $A' \in \mathcal{A}$, $A' \cup A \in \mathcal{A} \subseteq \mathcal{C}$ so $A' \in \mathcal{S}(A)$.

Theorem 2.9.2 (Fubini-Tonelli for $\mathcal{I} \otimes \mathcal{J}$)**.** Let $(X, \mathcal{I}, \mu)$ and $(Y, \mathcal{J}, \nu)$ be σ-finite measure spaces. Let $f : X \times Y \to \overline{\mathbb{R}}$ be an $\mathcal{I} \otimes \mathcal{J}$-measurable function. Then:

(a) $f_x : Y \to \overline{\mathbb{R}}$ and $f^y : X \to \overline{\mathbb{R}}$ are $\mathcal{J}$- and $\mathcal{I}$-measurable for all $x \in X, y \in Y$.

(b) Tonelli. If $f \geq 0$, then

(i) $x \mapsto \int_Y f_x \, d\nu$ and $y \mapsto \int_X f_y \, d\mu$ and $\mathcal{I}$- and $\mathcal{J}$- measurable.

(ii) $\int_X \int_Y f_x(y) \, d\nu(y) \, d\mu(x) = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \int_X f^y(x) \, d\mu(x) \, d\nu(y)$

(c) Fubini. If f is $\mu \times \nu$-integrable, then:

(i) $x \mapsto \int_Y f_x \, d\nu$ and $y \mapsto \int_X f_y \, d\mu$ and $\mathcal{I}$- and $\mathcal{J}$- measurable and integrable.

(ii) $\int_X \int_Y f_x(y) \, d\nu(y) \, d\mu(x) = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \int_X f^y(x) \, d\mu(x) \, d\nu(y)$

Proof. We have already proved (a) and we know (b) for indicator functions, which implies (b) and (c) for simple functions by the linearity of integral. To conclude (b) for all $f \geq 0$, write f as an increasing limit of nonnegative simple functions and in (i) use the closedness of measurable functions under pointwise limits and MCT. For (ii), just use MCT three times. Finally, for (c), write $f = f_+ - f_-$. so f_+ and f_- are $\mu \times \nu$-integrable and apply (b) to f^+ and f^- individually, observing that the finiteness of $\int f_\pm \, d\mu \times \nu$ implies that the functions $x \mapsto \int f_x \, d\nu$ and $y \mapsto \int f^y \, d\mu$ are finite a.e. at (c) for f by linearity of integral. $\square$

2.10 *Infinite products*

We already learned how to construct the product of two measure spaces, and hence also finitely many measure spaces by induction. We would like to extend this construction to infinite products. Firstly let's learn/recall product topology:

Definition 2.10.1 (product topology)**.** Let $\{X_i\}_{i\in I}$ be a collection of topological spaces for I some index set (e.g. $I = \mathbb{N}$ or $I = \mathbb{R}$). Then the product of these spaces is:

$$\prod_{i\in I} X_i = \{(x_i)_{i\in I} : x_i \in X_i \forall i \in I\}$$

and the product topology on this set is the one generated by the open cylinders:

$$[i_1 \mapsto U_{i_1}, \dots, i_k \mapsto U_{i_k}] := \{(x_i)_{i\in I} \in \prod_{i\in I} X_i : x_{i_j} \in U_{i_j} \forall j = 1, \dots, k\}$$

for some finitely many indices $i_1, \dots, i_k$ and open sets $U_{i_j} \in X_{i_j}$.

In other words, in the product topology, the open sets are the arbitrary unions of open cylinders.

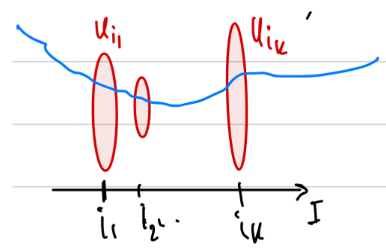


Figure 2.7: The open sets in the product topology are the blue "snakes" that pass through all the open sets $U_{i_1}, \dots, U_{i_k}$, represented by red "hoops"..

The product topology is the right topology to consider (as opposed to the box topology where nothing converges)

Theorem 2.10.1 (Tychonoff's theorem)**.** The product of compact spaces is compact. This is equivalent to the Axiom of Choice. In fact, the non-emptiness of the product is the statement of the Axiom of Choice.

Now given measure spaces $(X_i, \mathcal{B}_i, \mu_i)_{i\in I}$, we would like to build a measure on the σ-algebra generated by the cylinders:

$$[i_1 \mapsto B_{i_1}, \dots, i_k \mapsto B_{i_k}] := \{(x_i)_{i\in I} \in \prod_{i\in I} X_i : x_{i_j} \in B_{i_j} \forall j = 1, \dots, k\}$$

where $B_{i_j} \in \mathcal{B}_{i_j}$ satisfy that:

$$\mu([i_1 \mapsto B_{i_1}, \dots, i_k \mapsto B_{i_k}]) = \prod_{1\le j\le k} \mu_{i_j}(B_{i_j}) \cdot \prod_{i\in I, i\notin\{i_1,\dots,i_k\}} \mu_i(X_i) \quad (*)$$

Remark 2.10.2. *It only makes sense/suffices to consider probability spaces when dealing with infinite products. For simplicity, consider when $I = \mathbb{N}$. Indeed,*

(i) For $(*)$ to even make sense, we need the limit $\prod_{i\in\mathbb{N}} \mu(X_i) = \lim_{n\to\infty} \prod_{i<n} \mu(X_i)$ to exist, so we must assume we are dealing with probability spaces.

(ii) If the product of tails $\prod_{i\ge n} \mu_i(X_i)$ is 0 or ∞ for all n sufficiently large, then the restriction of the hypothetical measure μ to the algebra $\mathcal{A}$ generated by cylinders would only take values 0 and ∞, so it would be useless (basically equal to the measure that is 0 when any component is null and ∞ otherwise). In particular, $0 < \mu_n(X_n) < \infty$ for all large enough $n \in \mathbb{N}$, we can focus on defining μ in the case $0 < \mu_n(X_n) < \infty$ for all $n \in \mathbb{N}$

(iii) Since $0 < \prod_{n\in\mathbb{N}} \mu_n(X_n) < \infty$, after μ is proved to exist, we could scale it considering $\mu / \prod_{n\in\mathbb{N}} \mu_n(X_n)$ instead, in other words, scale each μ_n to $\mu_n/\mu_n(X_n)$ so we consider the product of probability measures.

So now we let $(X_i, \mathcal{B}_i, \mu_i)$, $i \in I$ be probability spaces. Let $\mathcal{A}$ be the algebra generated by the cylinders, hence $\mathcal{A}$ consists of finite disjoint unions of cylinders (because intersections of two cylinders is a cylinder and a complement of a cylinder is a finite disjoint union of cylinders). We can define a function μ on cylinders as above, i.e.

$$\mu([i_1 \mapsto B_{i_1}, \ldots, i_k \mapsto B_{i_k}]) = \prod_{1 \le j \le k} \mu_{i_j}(B_{i_j})$$

By the usual refining partitions and switching the sums argument, this extends to a well-defined finitely additive probability measure on $\mathcal{A}$. We would like to apply Carathéodory's theorem to this and get a unique measure on $\langle \mathcal{A} \rangle_\sigma$ by extending μ . But to apply Carathéodory's theorem, we need to show that μ is countably additive on $\mathcal{A}$. By finite additivity, μ is automatically countably superadditive and we need to show countable subadditivity.

i.e. is a premeasure

This was proved first by Kolmogorov for $I = \mathbb{N}$ and $(X_i, \mathcal{B}_i, \mu_i) = ([0,1], \mathcal{B}([0,1]), \lambda)$ in 1933; this is known in probability theory as Kolmogorov consistency theorem. This was later extended by Doob in 1938 to countable products of arbitrary Borel probability measures on $\mathbb{R}$. Both of these proofs rely on regularity and tightness, which are inherently topological notions, restricting the types of measure spaces they handle. However, in 1943 Kakutani published a topology-free completely abstract and short proof of the general case for all I and all probability spaces (4 pages, really two pages). Here we will present a slight extension of Doob's theorem.

Theorem 2.10.2 (Kolmogorov consistency)**.** A countable product of standard probability spaces exists (and is unique)

Proof. Let $(X_n, \mathcal{B}_n, \mu_n)$ be standard probability spaces, hence it is isomorphic to the Lebesgue measure on $[0,1]$, so WLOG assume that each $X_n = [0,1], \mathcal{B}_n = \mathcal{B}([0,1]), \mu_n = \lambda$. To show countable subadditivity on the algebra $\mathcal{A}$ generated by cylinders, it is enough to take a countable partition

$$C = \bigsqcup_{m \in \mathbb{N}} C_m$$

of a cylinder $C := B_0 \times \ldots \times B_{\ell-1} \times [0,1]^{\mathbb{N} \setminus \ell}$ and cylinders $C_m := B_0^{(m)} \times \ldots \times B_{\ell-1}^{(m)} \times [0,1]^{\mathbb{N} \setminus \ell_m}$. We shrink C a bit to make it compact and we extend each C_m to make it open, as follows:

- By tightness of λ, for each $i < l$ there exist compact sets $K_i \subseteq B_i$ such that $\lambda(K_i) \approx \lambda(B_i)$ so that setting $C' := K_0 \times \ldots \times K_1 \times [0,1]^{\mathbb{N} \setminus \ell}$, we have $\mu(C') \approx_{\varepsilon/2} \mu(C)$. This C' is compact being a closed subset of the compact space $[0,1]^{\mathbb{N}}$ by Tychonoff's theorem.

- By strong regularity of λ, Corollary 1.16.1 for each $i \leq l_m$ there exist open sets $U_i^{(m)} \supseteq B_i^{(m)}$ such that $\lambda(U_i^{(m)}) \approx \lambda(B_i^{(m)})$, so that setting $\tilde{C}_m := U_i^{(m)} \times \ldots \times U_{l_m - 1}^{(m)} \times [0,1]^{\mathbb{N} \setminus l_m}$ we have that $\mu(\tilde{C}_m) \approx_{\varepsilon/2 \cdot \frac{1}{2^{m+1}}} \mu(C_m)$

Thus, using finite additivity of μ, we get that $\mu(C) \leq_{\varepsilon/2} \mu(C') \leq \sum_{m \in \mathbb{N}} \mu(\tilde{C_m}) \leq_{\varepsilon/2} \sum_{m \in \mathbb{N}} \mu(C_m)$ which shows countable subadditivity since $\varepsilon > 0$ is arbitrary. $\square$

3 Differentiation of measures

3.1 Orthogonality and equivalence

We now let μ, ν be measures on the same measurable space $(X, \mathcal{B})$, and try to understand their relationship.

Recall absolute continuity: $\nu \ll \mu$ if every μ-null set is ν-null. Also recall that when ν is finite, then $\nu \ll \mu$ is equivalent to:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ such that } \mu(B) < \delta \implies \nu(B) < \varepsilon \quad \forall B \in \mathcal{B}$$

Definition 3.1.1 (equivalent, orthogonal)**.** Say that ν and μ are:

- equivalent if they have the same null sets, i.e. $\nu \ll \mu$ and $\mu \ll \nu$. This property is denoted $\nu \sim \mu$.
- orthognal if $X = X_\nu \sqcup X_\mu$ where $\nu(X_\mu) = 0$ and $\mu(X_\nu) = 0$. We denote this by $\nu \perp \mu$. This property is also called μ and ν are mutually singular.

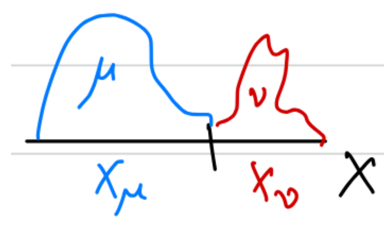


Figure 3.1: $\nu \perp \mu$.

Example 3.1.1 (examples of orthogonality)**.**

(a) if $\nu = \delta_{x_0}$ is a Dirac measure and μ is atomless then $\nu \perp \mu$ because $X = \{x_0\} \bigsqcup (X \setminus \{x_0\})$

(b) Let ν_p be the pushforward of the Bernoulli(p) measure from $2^{\mathbb{N}}$ to $\{0,1\}$. by the usual ternary representation map. ϕ is a homeomorphism between $2^{\mathbb{N}}$ and the standard Cantor set $C \subseteq [0,1]$, so $\nu_p(C) = 1$ and $\nu_p([0,1] \setminus C) = 0$. Thus $\nu_p \perp \lambda$ since $\lambda(C) = 0$.

(c) Let (X, μ) be a measure space and $f \geq 0$ be a μ-integrable function on X. Partition X into μ-measurable sets X_0, X_1 and set $f_i := \mathbb{1}_{X_i} \cdot f$ so then $\mu_{f_0} \perp \mu_{f_1}$.

Theorem 3.1.1 (Lebesgue decomposition theorem)**.** For ν, μ σ-finite measures on a measurable space $(X, \mathcal{B})$, we have a decomposition $\nu = \nu_0 + \nu_1$ where $\nu_0 \ll \mu$ and $\nu_1 \perp \mu$. In fact, $X = X_0 \sqcup X_1$ with $X_0, X_1 \in \mathcal{B}$ such that $\nu|_{X_0} \ll \mu|_{X_0}$ and $\nu(X_1) = 0$ (in particular, $\nu|_{X_1} \perp \mu|_{X_1}$)

Proof. HW □

Corollary 3.1.1. For σ-finite measures ν, μ on a measurable space $(X, \mathcal{B})$ there is a partition $X = X_0 \sqcup X_1$ into sets in $\mathcal{B}$ such that $\mu|_{X_0} \sim \nu|_{X_0}$ and $\nu|_{X_1} \perp \mu|_{X_1}$.

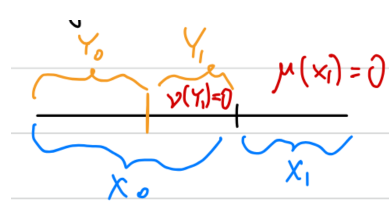


Figure 3.2: $\mu|_{Y_0} \sim \nu|_{Y_0}$ and $\mu|_{Y_1 \cup X_1} \perp \nu|_{Y_1 \cup X_1}$.

Theorem 3.1.2 (Radon-Nikodym)**.** Let ν, μ be σ-finite measures on $(X, \mathcal{B})$. If $\nu \ll \mu$ then there is a μ-measurable $f : X \to [0, \infty)$ such that $\nu = \mu_f$ i.e. $\nu(B) = \int_B f \, d\mu$ for all $B \in \mathcal{B}$.

To prove this elegantly, we will consider differences $\nu - \mu$, which give the abstract concept of signed measures. Thus, we take a detour into signed measures, prove a decomposition theorem for them, from which we will derive the Radon-Nikodym theorem.

3.2 *Signed measures*

Definition 3.2.1 (signed measure)**.** Let $(X, \mathcal{B})$ be a measurable space. A signed measure on $(X, \mathcal{B})$ is a function $\xi : \mathcal{B} \to \overline{\mathbb{R}}$ such that:

(i) $\xi(\emptyset) = 0$

(ii) Countable additivity: $\xi(B) = \sum_{n \in \mathbb{N}} \xi(B_n)$ whenever $B, B_n \in \mathcal{B}$ and $B = \bigsqcup_{n \in \mathbb{N}} B_n$ so that if $|\xi(B)| < \infty$, then $\sum_{n \in \mathbb{N}} |\xi(B_n)| < \infty$

(iii) ξ doesn't attain at least one of $+\infty$, $-\infty$

Example 3.2.1 (examples of signed measures)**.**

(a) All measures are signed measures

(b) If ν and μ and measures on $(X, \mathcal{B})$ and one of them is finite, then $\nu - \mu$ is a signed measure.

(c) Let μ be a measure on $(X, \mathcal{B})$ and $f \in L^1(X, \mu)$, then $\mu_f := \int_B f \, d\mu$ is a finite signed measure. Note that in the case, $\mu_f = \mu_{f_+} - \mu_{f_-}$.

It turns out that the decomposition from the last example always exists.

Theorem 3.2.1 (Jordan Decomposition)**.** For any signed measure ξ on a measurable space $(X, \mathcal{B})$, we have $\xi = \nu - \mu$ for some measure ν, μ on $(X, \mathcal{B})$

We will prove a stronger version of this below, for which we need the following notions:

Definition 3.2.2 ((purely) positive/negative/null set)**.** For a signed measure ξ on a measurable space $(X, \mathcal{B})$, call a set $B \in \mathcal{B}$

- positive (resp. negative) if $\xi(B) \geq 0$ (resp. $\xi(B) \leq 0$)
- purely positive (resp. purely negative, purely null) if $\xi|_B \geq 0$ (resp. $\xi|_B \leq 0, \xi|_B = 0$). In particular, $\xi|_B$ (resp. -$\xi|_B$) is a measure

Caution 3.2.1. The union of two positive sets A, B may not be positive.

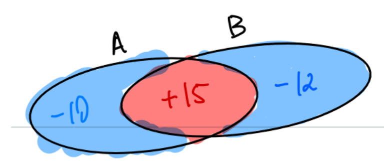


Figure 3.3: Both A and B have positive measure but $A \cup B$ has negative measure..

Theorem 3.2.2 (Hahn Decomposition)**.** For any signed measure ξ on a measurable space $(X, \mathcal{B})$, there is a partition $X = X_+ \sqcup X_-$ into purely positive and purely negative sets, so $\xi = \xi|_{X_+} - \xi|_{X_-}$ is a Jordan decomposition. Such a decomposition is unique up to purely null sets, i.e. if $X = Y_+ \sqcup Y_-$ is another such decomposition, then $Y_+ \triangle X_+$ and $Y_- \triangle X_-$ are purely null.

Proof (of uniqueness). Let $X = P \cup N = P' \cup N'$ be two Hahn decompositions for ξ. Since $P \cap N'$ is a subset of both P and N', it must be both positive and negative, hence null. Similarly, $P' \cap N$ is null. Then $P \triangle P' = (P \setminus P') \cup (P' \setminus P) = (P \cap N') \cup (P' \cap N)$ which is a union of null sets. Similarly, $N \triangle N'$ is null. □

Proof. The uniqueness is straightforward and is left as an exercise. To prove the existence, assume WLOG $\xi < \infty$. We will use $\frac{1}{2}$-exhaustion to obtain a "maximal" purely positive set X_+ and then show that $X_- := X \setminus X_+$ is purely negative. For the latter, we need the following:

Claim. Every positive set $P \in \mathcal{B}$ contains a purely positive $P_+ \subseteq P$ with $\xi(P_+) \geq \xi(P)$ (i.e. $\xi(P \setminus P_+) \leq 0$)

Proof of claim. Use $\frac{1}{2}$-exhaustion to build a "maximal" negative subset of P. Namely, define a disjoint sequence $(N_k)_{k \in \mathbb{N}}$ of negative subsets of P such that for each $k \in \mathbb{N}$,

$$-\xi(N_k) \geq \frac{1}{2}\overline{\sup}\{-\xi(N) : N \subseteq P \setminus \bigsqcup_{i<k} N_i \text{ and } \xi(N) \leq 0\}$$

where $\overline{\sup} := \min\{\sup, 1\}$.

Let $P_+ := P \setminus \bigsqcup_{k \in \mathbb{N}} N_k$ so $0 \leq \xi(P) = \xi(P_+) + \sum_{k \in \mathbb{N}} \xi(N_k) = \xi(P_+) - \sum_{k \in \mathbb{N}} |\xi(N_k)| \leq \xi(P_+)$, hence $\xi(P_+) \geq \xi(P)$ and $\sum_{k \in \mathbb{N}} |\xi(N_k)| \leq \xi(P_+) < \infty$. In particular, $\xi(N_k) \to 0$ as $k \to \infty$, so if there were a non-null negative set $N \subseteq P_+$, then $|\xi(N_k)| < \frac{1}{2} \min\{|\xi(N_k)|, 1\}$ for some $k \in \mathbb{N}$, contradicting the choice of k. □

Now we use $\frac{1}{2}$-exhaustion to get a "maximal" purely positive set, i.e. take a disjoint sequence $(P_k)_{k \in \mathbb{N}}$ of purely positive sets satisfying, for each $k \in \mathbb{N}$,

$$\xi(P_k) \geq \frac{1}{2}\overline{\sup}\{\xi(P_+) : P_+ \text{ is purely positive and } P_+ \subseteq X \setminus \bigsqcup_{i<k} P_i\}$$

Then, letting $X_+ := \bigsqcup_{k \in \mathbb{N}} P_k$, we have $\sum_{k \in \mathbb{N}} P_k = \xi(X_+) < \infty$, so $\xi(P_k) \to 0$ as $k \to \infty$, which implies that $X_- := X \setminus X_+$ is purely negative; indeed, if $P \subseteq X_-$ were such that $\xi(P) > 0$, then the claim would give a purely positive $P_+ \subseteq P$ with $\xi(P_+) \geq \xi(P) > 0$, so $\xi(P_k) < \frac{1}{2} \min\{\xi(P_+).1\}$ for some large enough $k \in \mathbb{N}$, contradicting the choice of P_k. □

Remark 3.2.3. *Note that condition (iii) in the definition of signed measure [Definition 3.2.1] says that WLOG, $\xi < \infty$, i.e. $\xi(B) < \infty$ for all $B \in \mathcal{B}$. But it's a priori possible that $\sup \xi(B) = \infty$. However, the Hahn decomposition theorem says that this isn't possible: $B \in \mathcal{B}, \xi(B) \leq \xi(X_+) < \infty$ for all $B \in \mathcal{B}$. More generally, $\xi(X_-) \leq \xi \leq \xi(X_+)$.*

We now prove a key lemma about (unsigned) measures used in the proof of the Radon-Nikodym theorem.

Lemma 3.2.1. Let ν and μ be finite measures on a measurable space $(X, \mathcal{B})$. Then either $\nu \perp \mu$ or $\nu|_Y \geq \varepsilon \cdot \mu|_Y$ for some μ-nonnull set $Y \in \mathcal{B}$ and $\varepsilon > 0$.

Proof. Let $\varepsilon_k := \frac{1}{k}$. Let $X = P_k \sqcup N_k$ be a Hahn decomposition for the signed measure $\nu - \varepsilon_k \mu$.

Case 1 If $P := \bigcup_{k \in \mathbb{N}} P_k$ is not μ-null, then $\mu(P_k) > 0$ for some k, so $Y := P_k$ and $\xi := \xi_k$ works, i.e. $\nu|_{P_k} - \xi_k \nu|_{P_k} \geq 0$, hence $\nu|_{P_k} \geq \xi_k \nu|_{P_l}$.

Case 2 If $P := \bigcup_{k \in \mathbb{N}} P_k$ is μ-null, then it remains to observe that $N := X \setminus \bigcup_{k \in \mathbb{N}} P_k = \bigcap_{k \in \mathbb{N}} N_k$ is ν-null since $\nu(N) \leq \nu(N_k) \leq \xi_k \nu(N_k) \leq \xi_k \mu(X) \to 0$ as $k \to \infty$ where we used that $\mu(X) < \infty$.

□

Theorem 3.2.3 (Lebesgue-Radon-Nikodym)**.** Let ν and μ be two σ-finite measures on a measurable space $(X, \mathcal{B})$. Then $\nu = \mu_f + \nu_0$ where $\nu_0 \perp \mu$ and $f : X \to [0, \infty]$ is a $\mathcal{B}$-measurable function. This function f is essentially unique (i.e. if $\tilde{f}$ is another such function then $f = \tilde{f}$ a.e.) and is called the Radon-Nikodym derivative of ν and μ, denoted $\frac{d\nu}{d\mu}$

Proof (of uniqueness). HW 6 □

Proof. The uniqueness is straightforward to check and is left as an exercise. By the usual argument, we may assume that μ, ν are finite. We aim to find a desired function by a maximality (greedy) argument. Let

$$\mathcal{F} := \{f : X \to [0, \infty] : f \text{ is } \mathcal{B}\text{-measurable and } \mu_f \leq \nu\}$$

if $X = \bigsqcup_{n \in \mathbb{N}} X_n = \bigsqcup_{m \in \mathbb{N}} Y_m$ where $\nu(X_n), \mu(X_m) < \infty$ then $X = \bigsqcup_{n,m \in \mathbb{N}} X_n \cap Y_m$ so $\nu(X_n \cap X_m), \mu(X_n \cap X_m) < \infty$ and it suffices to prove for the restrictions of ν, μ to $X_n \cap Y_m$.

Claim. $\exists f \in \mathcal{F}$ that satisfies $\int f \, d\mu = \sup\{\int g \, d\mu : g \in \mathcal{F}\}$

Proof of claim. $\mathcal{F} \neq \emptyset$ because $0 \in \mathcal{F}$. Also, $\mathcal{F}$ is closed under taking maximums: if $f_1, f_2 \in \mathcal{F}$, then $\max\{f_1, f_2\} \in \mathcal{F}$. Let $f_n \in \mathcal{F}$ such that $\lim_{n \to \infty} \int f_n \, d\mu = \sup\{\int g \, d\mu : g \in \mathcal{F}\}$. By replacing each f_n with $\max\{f_0, \ldots, f_n\}$, we may assume that (f_n) is increasing. Then $f_n \to f$ for some $\mathcal{B}$-measurable $f : X \to [0, \infty]$ and $\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu = \sup\{\int g \, d\mu : g \in \mathcal{F}\}$ by MCT. □

we can write $X = X_1 \sqcup X_2$ where $f_1|_{X_1} \geq f_2|_{X_1}$ and $f_2|_{X_2} \geq f_1|_{X_2}$ so $\nu|_{X_i} \geq \mu_{f_i|_{X_i}} = \mu_{\max(f_1, f_2)}|_{X_i}$

We now show that the function f from the claim works, i.e. that $\nu_0 = \nu - \mu_f$ where $\nu_0 \perp \mu$. If not, then the above lemma [Lemma 3.2.1] gives $\varepsilon > 0$ and μ-nonnull $Y \in \mathcal{B}$ with $\nu_0 \geq \varepsilon \cdot \mu|_Y$. But then $f + \varepsilon \cdot \mathbb{1}_Y \in f$ because $\mu_{f + \varepsilon \cdot \mathbb{1}_Y} = \mu_f + \varepsilon \cdot \mu|_Y \leq \mu_f + \nu_0 = \nu_1$ and $\int f + \varepsilon \cdot \mathbb{1}_Y \, d\mu > \int f \, d\mu$ contradicting the choice of f. □

Both ν_0 and $\nu - \mu_f$ are finite measures, so the lemma applies. By assumption we don't have case 1. so we are in case 2, so $\nu_0|_Y \geq \varepsilon \mu|_Y$ for some μ-nonnull set $Y \in \mathcal{B}$ and $\varepsilon > 0$.

Corollary 3.2.1. Let ν, μ be σ-finite measures on a measurable space $(X, \mathcal{B})$ and suppose that $\nu \ll \mu$. Then for each $\mathcal{B}$-measurable $g : X \to \mathbb{R}$ which is either ν-integrable or nonnegative, we have:

$$\int g \, d\nu = \int g \frac{d\nu}{d\mu} \, d\mu \quad (*)$$

Proof. $\nu \ll \mu$ implies $\nu = \mu_f$ where $f := \frac{d\nu}{d\mu}$, so for each $B \in \mathcal{B}$, we have:

$$\int \mathbb{1}_B \, d\nu = \nu(B) = \int_V f \, d\mu = \int \mathbb{1}_B f \, d\mu$$

□

Hence $(*)$ is true for indicator functions, so by the usual argument it is true for all nonnegative and ν-integrable functions.

$(*)$ is true for indicator functions, so by linearly it is true for simple functions. Then it is true for all nonnegative $\mathcal{B}$-measurable functions by the monotone convergence. Then it is true for all ν-integrable functions by linearly.

Corollary 3.2.2 (Chain rule)**.** Let ν, μ, ξ be σ-finite measures on a measurable space $(X, \mathcal{B})$. If $\nu \ll \mu$ and $\mu \ll \xi$ then $\nu \ll \xi$ and:

$$\frac{d\nu}{d\xi} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\xi} \text{ a.e.}$$

Proof. By the essential uniqueness of the Radon-Nikodym derivative, it is enough to to show that $\frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\xi}$ has the defining property, i.e. $\forall B \in \mathcal{B}$,

$$\nu(B) = \int_B \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\xi} \, d\xi$$

But by the last corollary,

$$\nu(B) = \int_B \frac{d\nu}{d\mu} \, d\mu = \int \left(\mathbb{1}_B \cdot \frac{d\nu}{d\mu} \right) d\mu = \int \left(\mathbb{1}_B \cdot \frac{d\nu}{d\mu} \right) \frac{d\mu}{d\xi} \, d\xi = \int_B \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\xi} \, d\xi$$

□

Corollary 3.2.3. If $\nu \sim \mu$ are σ-finite measures on a measurable space $(X, \mathcal{B})$, then $\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}$ a.e.

Proof. Follows from the chain rule applied to $\nu \ll \mu \ll \mu$:

$$\frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\mu} = \frac{d\nu}{d\nu} = 1 \text{ a.e.}$$

□

Remark 3.2.4. *The term derivative comes from the following example: if μ is a locally finite Borel measure on $\mathbb{R}$ with a continuously differentiable distribution $f : \mathbb{R} \to \mathbb{R}$ (i.e. a function satisfying $\mu((a, b])) = f(b) - f(a)$, then $\mu \ll \lambda$ and $\frac{d\mu}{d\lambda} = f'$ (HW)*

Remark 3.2.5. *Note that in the Radon-Nikodym theorem, if $\nu \ll \mu$ and ν is finite, then $\frac{d\nu}{d\mu}$ is μ-integrable because $\int \frac{d\nu}{d\mu} = \nu(X) > \infty$.*

Similarly, if X is a metric space and ν is a locally finite measure on it, then f is "locally integrable" which we now define.

Definition 3.2.6 (locally integrable)**.** Let X be a metric space and μ a Borel measure on X. A μ-measurable function $f : X \to \overline{\mathbb{R}}$ is called locally μ-integrable if $\mu_{|f|}$ is locally finite, i.e. each $x \in X$ admits a neighbourhood U containing x with $\int_U |f| \, d\mu = \mu_{|f|}(U) < \infty$. Denote by $\mathrm{Loc}^1(X, \mu)$ the space of these functions.

Recall that we have proved equivalences to local finiteness of measures under various conditions in X, and translate it now for functions [Proposition 1.15.2].

Proposition 3.2.1. Let X be a locally compact second countable metric space and μ a Borel measure on X. Let $f : X \to \overline{\mathbb{R}}$ be a μ-measurable functions. Then TFAE:

(1) f is locally μ-integrable

(2) $\exists X = \bigcup_{n \in \mathbb{N}} U_n, U_n$ open such that $\int_{U_n} |f| \, d\mu < \infty$

(3) $\int_K |f| \, d\mu < \infty$ for each compact $K \subseteq X$.

In particular, this holds for $X = \mathbb{R}^d$

3.3 *Radon-Nikodym derivatives w.r.t Lebesgue measure and Lebesgue differentiation*

LET λ denote the Lebesgue measure on $\mathbb{R}^d$ and let μ be another locally finite Borel measure on $\mathbb{R}^d$. Suppose $\mu \ll \lambda$ and try and find a formula for $\frac{d\mu}{d\lambda}$. Note that f is locally λ-integrable and for each ball $B_r(x)$, we have:

$$\mu(B_r(x)) = \int_{B_r(x)} f \, d\lambda$$

hence $\frac{\mu(B_r(x))}{\lambda(B_r(x))} = \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} \frac{d\mu}{d\lambda} \, d\lambda$. Since $\frac{d\mu}{d\lambda}$ measures the relative "weight" of a point of μ over λ, it is plausible that $\lim_{r \to 0} = \frac{d\mu}{d\lambda}(x)$ for a.e. $x \in \mathbb{R}^d$. Indeed, this is the case:

Theorem 3.3.1 (Lebesgue differentiation theorem)**.** For any $f \in \mathrm{Loc}^1(\mathbb{R}^d, \lambda)$,

$$\lim_{r \to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f \, d\lambda = f \text{ a.e.}$$

Equivalently, for any locally finite Borel measure μ on $\mathbb{R}^d$ with $\mu \ll \lambda$, we have,

$$\frac{d\mu}{d\lambda} = \lim_{r \to 0} \frac{\mu(B_r(x))}{\lambda(B_r(x))} \text{ a.e.}$$

We will prove this after a definition and lemmas.

Definition 3.3.1. Let $f \in \text{Loc}^1(\mathbb{R}^d, \lambda)$ and define the averaging operator of radius r A_r by:

$$A_r f := \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f \, d\lambda$$

We want to prove:

$$\lim_{r \to 0} A_r f = f \text{ a.e.}$$

Lemma 3.3.1 (Local-global bridge)**.** Let $f \in \text{Loc}^1(\mathbb{R}^d, \lambda)$. For each $r > 0$, we have:

(a) $\int f \, d\lambda = \int A_r f \, d\lambda$

(b) $\|A_r f\| \leq \|f\|_1$, i.e. A_r is an L^1-contraction

Proof. HW □

Lemma 3.3.2. If $g \in \text{Loc}^1(\mathbb{R}^d, \lambda)$ is continuous, then $\lim_{r \to 0} A_r g = g$ everywhere.

Proof.

$$\begin{aligned} |(A_r g)(x) - g(x)| &= \frac{1}{\lambda(B_r(x))} \left| \int g(y) - g(x) \, d\lambda(x) \right| \\ &\leq \frac{1}{\lambda(B_r(x))} \int |g(y) - g(x)| \, d\lambda(x) \\ &\leq \sup_{y \in B_r(x)} |g(y) - g(x)| \to 0 \text{ as } r \to 0 \end{aligned}$$

□

Proof of Lebesgue differentiation. By the usual argument it is enough to prove that the statement holds for $\mathbb{1}_{B_n(0)} \cdot f$ for all $n \in \mathbb{N}^+$. Hence fix $n \in \mathbb{N}$, so replacing f with $\mathbb{1}_{B_n}(0) \cdot f$, we may assume f is λ-integrable and forget $B_n(0)$. We aim to prove that $A^* f := \limsup_{r \to 0} A_r f = f$ a.e. since the argument for $\liminf_{r \to 0}$ would be analogous. □

The countable union of null sets is null and for each $x \in B_n(0)$, the small ball $B_r(x) \subseteq B_n(0)$ by openness and for all sufficiently small r.

Notation 3.3.1. *For a function $h : X \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, put $\{h > \alpha\} := \{x \in X : h(x) > \alpha\}$*

To show that $\{|A^* f - f| > 0\}$ is null, it is enough to show that $\{|A^* f - f| > \alpha\}$ is null for all $\alpha > 0$ by the usual argument. Fix $\alpha > 0$. Letting $g \in L^1(\mathbb{R}^d, \lambda)$ be a continuous function, we see that:

$$\begin{aligned} |A^* f - f| &= |A^* f - A^* g + A^* g - g + g - f| \\ &\leq |A^* f - A^* g| + |A^* g - g| + |g - f| \\ &= |A^* f - A^* g| + |g - f| \\ &\leq A^* |f - g| + |g - f| \end{aligned}$$

The usual argument: $\{|A^* f - f| > 0 = \bigcup_{n \in \mathbb{N}^+} \{|A^* f - f| > \frac{1}{n}\}\}$

To go from the second to the third line, the previous lemma implies $A^* g = g$ because g is continuous

To go from the third to the fourth line, observe: $|A^* f - A^* g| = |\limsup_{r \to 0} A_r f - \limsup_{r \to 0} A_r g| \leq |\limsup_{r \to 0} A_r f - A_r g| = |\limsup_{r \to 0} A_r (f - g)| = |A_r (f - g)| \leq A_r |f - g|$

Therefore, $\{|A^* f - f| > \alpha\} \subseteq \{A^*|f-g| > \alpha/2\} \cup \{|f-g| > \alpha/2\}$ so it is enough to show the last two sets have arbitrarily small measure for an appropriate choice of g. Because continuous functions are dense in $L^1(\mathbb{R}^d, \lambda)$, we can make $\|f-g\|_1$ arbitrarily small, so it would suffice to show that each of these two sets has measure constant $\cdot \|f-g\|_1$, where the constant doesn't depend on g

(a) By Chebyshev's inequality, $\lambda(\{|f-g| > \alpha/2\}) \le \frac{2}{\alpha}\|f-g\|_1 \to 0$ as $g \to_{L_1} f$

(b) Follows from the following theorem, so $\lambda(\{A^*|f-g| > \alpha/2\}) \le \frac{2}{\alpha}\|f-g\|_1 \to 0$ as $g \to_{L_1} f$

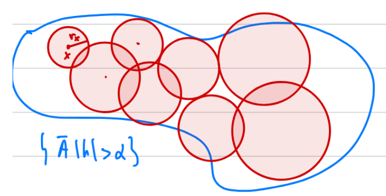
The only difference for the lim inf version of the proof is that we show $|A_* f - A_* g| \le A^*|f-g|$ which is true because: $|A_* f - A_* g| = |\liminf_{r\to 0} A_r f - \liminf_{r\to 0} A_r g| \le |\limsup_{r\to 0} A_r f - A_r g| = |\limsup_{r\to 0} A_r(f-g)| = |A_r(f-g)| \le A_r|f-g|$

Theorem 3.3.2 (Hardy-Littlewood Maximal Theorem)**.** Let $h \in L^1(\mathbb{R}^d, \lambda)$ and $\alpha > 0$. Then:

$$\lambda(\{A^*|h| > \alpha\}) \le \frac{3^d}{\alpha}\|h\|_1$$

In fact, we have $\lambda(\{\overline{A}|h| > \alpha\}) \le \frac{3^d}{\alpha}\|h\|_1$, where $\overline{A}|h| := \sup_{r\le 1} A_r|h| \ge A^*|h|$ is the Hardy-littlewood maximal function.

Proof. Note that for each $x \in \mathbb{R}^d$, we have $x \in \{\overline{A}|h| > \alpha\} \iff \exists r_x \in (0,1]$ such that $Ar_x|h| < \alpha$, i.e. that $\frac{1}{\lambda(B_{r_x}(x))}\int_{B_{r_x}} |h|\, d\lambda > \alpha$, i.e. $\lambda(B_{r_x}(x)) < \frac{1}{\alpha}\cdot\int_{B_{r_x}(x)} |h|\, d\lambda$.



It would be enough to get a countable subfamily of these balls $B_{r_x}(x)$ so that they are disjoint and cover a constant fraction (say half) of $\{\overline{A}|h| > \alpha\}$. This is exactly the content of the Vitali covering lemma below. Granted this lemma, we finish the proof as follows. For any $a > 0$ below $\lambda(\{\overline{A}|h| > \alpha\})$ and get a finite disjoint subcollection $\mathcal{C}_0 \subseteq \{B_{r_x}(x) : x \in \{\overline{A}|h| > \alpha\}\}$ with $\lambda\left(\bigsqcup_{B\in\mathcal{C}_0} \ge \frac{1}{3^\alpha}a\right)$. Then

$$\|h\|_1 \ge \int_{\bigsqcup_{B\in\mathcal{C}_0} B} |h|\, d\lambda > \alpha \cdot \sum_{B\in\mathcal{C}_0} \lambda(B) = \alpha\lambda\left(\bigsqcup_{B\in\mathcal{C}_0} B\right) \ge \frac{\alpha}{3^d}a$$

Then as $a \to \lambda(\{\overline{A}|h| > \alpha\}$,

$$\frac{\alpha}{3^d}a \to \frac{\alpha}{3^d}\lambda(\{\overline{A}|h| > \alpha\}$$

so

$$\lambda(\{\overline{A}|h| > \alpha \le \frac{3^d}{\alpha}\|h\|_1\})$$

□

Lemma 3.3.3 (Vitali Covering). Let $A \subseteq \mathbb{R}^d$ be any λ-measurable set of positive measure and let $\mathcal{C}$ be a family of balls that cover A. Then for each $0 < a < \lambda(A)$, there is a finite disjoint subcollection $\mathcal{C}_0 \subseteq \mathcal{C}$ such that:

$$\lambda\left(\bigsqcup_{B \in \mathcal{C}_0} B \geq \frac{1}{3^d} \cdot a\right)$$

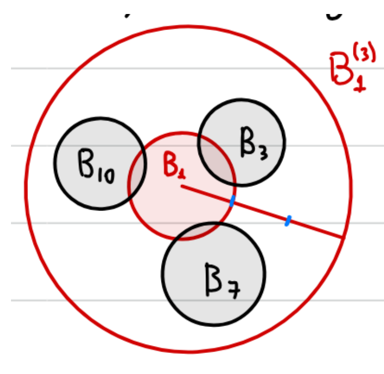


Figure 3.4: $B_1^{(3)}$ constructed from B_1 contains all the balls that intersect with B_1..

Proof. Fix $a < \lambda(A)$ and by regularity, get a compact set $K \subseteq A$ with $\lambda(K) > a$. $\mathcal{C}$ is still a cover of K, hence there is a finite subcover $\{B_1, \ldots, B_n\} \subseteq \mathcal{C}$ of K. Order these balls by decreasing radii. Put $B_{n_1} := B_1$ into $\mathcal{C}$, and delete the balls that intersect B_{n_1} and let B_{n_2} be a largest radius ball among the remaining balls. Delete the balls that intersect B_{n_2} and let B_{n_3} be the largest radius ball among what remains. Let $n_k := \min\{i < n : B_i \cap \bigcup_{j>k} B_{n_j} = \emptyset\}$. For a ball B, denote by $B^{(1)}$ the ball with the same centre but with 3 times the radius. After this algorithm finishes, we have obtained a disjoint collection $\mathcal{C}_0 := \{B_{n_1}, \ldots, B_{n_l}\}$ such that $\bigcup_{i \leq l} B_{n_i}^{(3)} \supseteq \bigcup_{j<n} B_j \supseteq K$ because $B_{n_i}^{(3)}$ contains all B_j for $j \geq n_i$ which intersect B_{n_i}. Thus,

$$\lambda(\bigsqcup_{i \leq l} B_{n_i}) = \sum_{i<l} \lambda(B_{n_i}) = \frac{1}{3^d} \sum_{i<l} \lambda(B_{n_i}^{(3)}) \geq \frac{1}{3^d} \lambda(\bigcup_{i<l} B_{n_i}^{(3)}) \geq \frac{1}{3^d} \lambda(K) \geq \frac{1}{3^d} a$$

□

Theorem 3.3.3 (Technical strengthening of Lebesgue differentiation theorem). For each $f \in \text{Loc}^1(\mathbb{R}^d, \lambda)$,

$$\lim_{r \to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| \, d\lambda(y) = 0 \quad \text{for } \lambda\text{-a.e.} x \in \mathbb{R}$$

This proof is not that important, and just strengthens what we have already shown.

Proof. What we have proved is

$$\lim_{r \to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y) - f(x) \, d\lambda(y) = 0$$

for a.e. $x \in \mathbb{R}^d$ but this doesn't help.

However, for each constant $c \in \mathbb{R}$, for a.e. $x \in \mathbb{R}^d$,

$$\lim_{r \to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f - c| \, d\lambda = |f(x) - c|$$

In particular, we have this for all rationals, so by intersecting countably many conull sets, we get a conull set $X \subseteq \mathbb{R}^d$ s.t. $\forall x \in X, \forall q \in \mathbb{Q}$,

$$\lim_{r \to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f - q| \, d\lambda = |f(x) - q|$$

Now fix $x \in X$ and put $c := f(x)$. Take $q \in \mathbb{Q}$ so $|f-c| \leq |f-q| + |q-c|$, hence

$$\begin{aligned}\frac{1}{\lambda(B_r(x))}\int_{B_r(x)} |f-c|\,d\lambda &\leq \frac{1}{\lambda(B_r(x))}\int_{B_r(x)} |f-q|\,d\lambda + \frac{1}{\lambda(B_r(x))}\int_{B_r(x)} |q-c|\,d\lambda \\ &= \frac{1}{\lambda(B_r(x))}\int_{B_r(x)} |f-c| + |q-c| \\ &\overset{r\to 0}{\to} |f(x)-q| + |q-c| \\ &= 2|q-r| \\ &\overset{q\to c}{\to} 0\end{aligned}$$

□

Definition 3.3.2. For $x \in \mathbb{R}^d$, we say that a family $\{B'_r(x)\}_{r>0}$ of λ-measurable subsets of $\mathbb{R}^d$ shrinks nicely to x if $\exists p > 0$ such that $\forall r > 0$,

(i) $B'_r(x) \subseteq B_r(x)$

(ii) $\lambda(B'_r(x)) \supseteq p\lambda(B_r(x))$

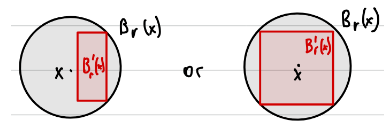


Figure 3.5: Both families shrink nicely to x. Note that x doesn't have to be in each set B' in the family..

Theorem 3.3.4 (Lebesgue differentiation theorem with non-balls)**.** Let $f \in \mathrm{Loc}^1(\mathbb{R}^d, \lambda)$. Then for a.e. $x \in \mathbb{R}^d$, for each family $\{B'_r(x)\}_{r>0}$ that shrink nicely to x, we have:

$$\lim_{r\to 0}\frac{1}{\lambda(B'_r(x))}\int_{B'_r(x)} f\,d\lambda = f(x) \quad \text{for } \lambda\text{-a.e.} x \in \mathbb{R}$$

In fact, we have,

$$\lim_{r\to 0}\frac{1}{\lambda(B'_r(x))}\int_{B'_r(x)} |f(y)-f(x)|\,d\lambda(y) = 0 \quad \text{for } \lambda\text{-a.e.} x \in \mathbb{R}$$

Equivalently, for all locally finite Borel measure $\mu \ll \lambda$,

$$\frac{d\mu}{d\lambda}(x) = \lim_{r\to 0}\frac{\mu(B'_r(x))}{\lambda(B'_r(x))}$$

Proof. We already know the statements hold for balls. Fix $x \in \mathbb{R}^d$ for which Lebesgue differentiation for balls holds and fix a family $\{B'_r(x)\}_{r>0}$ which shrinks nicely to x with the proportion $p > 0$ as in the definition. Then,

$$\frac{1}{\lambda(B'_r(x))}\int_{B'_r(x)} |f(y)-f(x)|\,d\lambda(y) \leq \frac{1}{p\lambda(B_r(x))}\int_{B_r(x)} |f(y)-f(x)|\,d\lambda(y)$$
$$\to 0 \text{ as } r \to 0 \text{ since } p \text{ is constant}$$

□

Before moving to understanding singular μ, we discuss an important corollary of Lebesgue differentiation theorem (original form)

3.4 Lebesgue density

APPLYING THE Lebesgue differentiation theorem to the indicator function of a λ-measurable set $A \subseteq \mathbb{R}^d$, we get:

Theorem 3.4.1 (Lebesgue density)**.** For every λ-measurable $A \subseteq \mathbb{R}^d$, we have for λ-a.e. $x \in \mathbb{R}^d$:

$$d_A(x) := \lim_{r \to 0} \frac{\lambda(A \cap B'_r(x))}{\lambda(B'_r(x))} = \mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \in \mathbb{R}^d \setminus A \end{cases}$$

for each family $\{B'_r(x)\}_{r>0}$ that shrinks nicely to x. We call d_A the Lebesgue density function of A.

Proof. Observe that $\frac{\lambda(A \cap B'_r(x))}{\lambda(B'_r(x))} = \frac{1}{\lambda(B'_r(x))} \int_{B'_r(x)} \mathbb{1}_A \, d\lambda$ so the statement is a special case for Lebesgue differentiation for non-balls. □

Lemma 3.4.1 (Strong 99% lemma)**.** For every λ-measurable $A \subseteq \mathbb{R}^d$ of positive measure, for λ-a.e. $x \in A$, for all small enough $r > 0$, we have:

$$\frac{\lambda(A \cap B'_r(x))}{\lambda(B'_r(x))} \geq 0.99$$

Definition 3.4.1 (Lebesgue density set)**.** For a λ-measurable set $A \subseteq \mathbb{R}^d$, put $\mathcal{D}_A := \{x \in \mathbb{R}^d : d_A(x) = 1\}$ and call this the Lebesgue density set of A.

Observation 3.4.2.

(a) $A =_\lambda \mathcal{D}_A$, *i.e.* $A \triangle \mathcal{D}_A$ *is null, for each* λ*-measurable* $A \subseteq \mathbb{R}^d$

(b) $A =_\lambda B \implies d_A = d_B \implies \mathcal{D}_A = \mathcal{D}_B$ *for all* λ*-measurable* $A, B \subseteq \mathbb{R}^d$.

Remark 3.4.3. *In other words, the map* $A \mapsto \mathcal{D}_A$ *is a canonical selector for the equivalence relation* $=_\lambda$ *on* Meas_λ

Remark 3.4.4 (Teaser)**.** *Call a* λ*-measurable set* $A \subseteq \mathbb{R}^d$ *Lebegue open if* $A \subseteq \mathcal{D}_A$. *Shockingly, these sets form a topology on* $\mathbb{R}^d$ *called the Lebesgue density topology, in particular arbitrary unions of Lebesgue open sets are Lebesgue open, hence* λ*-measurable. This topology is much finer than the usual (Euclidean) topology, in particular, it's not metrizable.*

3.5 Lebesgue differentiation for all measures

LETTING μ be a locally finite Borel measure on $\mathbb{R}^d$, we've shown that if $\mu \ll \lambda$ then $\lim_{r \to 0} \frac{\mu(B_r(x))}{\lambda(B_r(x))} = \frac{d\mu}{d\lambda}(x)$ for λ-a.e. $x \in \mathbb{R}^d$. What about $\mu \perp \lambda$?. For example, let μ be the pushforward of Bernoulli$(\frac{1}{2})$ from

$2^{\mathbb{N}}$ to the standard Cantor set $C \subseteq [0,1]$, which is λ-null. Then a.e. $x \in \mathbb{R}$ is not in C, hence is in the open set $U := \mathbb{R} \setminus C$, hence all small enough balls $B_r(x) \subseteq U$, hence $\mu(B_r(x)) = 0$ so $\frac{\mu(B_r(x))}{\lambda(B_r(x))} = 0$. In particular, $\lim_{r\to 0} \frac{\mu(B_r(x))}{\lambda(B_r(x))} = 0$ for a.e. $x \in \mathbb{R}$. This is true more generally:

Theorem 3.5.1 (Lebesgue differentiation for singular measures)**.** Let μ be a locally finite measure on $\mathbb{R}^d$ such that $\mu \perp \lambda$. Then for λ-a.e. $x \in \mathbb{R}^d$ and any family $\{B'_r(x)\}_{r>0}$ that shrinks nicely to x,

$$\lim_{r\to 0} \frac{\mu(B'_r(x))}{\lambda(B'_r(x))} = 0$$

Proof. It is enough to prove for balls since,

$$\frac{\mu(B'_r(x))}{\lambda(B'_r(x))} \leq \frac{\mu(B_r(x))}{p\lambda(B_r(x))}$$

and $p > 0$ is constant.

By $\mu \perp \lambda$, we have a partition $\mathbb{R}^d = X_\lambda \bigsqcup X_\mu$ where X_λ is λ-conull and X_μ is μ-conull, in particular, X_μ is λ-null. We show that for λ-a.e. $x \in X_\lambda$,

$$\limsup_{r\to 0} \frac{\mu(B_r(x))}{\lambda(B_r(x))} = 0$$

By the usual argument, it is enough to show that for any $\alpha > 0$, the set $Z_\alpha := \{x \in X_\lambda : \limsup_{r\to 0} \frac{\mu(B_r(x))}{\lambda(B_r(x))} > \alpha\}$ is λ-null. Fix $\alpha > 0$ and $\varepsilon > 0$. Since $Z_\alpha \subseteq X_\alpha$, we know that $\mu(Z_\alpha) = 0$, so by outer regularity there exists an open set $U \supseteq Z_\alpha$ with $\mu(U) \leq \varepsilon \cdot \frac{\alpha}{3^d}$. Now for each $x \in Z_\alpha$, there is $r_x > 0$ such that $\frac{\mu(B_{r_x}(x))}{\lambda(B_{r_x}(x)) > \alpha}$ $(*)$ and $B_{r_x}(x) \subseteq U$. Replacing U with $U_{x\in Z_\alpha} B_{r_x}(x)$ we may assume $U = \bigcup_{x\in Z_\alpha}$ to begin with. This is a cover $\mathcal{C}$ of U by balls, so for an arbitrary $u < \lambda(U)$, the Vitali covering lemma gives a finite collection $\mathcal{C}_0 \subseteq \mathcal{C}$ of pairwise disjoint balls such that $\lambda(\bigsqcup_{B\in\mathcal{C}_0}) \geq \frac{1}{3^d} u$. Then:

$$u \leq 3^d \cdot \lambda(\bigsqcup_{B\in\mathcal{C}_0} B) = 3^d \sum_{B\in\mathcal{C}_0} \lambda(B) \leq \frac{3^d}{\alpha} \sum_{B\in\mathcal{C}_0} \mu(B) = \frac{3^d}{\alpha} \mu(\bigsqcup_{B\in\mathcal{C}_0} B) \leq \frac{3^d}{\alpha} \mu(U) \leq \varepsilon$$

Taking $u \to \lambda(u)$, we get $\lambda(U) \leq \varepsilon$, so $\lambda(Z_\alpha) \leq \varepsilon$. Hence $\lambda(Z_\alpha) = 0$ because $\varepsilon > 0$ is arbitrary. □

Corollary 3.5.1 (Lebesgue differentiation for all measures)**.** Let μ be a locally finite Borel measure on $\mathbb{R}^d$ and let $\mu = \mu_\ll + \mu_\perp$ be a Lebesgue decomposition of μ with respect to λ, i.e. $\mu_\ll \ll \lambda$ and $\mu_\perp \perp \lambda$. Then for λ-a.e. $x \in \mathbb{R}^d$ and any family $\{B'_r(x)\}_{r>0}$ shrinking nicely to x, we have,

$$\lim_{r\to 0} \frac{\mu(B'_r(x))}{\lambda(B'_r(x))} = \frac{d\mu_\ll}{d\lambda}(x)$$

3.6 Functions on $\mathbb{R}$ as distributions of measures: the FTC

WE VIEWED locally integrable functions on $\mathbb{R}^d$ as Radon-Nikodym derivatives of locally finite Borel measures on $\mathbb{R}^d$ with respect to λ. We now focus on $\mathbb{R}$ and view functions as distributions of locally finite Borel signed measure on $\mathbb{R}$. We begin with unsigned measures and derive the theory for signed measures from unsigned using the decomposition of signed measures as as difference of unsigned measures.

Definition 3.6.1. A distribution of a locally finite Borel measure μ on $\mathbb{R}$ is a function $f : \mathbb{R}to\mathbb{R}$ such that

$$\mu((a,b]) = f(b) - f(a)$$

for all $a \leq b$. Such a function is increasing and right continuous, and unique up to an additive constant.

For example:

$$f(x) = \begin{cases} \mu((0,x]) & x > 0 \\ \mu((-x,0]) & x \leq 00 \end{cases}$$

It was a HW problem to prove that if a distribution f of μ is continuously differentiable, then $\mu \ll \lambda$ and $f' = \frac{d\mu}{d\lambda}$ λ-a.e.. This holds more generally for all $\mu \ll \lambda$.

Theorem 3.6.1 (characterization of measures via distributions)**.** Let μ be a locally finite Borel measure on $\mathbb{R}$ and f a distribution of μ. Let $\mu = \mu_{\ll} + \mu_{\perp}$ be a Lebesgue decomposition of μ with respect to λ, i.e. $\mu_{\ll} \ll \lambda$ and $\mu_{\perp} \perp \lambda$. Then $\frac{d\mu_{\ll}}{d\lambda} = f'$ λ-a.e. (in particular, f' exists λ-a.e.). Thus:

(a) $\mu \ll \lambda \iff$ the fundamental theorem of calculus (FTC) holds for f: for all $a \leq b$, $f(b) - f(a) = \int_{(a,b)} f' \, d\lambda$

(b) $\mu \perp \lambda \iff f' = 0 \; \lambda-$a.e

Proof. For each $x \in \mathbb{R}$, $f'(x) = \lim_{r \to 0} \frac{f(x+r)-f(x)}{r}$. To show that this exists, it's enough to show:

$$\lim_{r \to 0^+} \frac{f(x+r) - f(x)}{r} = \frac{d\mu_{\ll}}{d\lambda}(x) = \lim_{r \to 0^+} \frac{f(x) - f(x-r)}{r} \qquad (*)$$

But $r = \lambda((x, x+r])$, $f(x+r) - f(x) = \mu((x, x+r])$ and $f(x) - f(x - r) = \mu((x-r, x])$, and the families $((x, x+r])_{r>0}$ and $((x-r, x])_{r>0}$ shrink nicely to x, so by Lebesgue differentiation for all measures, we get $(*)$ for λ-a.e. $x \in \mathbb{R}$.

(a) It is clear by Radon-Nikodym, the definition of a distribution, and $(*)$ that:

$$\mu \ll \lambda \implies \mu((a,b]) = \int_{(a,b)} \frac{d\mu_{\ll}}{d\lambda}\, d\lambda \qquad \text{for all } a \le b$$
$$\iff f(b) - f(a) = \int_{(a,b)} f'\, d\lambda \qquad \text{for all } a \le b$$

To get the reverse of the first implication, note that since the intervals $(a,b]$ generate the Borel σ-algebra of $\mathbb{R}$, the uniqueness part of Carathéodory's theorem gives that $\mu = \lambda_{f'}$.

(b) $\mu \perp \lambda \iff \mu_{\ll} = 0 \iff \frac{d\mu_{\ll}}{d\lambda} = 0$ λ-a.e. $\iff f' = 0$ λ-a.e.

□

We would like to characterize exactly which functions are distributions of μ with $\mu \ll \lambda$, equivalently, which functions satisfy the FTC. If $f : \mathbb{R} \to \mathbb{R}$ is such a function then it is increasing and continuous because μ is atomless. Moreover, for each (a,b), $\mu|_{(a,b)}$ is finite so $\mu|_{(a,b)} \ll \lambda$ is equivalent to $\forall \varepsilon > 0 \exists \delta > 0$ s.t. for all Borel sets $B \subseteq (a,b).\lambda(B) < \delta \implies \mu(B)\varepsilon$. Writing this for B open and writing B as a disjoint union of open intervals, gives:

Definition 3.6.2 (absolutely continuous)**.** A function $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous on (a,b) if $\forall \varepsilon > 0 \exists \delta > 0$ such that for all open sets $U \subseteq (a,b)$, writing $U = \bigsqcup_{n \in \mathbb{N}} (a_n, b_b)$, we have:

$$\sum_{n \in \mathbb{N}} b_n - a_n \le \delta \implies \sum_{n \in \mathbb{N}} |f(b_n) - f(a_n)| \le \varepsilon$$

We say that f is locally absolutely continuous if f is absolutely continuous on every $(a,b) \subseteq \mathbb{R}$.

Remark 3.6.3. *This definition says that the vertical distances travelled by the graph is uniformly continuous with respect the horizontal distance it travels.*

Remark 3.6.4. *Bounded derivative $\implies$ Lipschitz $\implies$ absolutely continuous $\implies$ uniformly continuous*

Theorem 3.6.2 (characterization of distributions of $\mu \ll \lambda$)**.** For an increasing function $f : \mathbb{R} \to \mathbb{R}$, TFAE:

(1) f is a distribution of a locally finite Borel measure $\mu \ll \lambda$

(2) FTC holds for f, i.e. f' exists λ-a.e. and $f(b) - f(a) = \int_{(a,b)} f'\, d\lambda$ for all $(a,b) \subseteq \mathbb{R}$

(3) f is locally absolutely continuous

Proof. We have already proved $(1) \iff (2)$.

$(1) \implies (3)$. Supposing that $\mu \ll \lambda$ and fixing $(a,b) \subseteq \mathbb{R}$, we have that $\forall \varepsilon > 0 \exists \delta > 0$ such that for all Borel sets $B \subseteq (a,b)$,

$\lambda(B) \le \delta \implies \mu(B) \le \varepsilon$. Now let B be open, so $B = \bigsqcup_{n\in\mathbb{N}}(a_n, b_n)$, hence $\lambda(B) = \sum_{n\in\mathbb{N}} \lambda((a_n, b_n)) = \sum_{n\in\mathbb{N}} b_n - a_n$ and similarly $\mu(B) = \sum_{n\in\mathbb{N}} \mu((a_n, b_n)) = \sum_{n\in\mathbb{N}} \mu((a_n, b_n]) = \sum_{n\in\mathbb{N}} f(b_n) - f(a_n) \sum_{n\in\mathbb{N}} |f(b_n) - f(a_n)|$ because μ is atomless and f is increasing. Thus, f is absolutely continuous on (a, b).

(3) $\implies$ (1). This is by the regularly of λ. Suppose f is locally absolutely continuous. To show $\mu \ll \lambda$, if suffices to show that $\mu|_{(a,b)} \ll \lambda|_{(a,b)}$ because $\mathbb{R} = \bigcup_{n\in\mathbb{Z}}(n, n+2)$. Fix (a, b) and $\varepsilon > 0$. Let δ be the witness to absolutely continuity of $f|_{(a,b)}$. Then for each Borel set $B \subseteq (a, b)$, if $\lambda(B) \le \frac{\delta}{2}$ then by the regularity of λ there is an open set U such that $B \subseteq U \subseteq (a, b)$ with $\lambda(U) \le \delta$ and hence $\mu(U) \le \varepsilon$ because $\mu(U) = \sum_{n\in\mathbb{N}} f(b_n) - f(a)$ where $U = \bigsqcup_{n\in\mathbb{N}}(a_n, b_n)$. But then $\mu(B) \le \mu(U) \le \varepsilon$ so we are done. □

Example 3.6.1 (the devil's staircase)**.** Let $f : [0, 1] \to [0, 1]$ be the distribution with $f(0) = 0$ of the pushforward measure μ of the Bernoulli($\frac{1}{2}$) measure through the the usual homeomorphism $2^{\mathbb{N}} \to C \subseteq [0, 1]$ where C is the standard Cantor set. Because $\lambda(C) = 0$ while $\mu(C) = 1$, so C is μ-conull, $\mu \perp \lambda$. Thus $f' = 0$ a.e. and f isn't absolutely continuous because $\mu \not\ll \lambda$. But f is uniformly continuous since it is continuous on the compact interval $[0, 1]$ because μ is atomless.

Now we switch to signed measure on $\mathbb{R}$ to characterize all, not just increasing, functions for which the FTC holds. Because the definition of signed measures is asymmetric (only $+\infty$ or $-\infty$ is allowed), we restrict to finite signed measures for the sake of notation convenience. However, everything still works for non-finite signed measures.

Definition 3.6.5 (total variation)**.** For a signed measure ν on some measurable space, let $\nu = \nu_+ - \nu_-$ be it's Hahn decomposition into unsigned measures. The total variation ν_* of ν is defined $\nu_* := \nu_+ + \nu_-$. We say that ν is finite if ν_* is finite.

For a finite Borel signed measure ν on $\mathbb{R}$, we call $f : \mathbb{R} \to \mathbb{R}$ a distribution of ν if $\nu((a, b]) = f(b) - f(a)$ for all $a \le b$. Such functions differ by additive constants and writing $\nu = \nu_+ - \nu_-$ for unsigned finite Borel measures $\nu_+ - \nu_-$, letting g and h be distributions of ν_+ and ν_-, we get that $f := g - h$ is a distribution of ν.

For example,

$$f(x) := \begin{cases} \nu((0, x]) & x > 0 \\ \nu((-x, 0]) & x \le 0 \end{cases}$$

is a distribution function.

From the Hahn decomposition of signed measures, we get:

$$\text{finite Borel measures on } \mathbb{R} \overset{\text{distribution}}{\longleftrightarrow} \text{bdd increasing right continuous functions}$$

$$\text{finite signed measures on} \mathbb{R} \overset{\text{distribution}}{\longleftrightarrow} \text{differences of bdd increasing right-continuous functions}$$

But which functions are a difference of two bdd increasing right-continuous functions? Spelling out what is means for ν_* to be finite on open sets gives the following:

Definition 3.6.6 (total variation, bounded variation)**.** For $f : \mathbb{R} \to \mathbb{R}$, define the total variation of f, $T_f : \mathbb{R} \to [0, \infty]$ by

$$T_f(x) := \sup\{\sum_{i<n} |f(x_{i+1}) - f(x_i)| : -\infty < x_0 < \ldots < x_n \leq x\}$$

f has bounded variation if:

$$T_f(\infty) := \lim_{x \to \infty} T_f(x) < \infty$$

Remark 3.6.7. *Note that for $(a, b] \subseteq \mathbb{R}$, $T_f(b) - T_f(a) = \sup\{\sum_{i<n} |f(x_{i+1}) - f(x_i)| : a < x_0 < \ldots < x_n \leq b\}$ is the vertical distance travelled by the graph of f on $(a, b]$.*

Proposition 3.6.1. For $f : \mathbb{R} \to \mathbb{R}$, TFAE:

(1) f has bounded variation

(2) f is a difference of two bounded increasing functions, namely: $f = \frac{1}{2}(T_f + f) - \frac{1}{2}(T_f - f)$

Proof. (2) $\implies$ (1). The triangle inequality implies that linear combinations of bounded variation functions have bounded variation, so it suffices to check that every bounded increasing function $g : \mathbb{R} \to \mathbb{R}$ has bounded variation, which follows from $T_g(\infty) = g(\infty) - g(-\infty) = \lim_{x\to\infty} g(x) - \lim_{x\to\infty} g(-x) < \infty$
(1) $\implies$ (2) (HW) □

Proposition 3.6.2. If $f : \mathbb{R} \to \mathbb{R}$ is right-continuous and has bounded variation, then T_f is right-continuous

Proof. HW □

Corollary 3.6.1 (characterization of distributions of finite signed measures)**.** For $f : \mathbb{R} \to \mathbb{R}$, TFAE:

(1) f is a distribution of a finite Borel signed measure on $\mathbb{R}$

(2) f is a difference of bounded increasing right continuous functions

(3) f is right continuous and has bounded variation

Proof. (1) $\iff$ (2) follows from Hahn decomposition of signed measures and the fact that bounded increasing right continuous functions are exactly the distributions of finite Borel (unsigned) measures on $\mathbb{R}$. (2) $\iff$ (3) follows from the previous two propositions. □

Definition 3.6.8. Let $(X, \mathcal{B})$ be a measurable space, μ an (unsigned) measure on it and ν a signed measure on it. We say that ν is absolutely continuous with respect to μ and write $\nu \ll \mu$, if $\nu_* \ll \mu$.

Finally, combining this with the characterization of distributions of finite (unsigned) Borel measure $\mu \ll \lambda$, we obtain:

Theorem 3.6.3 (characterization of distributions of finite signed $\nu \ll \lambda$)**.** For $f : \mathbb{R} \to \mathbb{R}$, TFAE:

(1) f is a distribution of finite Borel signed measure $\nu \ll \lambda$

(2) The FTC holds for fi.e. f' exists λ-a.e. and $f(b) - f(a) = \int_{(a,b)} f' \, d\lambda$ for all $(a, b) \subseteq \mathbb{R}$.

(3) f is absolutely continuous and has bounded variation

Remark 3.6.9. *For $f : \mathbb{R} \to \mathbb{R}$ and $[a, b] \subseteq \mathbb{R}$, if $f|_{[}a, b]$ is absolutely continuous, then $f|_{[}a, b]$ has bounded variation. However, f being absolutely continuous on $\mathbb{R}$ does not imply bounded variation on $\mathbb{R}$.*

4 Extras

THE FOLLOWING are not part of the course but are things I found generally helpful to refer to:

4.1 Selected results proved in homework

Theorem 4.1.1. In a metric space (X, d), every open set is F_σ and every closed set is G_δ.

Theorem 4.1.2. All the following collections generate $\mathcal{B}(\mathbb{R}^d)$:

(i) Balls with rational centres and rational radii

(ii) Bounded open boxes

(iii) Bounded closed boxes

Theorem 4.1.3. Every Cantor set is closed. The standard Cantor set is λ-null. $2^{\mathbb{N}}$ is homeomorphic to the Cantor set. It is possible to construct cantor sets of positive Lebesgue measure.

Definition 4.1.1. Let X be a set. A subset $F \subseteq X$ separates $x, y \in X$ if it contains exactly one of x, y. A family $\mathcal{F}$ of subsets of X separates points in X if any two distinct points $x, y \in X$ are separated by some $F \in \mathcal{F}$

Lemma 4.1.1. If $\mathcal{B}$ is a σ-algebra on X containing all singletons, then every generating family $\mathcal{F}$ for $\mathcal{B}$ separates points in X.

Lemma 4.1.2. If $(X, \mathcal{B}, \mu)$ is a σ-finite measure space and $\mathcal{B}$ contains all singletons, then every μ-atom is $=_\mu$ to a singleton.

Theorem 4.1.4. Every σ-finite Borel measure μ on a second countable metric space X decomposes into purely atomic and atomless parts, i.e. $\mu = \mu_0 + \mu_1$ where μ_1 is an atomless Borel measure on X and μ_1 is a purely atomic Borel measure on X.

Theorem 4.1.5 (Steinhaus theorem)**.** For every Lebesgue measurable non-null set $A \subseteq \mathbb{R}^d$, the difference set $A - A := \{a_0 - a_1 : a_0, a_1 \in A\}$ contains an open neighbourhood of 0.

Theorem 4.1.6. Let (X, μ) be a measure space and Y, Z topological/metric spaces. If $f : X \to Y$ is μ-measurable and $g : Y \to Z$ is Borel then $g \circ f : X \to Z$ is μ-measurable.

Definition 4.1.2. Let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \to X$ be a $(\mathcal{B}$-$\mathcal{B})$-measurable μ-preserving transformation. Call a set $W \subseteq X$ T-wandering if the preimages T^{-n} are pairwise disjoint for all $n \in \mathbb{N}$. Call a set $W \subseteq X$ T-recurrent if for a.e. $x \in B$ there are infinitely many $n \in \mathbb{N}$ such that $T^n(x) \in B$.

Theorem 4.1.7 (Poincaré recurrence theorem)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \to X$ be a $(\mathcal{B}$-$\mathcal{B})$-measurable μ-preserving transformation. Then every μ-measurable set $B \subseteq X$ is T-recurrent.

Theorem 4.1.8. Let $f : [a, b] \to \mathbb{R}$. f is Riemann integrable $\iff$ f is bounded and continuous a.e.

Theorem 4.1.9. For a countably generated measure space $(X, \mathcal{B}, \mu)$, there is a countable collection of simple functions which are dense in $L^1(X, \mu)$ in the L^1 metric. In particular $L^1(X, \mu)$ is separable.

Theorem 4.1.10 (Generalized dominated convergence theorem)**.** Let (X, μ) be a measure space and f_n, f μ-measurable functions such that $f_n \to f$ a.e. If there are nonnegative functions $\{g_n\}, g \in L^1(X, \mu)$ such that $g_n \to g$ a.e. and $|f_n| \leq g_n$ for each $n \in \mathbb{N}$, then $\{f_n\}, f \in L^1$ for all $n \in \mathbb{N}$ and $f_n \to_{L_1} f$.

Theorem 4.1.11 (Change of variables)**.** Let $(X, \mathcal{B}), (Y, \mathcal{C})$ be measurable spaces and $T : X \to Y$ a $(\mathcal{B}$-$\mathcal{C})$-measurable function. Then for any measure μ on $\mathcal{B}$ and measurable function $f \in L^1(Y, \mathcal{C}, T_*\mu)$,

$$\int_X (f \circ T)\, d\mu = \int_Y f\, d(T_*\mu)$$

Corollary 4.1.1. Let $T_A : \mathbb{R}^d \to \mathbb{R}^d$ be the linear transformation given by a $d \times d$ invertible matrix A. Let λ be the Lebesgue measure on $\mathbb{R}^d$. Then:

$$\int f \circ T_A\, d\lambda = \int f |\det A|^{-1}\, d\lambda$$

Theorem 4.1.12. BS is dense in $L^1(\mathbb{R}^d, \lambda)$, where BS is the set of simple functions that are finite linear combinations of indicator functions of bounded boxes in $\mathbb{R}^d$

Theorem 4.1.13. The set of continuous functions $\mathbb{R}^d \to \mathbb{R}$ if bounded support are dense in $L^1(\mathbb{R}^d, \lambda)$.

Theorem 4.1.14 (Egorov)**.** Let (X, μ) be a finite measure space and $f_n \to f$ a.e. for some μ-measurable functions $f_n, f : X \to \mathbb{R}$. Then $\forall \varepsilon > 0$ there is a μ-measurable set $X' \subseteq X$ such that $\mu(X \setminus X') < \varepsilon$ such that $f_n|_{X'} \to f|_{X'}$ uniformly.

Theorem 4.1.15. Let (X, μ) be a measure space and $f, g : X \to \mathbb{R}$ be μ-measurable functions. Then:

$$d_\mu(f, g) := \inf_{\alpha > 0}(\delta_\alpha(f, g) + \alpha)$$

is a pseudo-metric on the set of μ-measurable functions. Also,
(i) $f_n \to f$ in measure $\iff$ $f_n \to f$ in d_μ
(ii) f_n is Cauchy in measure $\iff$ f_n is d_μ-Cauchy
This justifies the metric space terminology (such as Cauchy sequences) used for the topology of convergence in measure.

Theorem 4.1.16. $\mu \times \nu$-measurable Fubini-Tonelli works.

Theorem 4.1.17 (Namioka's trick)**.** Let $f : (0, \infty) \to \mathbb{R}$ be a Lebesgue integrable function. Then:

$$\int_{(0,\infty)} \int_{(x,\infty)} t^{-1} f(t)\, d\lambda(t)\, d\lambda = \int_{(0,\infty)} f\, d\lambda$$

Theorem 4.1.18 (Conditional expectation)**.** Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space and $\mathcal{C} \subseteq \mathcal{B}$ be a sub-σ-algebra witnessing the σ-finiteness of μ. Thus, the restriction $\nu := \mu|_{\mathcal{C}}$ is a σ-finite measure on $(X, \mathcal{C})$. Then:
(a) $\int g\, d\mu = \int g\, d\nu$ for every $\mathcal{C}$-measurable function $g : X \to \overline{\mathbb{R}}$ which is non-negative or μ-integrable. In particular, $\mathcal{C}$-measurable μ-integrable functions are ν-integrable
(b) For every $\mathcal{B}$-measurable μ-integrable function, there exists a $\mathcal{C}$-measurable μ-integrable function $\tilde{f}$ such that $\int_C f\, d\mu = \int_C \tilde{f}$ for each $C \in \mathcal{C}$. $\tilde{f}$ is unique up to a μ-null set and is called the conditional expectation of f with respect to $\mathcal{C}$. In particular, if f was already $\mathcal{C}$-measurable, then $\tilde{f} = f$ a.e.

Theorem 4.1.19 (Measure disintegration)**.** Let $(X, \mathcal{B}), (Y, \mathcal{C})$ be standard Borel spaces. Let $\pi : X \to Y$ be a ($\mathcal{B}$-$\mathcal{C}$)-measurable function and $P(X, \mathcal{B})$ the set of all probability measures on $(X, \mathcal{B})$ Then:
(a) $\pi_* \mu$ admits measure disintegration, i.e. a map $\phi : Y \to P(X, \mathcal{B})$ such that:
  (i) for each $B \in \mathcal{B}$, the function $\phi_B : y \mapsto \mu_y(B)$ is $\mathcal{C}$-measurable and
  $$\mu(B) = \int_Y \mu_y(B)\, d(\pi_* \mu)$$
  (ii) $\mu_y(\pi^{-1}(y)) = 1$ for $\pi_* \mu$-a.e. $y \in Y$.
(b) ϕ is unique up to $\pi_* \mu$-null sets.

Theorem 4.1.20 (Locally finite Borel measures on $\mathbb{R}$)**.** Let μ be a locally finite Borel measure μ on $\mathbb{R}$. Then μ admits a distribution that is unique up to an additive constant, each distribution of μ is an increasing right-continuous function. Also, every increasing right continuous function $f : \mathbb{R} \to \mathbb{R}$ is a distribution of some locally finite Borel measure μ on $\mathbb{R}$.

4.2 *Translations between measure theory and probability theory*

Measure theory concept	Probability theory concept
Measure space $(X, \mathcal{B}, \mu)$, $\mu(X) = 1$	Probability space $(\Omega, \mathcal{F}, P)$
Measurable set $B \in \mathcal{B}$	Event $E \in \mathcal{F}$
$\mathcal{B}$-measurable function $f : X \to \mathbb{R}$	Random variable $X : \Omega \to \mathbb{R}$
$\mu_f(X) = \int_X f \, d\mu$	Expected value $E[X] = \int_\Omega X(w) \, dP(\omega)$
pushforward measure of f on $\mathbb{R}$, i.e. $f_*\mu(B) = \mu(f^{-1}(B))$	distribution of X
$f_*\mu(\{x\})$	pmf $f_X(x) = P(X = x)$
$x \mapsto f_*\mu((-\infty, x])$ (which is a distribution function of $f_*\mu$)	cdf $F_X(x) = P(X \leq x)$
Radon-Nikodym derivative $\frac{df_*\mu}{d\lambda}(x)$	pdf/pmf $f_X(x)$, when $X_*P \ll \lambda$
$f_n \to f$ a.e.	$X_n \to X$ a.s.
$f_n \to f$ in measure	$X_n \to X$ in probability
$f_n \to_{L_1} f$	$X_n \to_{L^r} X$ (converges in r-th mean)
$f_{n*}\mu \to f_{n*}\mu$ weakly (not covered)	$X_n \to_{L^r} X$ in distribution
Change of variable	$E[X] = \int_{\mathbb{R}} x f_X(x) \, d\lambda(x)$
Conditional expectation (Theorem 4.1.20)	Conditional expectation $E[X\|Y]$, Iterated expectation $E[X] = E[E[X\|Y]]$

Example 4.2.1. Consider a discrete probability space $(\Omega, \mathcal{P}(\Omega), P)$ (e.g. $\Omega = \mathbb{N}$) and let $X : \Omega \to \mathbb{R}$ be a function (automatically $\mathcal{P}(\Omega)$-measurable, hence a r.v.). The pushforward measure of X on $\mathbb{R}$, $X_*P : \mathcal{B}(\mathbb{R}) \to [0, \infty]$, is:

$$X_*P(B) = P(X^{-1}(B))$$

It is easy to see that X_*P is a probability measure on $\mathbb{R}$, indeed,

$$X_*P(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$$

Hence since both $\nu := X_*P$ and λ (the Lebesgue measure on $\mathbb{R}$) are σ-finite. Also, note that $X_*P \perp \lambda$ because $\mathbb{R} = X(\Omega) \sqcup \mathbb{R} \setminus (X(\Omega))$ and we have that $\lambda(X(\Omega)) = 0$ because $X(\Omega$ is countable and $X_*P((\mathbb{R} \setminus X(\Omega))) = P(\emptyset) = 0$.

By applying the Lebesgue-Radon-Nikodym theorem, we can write $\nu = \nu_f + \lambda_0$ where $\lambda_0 \perp \nu$ and $\nu_f \ll \lambda$. However, since $\nu \perp \lambda$, we have that $\nu_f = 0$, where $f := \frac{d\nu}{d\mu}$ is meaningless since any choice of f such that $f|_{\mathbb{R}\setminus X(\Omega)} = 0$ works. Hence, in this setting, it would be wrong to think about the pmf of X as a Radon-Nikodym derivative of the pushfoward measure $X_*\mu$ with respect to λ. Instead, the correct definition of the pmf of X is literally $\nu = X_*\mu$.

Example 4.2.2. Consider a discrete probability space $(\Omega, \mathcal{P}(\Omega), P)$ (e.g. $\Omega = \mathbb{N}$) and let $X : \Omega \to X(\Omega)$ be a function (automatically $\mathcal{P}(\Omega)$-measurable, hence a r.v.). Also, equip $X(\Omega)$ with the counting measure μ, then μ is a σ-finite measure on the measurable space $(X(\Omega), \mathcal{P}(X(\Omega)))$ since $X(\Omega)$ is at most countable.

The pushforward measure of X on $X(\Omega)$, $X_*P : \mathcal{P}(X(\Omega)) \to [0, \infty]$, is:

$$X_*P(B) = P(X^{-1}(B))$$

Like before, X_*P is a probability measure. Hence both measure μ and $\nu := X_*P$ are σ-finite. Also, now $\nu \ll \mu$ since the only μ-null set is the empty set which is also ν-null. Hence, by applying the Lebesgue-Radon-Nikodym theorem, we can write $\nu = \nu_f$ where $f := \frac{d\nu}{d\mu}$ is a Radon-Nikodym derivative. Hence, in this setting, it makes sense to think about the pmf of X as a Radon-Nikodym derivative of the pushfoward measure $X_*\mu$ with respect to λ. Doing this makes the proofs that involve Radon-Nikodym derivatives as pdfs work for discrete random variables as well.

Example 4.2.3. Consider a continuous probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X : \Omega \to \mathbb{R}$ such that the pushforward measure $\nu := X_*P \ll \lambda$. Applying the Radon-Nikodym theorem, we get $\nu = \nu_f$, and set $\frac{d\nu}{d\lambda}(x) = f$ so that $\nu(A) = \int_A \frac{d\nu}{d\lambda}(x)\, d\lambda(x)$. In this case, it makes sense to set the pdf f_X of X as $f_X(x) = \frac{dP_*\lambda}{d\lambda}(x)$.

The cdf $F_X : \mathbb{R} \to \mathbb{R}$ of X defined $F(x) = P(X^{-1}((-\infty, x)))$ is a distribution function of ν_f. Then, using again that $\nu \ll \lambda$, we get that $F_X' = \frac{d\nu}{d\lambda} = f_X$ λ-a.e. (by characterization of measures via distribution).

Example 4.2.4 (Continuous random variable such that $X_*P \not\ll \lambda$)**.**

Let $(2^{\mathbb{N}}, \mathcal{B}, P)$ be a probability space where $\mathcal{B}$ is the σ-algebra generated by cylinders and P is the Bernoulli($\frac{1}{2}$) measure. Let $X : 2^{\mathbb{N}} \to C \subseteq [0, 1]$ be the typical mapping of $2^{\mathbb{N}}$ to the Cantor set $C \subseteq [0, 1]$, in particular, it is a $\mathcal{B}$-measurable function hence a random variable. It is easy to verify that $X_*P \perp \lambda$. Note that $X_*P \not\ll \lambda$ since $\lambda(C) = 0$ but $X_*P(C) = 1$. Hence it does not make sense to define a pdf through Radon-Nikodym derivatives for X like in the previous example because the Radon-Nikodym derivative does not exist.

By the characterization of measures via distributions, letting $F_X : \mathbb{R} \to \mathbb{R}, x \mapsto P(X^{-1}((-\infty, x)))$ be the cdf of X, we have that $F_X' = 0$ λ-a.e.

Example 4.2.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X \in L^1(\Omega, P)$. Suppose that $X_*P \ll \lambda$ so that the pdf $f_X(x)$ exists. The goal is to show the familiar identity:

$$E[X] = \int_\Omega X\, dP = \int_\mathbb{R} x f_X(x)\, d\lambda(x)$$

Note that:

$$f_X(x) = \frac{dX_*P}{d\lambda}(x)$$

By a corollary of Lebesgue-Radon-Nikodym (since X is X_*P-integrable, $X_*P \ll \lambda$, and both X_*P and λ are σ-finite),

$$\int_\mathbb{R} x f_X(x)\, d\lambda(x) = \int_\mathbb{R} x\, dX_*P(x)$$

Then by change of variables,

$$\int_\Omega X\, dP = \int_\mathbb{R} x\, dX_*P(x)$$

Note that this can also be extended to derive the identity:

$$E[g(X)] = \int g(x) f_X(x)\, d\lambda(x)$$

Which gives the familiar formula for the n-th moments (when they are well-defined):

$$\int_\Omega X(\omega)^n\, dP(\omega) = \int x^n f_X(x)\, d\lambda(x)$$

Example 4.2.6 (Iterated expectation)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ random variables (i.e. $\mathcal{F}$-measurable functions) such that $E[X]$ exists (i.e. $X \in L^1(\Omega, P)$). Let $\mathcal{C}$ be the sub-σ-algebra generated by Y (the preimage under Y of each Borel set is a set in $\mathcal{A}$, then $\mathcal{C}$ is the σ-algebra generated by all these sets.). Applying Theorem 4.1.20, we get that there exists a function $\tilde{X}$ (the conditional expectation of X with respect to $\mathcal{C}$) that is unique up to a P-null set such that for every set $C \in \mathcal{C}$,:

$$\int_C X\, dP = \int_C \tilde{X}\, dP$$

In probability, the function $\tilde{X}$ is denoted $E[X|Y]$. Clearly, we also have that:

$$E[X] = \int_\Omega X\, dP = \int_\Omega E[X|Y]\, dP = E[E[X|Y]]$$