Measure Theory
Princeton University MAT425
Lecture Notes

shapiro@math.princeton.edu

Created: Jan 10th 2025, Last Typeset: May 15, 2025

Abstract

These lecture notes correspond to a course given in the Spring semester of 2025 in the math department of Princeton University.

Contents

Syllabus

- Main source of material for the lectures: this very document (to be published and weekly updated on the course website–please do not print before the course is finished and the label "final version" appears at the top).
- Official course textbook: No one, main official text will be used but in preparing these notes; I will probably make heavy use of [Rud86] and [SS05].
- Other books one may consult are [Fol99, FR10, Sim15, Par67, Bog06].
- Two lectures per week: Tue and Thur, 1:30pm–2:50pm in Fine Hall 314.
- People involved:
  - Instructor:
    Jacob Shapiro shapiro@math.princeton.edu
    Office hours: In Fine 603, Thursdays 4pm-5pm starting Feb 13th, before that over zoom on Mondays at 4pm (send email beforehand).
  - Assistants:
    Serban Eugen Cicortas cicortas@princeton.edu
    Office hours: Thursdays 3pm-4pm Fine 401
    ~~Hyungjun Choi hc9325@princeton.edu~~
    ~~Office hours: Tuesdays from 4:30pm - 5:30pm in the common area in front of Jadwin A10.~~

Chayim Lowen cl3172@princeton.edu
Office hours: Tuesdays 4pm-5pm Fine 401

  - Undergraduate assistant:
    Lucas Shedlock ls2597@princeton.edu
    Office hours: ?
    Recitation sessions run by Lucas on Fridays 6pm-9pm in Fine 1001.

- HW will be published on a regular basis but is NOT to be submitted: do it for your own good. Sample solutions will be published one week later.
- Grade: 50% midterm (written in-person) scheduled-midterm; 45% final exam (oral, in person), 5% bonus.
- Attendance policy: *some* extra credit to students who attend lectures regularly and ask questions or point out mistakes.
- Anonymous Ed discussion enabled. Use it to ask questions or to raise issues (technical, academic, logistic) with the course.
- If you alert me about typos and mistakes in this manuscript (unrelated to the sections marked [todo]) I'll grant you *some* extra credit. In doing so, please refer to a version of the document by the date of typesetting.
  - Thanks goes to: Akshat Agarwal (×22), Heyu Li (×9), Natalia Khotiaintseva (×7), Vernon Hughes (×14), Eva Engel (×13), Tal Spiegel (×19), Joshua Lin (×4), Ary Cheng (×5), Jishnu Roychoudhury, Zhuokai Huang (×4), Lydia Boubendir (×21), Olivia Kwon (×∞), Selina Marvit (×5), Kareem Jaber (×2), Rodrigo Salgado Domingos Porto (×2), Joshua Cheng (×4), Kashti Satish Umare (×2), Ron Shvartsman, Kevin Xu, Emmet Weisz (×2).

Semester plan

List of (big) theorems and topics aimed at being included:

- Abstract measure theory.
- The Lebesgue integral: see Section 2.7.
- Radon-Nikodym derivative: see Section 5.
- Fubini, dominated convergence (see Theorem 2.61), monotone convergence (see Theorem 2.47), Fatou (see Lemma 2.53).
- Borel-Cantelli appears in HW3Q9.
- Ergodic theorems.
- Carathéodory's theorem (see Theorem 2.70 and Theorem 2.76).
- The Lebesgue-Stieltjes integral appears in HW3Q5.
- Tempered distributions.
- Hilbert space theory and applications to Fourier Transforms, and partial differential equations.
- Some probability theory?
- Introduction to fractals? Maybe.

Semester plan by date:

- Jan 28th 2025: introduction and abstract measure theory
- Jan 29th 2025: abstract measure theory.

1 Soft introduction

1.1 The Riemann integral and its inadequacies

In a single-value analysis class we are introduced to the rigorous definition of the Riemann integral, which is a $\mathbb{C}$-linear map from functions

$$f : [a, b] \to \mathbb{R}$$

into numbers. In particular, the integral is interpreted in multiple ways as:

1. The average value the function takes:

$$\overline{f} \quad = \quad \frac{1}{b-a} \int_{[a,b]} f \, .$$

2. The (signed) area enclosed between the graph of f, the horizontal axis, and the vertical lines $x = a$, $x = b$.

3. The appropriate continuum generalization to the discrete sum

$$\sum_{n=1}^{N} f(n)$$

   understood in some appropriate sense.

There are various ways to rigorously define the Riemann integral [Rud76]. Let us proceed somewhat informally. The minimal assumption we make on f is that it is bounded (otherwise we do not even ask whether it is Riemann integrable or not). To avoid the complication of partitions[1], let us always consider regular subdivisions of $[a, b]$. Then the lower / upper Riemann sum at N subdivisions is given by

$$L_N(f) := \frac{b-a}{N} \sum_{n=0}^{N-1} \inf\left(\left\{ f(x) \,\middle|\, x \in \left(a + [n, n+1]\frac{b-a}{N}\right) \right\}\right)$$

and

$$U_N(f) \quad := \quad \frac{b-a}{N} \sum_{n=0}^{N-1} \sup\left(\left\{ f(x) \,\middle|\, x \in \left(a + [n, n+1]\frac{b-a}{N}\right) \right\}\right) .$$

Definition 1.1. If the limits $\lim_N L_N(f)$ and $\lim_N U_N(f)$ exists and are equal, we say that f is Riemann integrable on $[a, b]$ and define its Riemann integral as equal to the result of these equal limits:

$$\int_{[a,b]} f := \lim_N L_N(f) = \lim_N U_N(f) \, .$$

We remind the reader of Lebesgue's theorem. For it we need the notion of measure zero set:

Definition 1.2 (Zero measure sets)**.** Let $S \subseteq \mathbb{R}$ be given. We say that S has *zero measure* iff for any $\varepsilon > 0$ there exists a countable collection of open intervals $\{ U_n \}_{n\in\mathbb{N}}$ such that both conditions below hold:

$$\sum_n |U_n| \quad < \quad \varepsilon$$

$$S \quad \subseteq \quad \bigcup_{n\in\mathbb{N}} U_n \, .$$

[1] We note in passing that while we are allowed to restrict to regular subdivisions, we are not allowed to restrict to *both* regular subdivisions *and* always sample at the starting / ending point of each sub-interval.

Theorem 1.3 (Lebesgue's theorem)**.** *The bounded function $f : [a,b] \to \mathbb{R}$ is Riemann integrable iff its set of discontinuities on $[a,b]$ has measure zero.*

Armed with this theorem, it is easy to come up with some examples and counter-examples of Riemann integrable functions:

1. Any continuous function is Riemann integrable.
2. The indicator function on the cantor Set C, $\chi_C : [0,1] \to \mathbb{R}$, is Riemann integrable. Its set of discontinuities is the Cantor set C which has measure zero (though it is uncountable).
3. The indicator function on a fat Cantor set is not Riemann integrable.
4. The indicator function onto the rationals $\chi_{\mathbb{Q}} : [0,1] \to \mathbb{R}$ is not Riemann integrable since it is discontinuous everywhere.

This last example is especially heinous: the set on which $\chi_{\mathbb{Q}}$ is different than zero is countable, it should somehow integrate to zero, since the countable set should not interfere with the uncountablity of the whole interval. Hence, already we see some deficiencies of the Riemann integral: what if the function we are trying to integrate doesn't have zero measure? Couldn't we still say something about its average value? This brings us to the study of just which sets are measurable at all, which we will get to eventually. Another question is what about unbounded functions? The improper Riemann integral addresses this to an extent.

Example 1.4. Consider the function $f : (0,1) \to \mathbb{R}$ given by $x \mapsto \frac{1}{\sqrt{x}}$ which is clearly *unbounded*. However, we may make sense of it formally by defining $f_n : \left[\frac{1}{n}, 1\right] \to \mathbb{R}$ by $x \mapsto \frac{1}{\sqrt{x}}$. For finite $n \in \mathbb{N}$, the function f_n is bounded and Riemann integrable, and

$$\int_{\left[\frac{1}{n},1\right]} f_n = \int_{x\in\left[\frac{1}{n},1\right]} \frac{1}{\sqrt{x}}\mathrm{d}x = 2x^{\frac{1}{2}}\Big|_{x=\frac{1}{n}}^{1} = 2 - \frac{2}{\sqrt{n}} \to 2\,.$$

If we had a finite number of integrable blow ups like this we could somehow manage. But this approach can go horribly wrong:

Example 1.5. Since $(0,1)\cap\mathbb{Q}$ is countable, let $\eta : \mathbb{N} \to (0,1)\cap\mathbb{Q}$ be the bijection which enumerates this set. Define then a sequence of functions $f_n : [0,1] \to \mathbb{R}$ via

$$f_n(x) \quad := \quad \begin{cases} (x-\eta_n)^{-\frac{1}{2}} & x > \eta_n \\ 0 & x \leq \eta_n \end{cases} \qquad (n \in \mathbb{N}, x \in [0,1])\ .$$

Then define $f : [0,1] \to [0,\infty]$ via

$$f(x) := \sum_{n=1}^{\infty} \frac{f_n(x)}{2^n} \qquad (x \in [0,1])\ .$$

f has the weird property that it is unbounded on every open subinterval of $[0,1]$, since each one contains a rational number. *Hence f is not Riemann integrable on every subinterval of $[0,1]$ which is not a singleton.*

But somehow we still feel like we should be able to assign an area under the graph of f, since we can do so for each f_n:

$$\begin{aligned} \int_{[0,1]} f_n &= \lim_{\varepsilon\to 0^+} \int_{[0,\eta_n-\varepsilon]} f_n + \int_{[\eta_n+\varepsilon,1]} f_n \\ &= \lim_{\varepsilon\to 0^+} \int_{x\in[\eta_n+\varepsilon,1]} (x-\eta_n)^{-\frac{1}{2}}\,\mathrm{d}x \\ &= \lim_{\varepsilon\to 0^+} 2(x-\eta_n)^{+\frac{1}{2}}\Big|_{x=\eta_n+\varepsilon}^{1} \\ &= 2\sqrt{1-\eta_n}\,. \end{aligned}$$

and somehow it should equal

$$\int_{[0,1]} f = \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} f_n = \sum_{n=1}^{\infty} 2^{-n+1} \sqrt{1-\eta_n} \leq \sum_{n=1}^{\infty} 2^{-n+1} < \infty\,.$$

From the more practical and less theoretical perspective, a much more severe limitation of the Riemann integral is how it behaves with limits. Namely, we have

Theorem 1.6. *Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of bounded Riemann integrable functions which converges* uniformly *to the bounded function $\lim_n f_n : [a,b] \to \mathbb{R}$. Then $\lim_n f_n : [a,b] \to \mathbb{R}$ is also Riemann integrable, and*

$$\lim_n \int_{[a,b]} f_n = \int_{[a,b]} \lim_n f_n\,.$$

However, establishing *uniform converges* is notoriously difficult, in fact it is false in many interesting applications. For instance, letting $\eta : \mathbb{N} \to (0,1) \cap \mathbb{Q}$ again be the bijection which enumerates its codomain, define

$$f_n := \chi_{\{\, \eta_j \mid j \in [1,n] \cap \mathbb{Z} \,\}}\,.$$

Clearly each f_n is bounded and Riemann integrable. Also, $\lim_n f_n = \chi_{\mathbb{Q}\cap[0,1]}$ pointwise. But as we saw above, this limit is *not* Riemann integrable. We are looking for a way to exchange integration and limit without uniform convergence. We shall see that to do so we need to invent a new, more robust notion of integration.

1.2 Intuitive difference between Riemann and Lebesgue integration

We will see that conceptually, while the Riemann integral divides the *domain* into small pieces and measures the area of each small rectangle, the Lebesgue integral does things somewhat sophisticatedly. To calculate the Lebesgue integral, we first need the notion of a *measure* which generalizes volume on Euclidean space to arbitrary spaces. Then we divide the *codomain* into small chunks and ask what is the measure of the preimage of that chunk in the domain. This turns out to give a more robust definition of the integral, which is not so susceptible to discontinuities and behaves better with limits. For that reason we now turn to abstract measure theory.

2 Abstract measure theory [Rudin]

We now want to define the concept of *measurability* and ultimately assign a *measure* to measurable sets. This will be useful when we define the Lebesgue integral, and furthermore, this has applications in probability theory where measurable sets may be considered as those events for which a probability can be calculated.

2.1 Measurable sets and measurable functions

On a set X, we now want to define a *system of subsets* much like $\text{Open}(X)$ is a system of subsets with certain axioms.

Definition 2.1 (σ-algebra)**.** Let X be a set. A collection $\mathfrak{M} \subseteq \mathscr{P}(X)$ is called *a σ-algebra in X* iff $\mathfrak{M}$ obeys the following conditions:

1. $X \in \mathfrak{M}$ (contains the whole space).
2. $X \setminus A \in \mathfrak{M}$ for each $A \in \mathfrak{M}$ (closed under complements).
3. If $\{ A_n \}_{n\in\mathbb{N}}$ is a sequence of subsets such that $A_n \in \mathfrak{M}$ for each $n \in \mathbb{N}$ then

$$\bigcup_{n\in\mathbb{N}} A_N \in \mathfrak{M}\,.$$

(closed under countable unions).

The tuple $(X, \mathfrak{M})$ where $\mathfrak{M}$ is a σ-algebra on X, is together called *a measurable space.*

Note that this definition automatically implies: (1) closure with respect to countable intersections via De Morgan and (2) $\varnothing \in \mathfrak{M}$.

Remark 2.2 (Etymology). The prefix σ denotes the closure w.r.t. countable unions. If we had merely closure wr.t. finite unions this would be called an *algebra*.

Contrast this with the notion of a *topology* on a given set X:

Definition 2.3 (Topology)**.** Let X be a set. A collection $\mathscr{T} \subseteq \mathscr{P}(X)$ is called a topology on X iff $\mathscr{T}$ obeys the following conditions:

1. $X, \varnothing \in \mathscr{T}$ (contains the whole space and the empty set).
2. $\bigcap_{j=1}^{n} U_j \in \mathscr{T}$ if $U_1, \cdots, U_n \in \mathscr{T}$ (closed under finite intersections).
3. $\bigcup_{\alpha \in \mathscr{I}} U_\alpha \in \mathscr{T}$ if $U_\alpha \in \mathscr{T}$ for any $\alpha \in \mathscr{I}$, where $\mathscr{I}$ is an arbitrary set (*not* necessary countable) (closed under arbitrary unions).

The tuple $(X, \mathscr{T})$, if $\mathscr{T}$ is a topology on X, is together called *a topological space*.

When dealing with a topological space X, it is often convenient to denote its (already defined) topology as $\operatorname{Open}(X)$. Similarly, given a measure space X, we denote by $\operatorname{Msrbl}(X)$ the σ-algebra in it, should it be understood from the context.

Definition 2.4 (Measurable function)**.** Let $f : X \to Y$ be given where X, Y are two measure spaces. We say that f is *measurable* iff $f^{-1}(A) \in \operatorname{Msrbl}(X)$ for each $A \in \operatorname{Msrbl}(Y)$.

Note that Rudin [Rud86] defines measurable function slightly differently (his codomains are always topological spaces).

Claim 2.5. The composition of two measurable functions is again measurable.

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be two measurable functions between measure spaces. Let $A \in \operatorname{Msrbl}(Z)$. Then $g^{-1}(A) \in \operatorname{Msrbl}(Y)$. But then $f^{-1}\left(g^{-1}(A)\right) \in \operatorname{Msrbl}(X)$. But $f^{-1}\left(g^{-1}(A)\right) = (g \circ f)^{-1}(A)$ so we conclude $g \circ f$ is measurable. □

Example 2.6 (The trivial σ-algebra)**.** Given a set X, we may consider its power set $\mathscr{P}(X)$ as a σ-algebra on it. It is called *the trivial* or *largest* σ-algebra on X. The smallest one is of course $\{ \varnothing, X \}$.

Example 2.7. Take $X := \{ 1, 2, 3, 4 \}$. Then a possible σ-algebra is $\{ \varnothing, \{ 1, 2 \}, \{ 3, 4 \}, \{ 1, 2, 3, 4 \} \}$.

Example 2.8. Let $A \in \mathscr{P}(X)$. Then $\{ \varnothing, A, X \setminus A, X \}$ is the smallest σ-algebra which contains A.

We may consider the category of measure spaces, in which measurable functions are precisely the morphisms.

Remark 2.9. A topology *need not* be a σ-algebra: it could fail to contain complements.

Claim 2.10. An arbitrary intersection of σ-algebras is again a σ-algebra. Not so for unions.

Proof. TODO, fix this: (even though a-priori it lies within an intersection of σ-algebras). Let then $A_n \in \sigma(\mathfrak{F})$ for every $n \in \mathbb{N}$. Let $\mathfrak{M} \in \Omega$. Then $A_n \in \mathfrak{M}$ by definition, so $\bigcup_n A_n \in \mathfrak{M}$, as $\mathfrak{M}$ is itself a σ-algebra. But since $\mathfrak{M} \in \Omega$ was arbitrary, the union lies in the intersection $\sigma(\mathfrak{F})$. The other two properties, complements and the entire space, are verified in the same manner.
TODO: provide a counter-example. □

Definition 2.11 (σ-algebra generated by a function)**.** Let $f : X \to Y$ with Y a measure space and X a set. Then the σ-algebra generated by f is a σ-algebra on X, denoted by $\sigma(f)$, given by

$$\sigma(f) := \left\{ f^{-1}(A) \,\middle|\, A \in \operatorname{Msrbl}(Y) \right\} .$$

One may then rephrase and say that, if X already had a measure space structure, then f is measurable w.r.t. it iff $\sigma(f) \subseteq \operatorname{Msrbl}(X)$. Cf. with initial topology.

Remark 2.12. Recall that arbitrary intersections of σ-algebras are again σ-algebras, Claim 2.10. As such, if $A \in \text{Msrbl}(X)$ then we have a σ-algebra structure on A given by the inclusion map $\iota : A \to X$ and $\sigma(\iota)$.

Example 2.13 (Baire σ-algebra)**.** Let X be a *topological space*. Define a σ-algebra $\mathcal{M}$ in X via the following criterion: $\mathcal{M}$ is the smallest σ-algebra so that all functions $f : X \to \mathbb{C}$ which are continuous and compactly supported are measurable.

Theorem 2.14 (σ-algebra generated by a collection of subsets)**.** *Let $\mathcal{F} \subseteq \mathcal{P}(X)$ with X some set. Then, there exists a* smallest *(in the sense of set inclusion) σ-algebra $\sigma(\mathcal{F})$ in X such that $\mathcal{F} \subseteq \sigma(\mathcal{F})$. We call $\sigma(\mathcal{F})$ the σ-algebra generated by $\mathcal{F}$.*

Proof. (See [Rud86] Theorem 1.10) Let Ω be the family of all σ-algebras in X which contain $\mathcal{F}$. Of course $\mathcal{P}(X)$ is in Ω, so it is not empty. Define the *set*

$$\sigma(\mathcal{F}) := \bigcap_{\mathfrak{M} \in \Omega} \mathfrak{M}.$$

Clearly $\mathcal{F} \subseteq \sigma(\mathcal{F})$ by construction. The fact that $\sigma(\mathcal{F})$ is itself a σ-algebra and not just a set follows via Claim 2.10. □

Theorem 2.15. *If $\mathcal{F} \subseteq \mathcal{P}(X)$ is countable (i.e., $|\mathcal{F}| = \aleph_0$) then $\sigma(\mathcal{F}) \subseteq \mathcal{P}(X)$ is either finite or $|\sigma(\mathcal{F})| = 2^{\aleph_0}$.*

Proof. (*Chayim Lowen*) See HW3Q3. □

Definition 2.16 (Borel sets)**.** Given a topology on X, by Theorem 2.14 there is a σ-algebra generated by $\text{Open}(X)$: $\sigma(\text{Open}(X))$. The elements of $\sigma(\text{Open}(X))$ are called *the Borel sets of X*. In particular:

- Closed sets are also Borel sets, since they are the complements of open sets.
- Countable unions of closed sets are also Borel sets. These are called F_σ's (F=closed, σ=union (summe)). For example $[a, b)$ is a F_σ set of $\mathbb{R}$ with its standard topology.
- Countable intersections of open sets are also Borel sets. These are called G_δ's (G=open, δ=intersection (durchschnitt)). For example $[a, b)$ is also a G_δ set of $\mathbb{R}$ with its standard topology.

We denote this special σ-algebra of Borels sets by $\mathcal{B}(X) := \sigma(\text{Open}(X))$.

Thus, given a topology on X we are automatically provided with the Borel σ-algebra on it! If we don't specify any other σ-algebra on a (otherwise topological) space, we shall always mean the Borel σ-algebra.

Claim 2.17. Let $f : X \to Y$ be a mapping between two measurable spaces where

$$\text{Msrbl}(Y) = \sigma(\mathcal{F})$$

for some $\mathcal{F} \subseteq \mathcal{P}(Y)$. Assume further that

$$f^{-1}(F) \in \text{Msrbl}(X) \qquad (F \in \mathcal{F}).$$

Then f is measurable.

Proof. We may consider the set

$$\mathfrak{M} := \left\{ A \in \mathcal{P}(Y) \mid f^{-1}(A) \in \text{Msrbl}(X) \right\}.$$

Cf. with final topology. We may verify it is stable under complements and countable unions: If $A \in \mathfrak{M}$, we want to show that $A^c \in \mathfrak{M}$, i.e., that $f^{-1}(A^c) \in \text{Msrbl}(X)$. But

$$f^{-1}(A^c) = \left[f^{-1}(A)\right]^c$$

and $\text{Msrbl}(X)$ is closed under complements so we are finished. Next, if $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathfrak{M}$ then

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(A_n)$$

and since $\text{Msrbl}(X)$ is closed under countable unions, we have $f^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \in \text{Msrbl}(X)$ so

$$\bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{M}$$

so, $\mathfrak{M}$ is itself a σ-algebra in Y. By hypothesis, $\mathcal{F} \subseteq \mathfrak{M}$ and so actually

$$\mathcal{F} \subseteq \sigma(\mathcal{F}) \subseteq \mathfrak{M}$$

since, by construction, $\sigma(\mathcal{F})$ is the *smallest* σ-algebra which contains $\mathcal{F}$. But by $\sigma(\mathcal{F}) \subseteq \mathfrak{M}$ we learn that f is measurable. □

Corollary 2.18. *Let $f : X \to Y$ be a mapping between where X is a measurable space and Y is a topological space, such that $f^{-1}(U) \in \text{Msrbl}(X)$ for any $U \in \text{Open}(Y)$. Then f is measurable w.r.t. $\text{Msrbl}(X)$ and $\mathcal{B}(Y)$ respectively. Similarly, if $f^{-1}(F) \in \text{Msrbl}(X)$ for any $F \in \text{Closed}(Y)$ then f is measurable w.r.t. $\text{Msrbl}(X)$ and $\mathcal{B}(Y)$ respectively.*

Proof. We know that the Borel σ-algebra is generated by the open sets

$$\mathcal{B}(Y) \equiv \sigma(\text{Open}(Y))$$

but in fact it may also be generated by the closed sets (one may verify this...), i.e.,

$$\sigma(\text{Open}(Y)) = \sigma(\text{Closed}(Y)) .$$

□

This then coincides with Rudin's definition of measurable function, since he *only* considers maps whose codomains are topological spaces and then restricts to the special case of the Borel σ-algebra on them.

Theorem 2.19 (Rudin's Theorem 1.8)**.** *Let $u, v : X \to \mathbb{R}$ be two measurable functions ($\mathbb{R}$ is considered a measure space w.r.t. $\mathcal{B}(\mathbb{R})$). Let $\varphi : \mathbb{R}^2 \to Y$ be continuous where Y is some topological space. Let $h : X \to Y$ be given by*

$$X \ni x \mapsto \varphi(u(x), v(x)) \in Y .$$

Then h is measurable w.r.t. $\text{Msrbl}(X)$ and $\mathcal{B}(Y)$.

Proof. The function $f : X \to \mathbb{R}^2$ given by $u \times v$. We have $h = \varphi \circ f$, so we only have to show f is measurable. Let R be any open rectangle on the plane with sides parallel to the axes: $R = I_1 \times I_2$ for two open intervals I_1, I_2 and so

$$f^{-1}(R) \quad = \quad u^{-1}(I_1) \cap v^{-1}(I_2)$$

which is measurable by assumption on u, v. Since every open set $V \in \text{Open}\left(\mathbb{R}^2\right)$ is the countable union of such rectangles R_i, we find

$$f^{-1}(V) = f^{-1}\left(\bigcup_{i=1}^{\infty} R_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(R_i)$$

and hence $f^{-1}(V)$ is measurable and so is f. □

Theorem 2.20 (Rudin's Theorem 1.9)**.** *Let X be a measure space. Then*

1. *If $u, v : X \to \mathbb{R}$ are measurable then $f : X \to \mathbb{C}$ defined by $f := u + iv$ is measurable.*
2. *If $f : X \to \mathbb{C}$ is measurable then $\mathbb{Re}\{f\}, \mathbb{Im}\{f\}$ and $|f|$ are measurable functions from $X \to \mathbb{R}$.*
3. *If $f, g : X \to \mathbb{C}$ are measurable then $f + g$ and fg are too.*
4. *If $A \in \operatorname{Msrbl}(X)$ then $\chi_A : X \to \mathbb{R}$ is a measurable function.*
5. *If $f : X \to \mathbb{C}$ is measurable then there exists some $\alpha : X \to \mathbb{C}$ measurable such that $f = \alpha |f|$.*

Proof. We only prove the last statement. Set $E := f^{-1}(\{0\})$ (a measurable set) and $Y := \mathbb{C} \setminus \{0\}$. Let

$$\begin{aligned} \varphi : Y &\to \mathbb{C} \\ z &\mapsto \frac{z}{|z|} \, . \end{aligned}$$

Define

$$\alpha(x) := \varphi(f(x) + \chi_E(x)) \qquad (x \in X) \, .$$

Show that φ is continuous on Y to conclude. □

In what follows, it will be convenient to consider the *extended real line* $[-\infty, \infty]$, see Appendix A. In particular we shall always consider it as a measure space w.r.t. $\mathcal{B}([-\infty, \infty])$ unless otherwise specified.

Theorem 2.21. *Let $f : X \to [-\infty, \infty]$ be a map with X a measure space. Here we consider $[-\infty, \infty]$ as the extended real line with its topology, see Appendix A. Then if*

$$f^{-1}((\alpha, \infty]) \in \operatorname{Msrbl}(X) \qquad (\alpha \in \mathbb{R})$$

then f is measurable w.r.t. $\operatorname{Msrbl}(X)$ and $\mathcal{B}([-\infty, \infty])$.

Proof. The set $(\alpha, \infty]$ is already open in $[-\infty, \infty]$ so our goal is to build *any* of the basis elements of $[-\infty, \infty]$ using this basic open set. To that end, let

$$\Omega := \left\{ E \subseteq [-\infty, \infty] \,\middle|\, f^{-1}(E) \in \operatorname{Msrbl}(X) \right\} \, .$$

Let $\alpha \in \mathbb{R}$ and $\{\alpha_n\}_n \to \alpha$ *from below*. Then $(\alpha_n, \infty] \in \Omega$ by hypothesis, and we have

$$[-\infty, \alpha) = \bigcup_{n=1}^{\infty} [-\infty, \alpha_n] = \bigcup_{n=1}^{\infty} (\alpha_n, \infty]^c$$

so we get the other type of basic open set, $[-\infty, \alpha)$. Next, using

$$(\alpha, \beta) = [-\infty, \beta) \bigcap (\alpha, \infty]$$

we see that since every open set of $[-\infty, \infty]$ is a *countable* union of segments of the above types, so that Ω contains all open sets of $[-\infty, \infty]$ and hence f is measurable. □

2.2 Limits of measurable functions

Recall the definition of the lim inf and lim sup: Let $\{a_n\}_{n \in \mathbb{N}} \subseteq \overline{\mathbb{R}}$ be a given sequence. Then

$$\liminf_{n \to \infty} a_n \equiv \lim_{n \to \infty} \left(\inf_{m \geq n} a_m \right) = \sup_{n \in \mathbb{N}} \inf_{m \geq n} a_m \, . \tag{2.1}$$

Similarly,

$$\limsup_{n \to \infty} a_n \equiv \lim_{n \to \infty} \left(\sup_{m \geq n} a_m \right) = \inf_{n \in \mathbb{N}} \sup_{m \geq n} a_m \, .$$

Evidently, we always have

$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$$

and if the limit of $\{ a_n \}_n$ actually exists then both are equal to that limit.

Theorem 2.22. *If $f_n : X \to [-\infty,\infty]$ is a sequence of measurable functions then $\sup_{n\in\mathbb{N}} f_n : X \to [-\infty,\infty]$ defined by*

$$X \ni x \mapsto \sup_{n\in\mathbb{N}} f_n(x)$$

and $\limsup_{n\to\infty} f_n : X \to [-\infty,\infty]$ defined by

$$X \ni x \mapsto \limsup_{n\to\infty} (f_n(x))$$

are both measurable.

Proof. Let us denote $g := \sup_{n\in\mathbb{N}} f_n$ and $h := \limsup_{n\to\infty} f_n$. Then, from the definition of g it follows that

$$g^{-1}((\alpha,\infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}((\alpha,\infty]) \, .$$

Indeed, let us show this. If $x \in g^{-1}((\alpha,\infty])$ then $g(x) > \alpha$. That means $\sup_{n\in\mathbb{N}} f_n(x) > \alpha$ so in particular there must exist $n \in \mathbb{N}$ so that $f_n(x) > \alpha$. Alternatively, if $x \in \bigcup_{n=1}^{\infty} f_n^{-1}((\alpha,\infty])$ then there exists some $n \in \mathbb{N}$ for which $f_n(x) > \alpha$. This in particular implies $g(x) > \alpha$.

We conclude that g is measurable. We write

$$h = \inf_{k\geq 1} \sup_{i\geq k} f_i$$

so that h is also measurable by similar representations. □

Corollary 2.23. *We have*

1. *The limit of every pointwise convergent sequence of complex measurable functions is measurable.*
2. *If $f, g : X \to \overline{\mathbb{R}}$ are measurable then so are $\max(\{ f, g \})$ and $\min(\{ f, g \})$.*
3. *In particular, so are $f^+ \equiv \max(\{ f, 0 \})$ and $f^- = -\min(\{ f, 0 \})$.*

We may always decompose any $\overline{\mathbb{R}}$-valued function into its positive and negative parts as follows

$$f = f^+ - f^-$$

with $f^\pm$ the positive and negative parts of f, and $|f| = f^+ + f^-$ [2].

2.3 Simple functions

We shall build a theory of integration starting from primitive functions and then take limits. This will proceed as follows. Given any function

$$f : X \to \mathbb{C}$$

we write it as

$$f = \mathbb{Re}\{f\} + \mathrm{i}\,\mathbb{Im}\{f\} \, .$$

Then we write

$$\mathbb{Re}\{f\} = \mathbb{Re}\{f\}^+ - \mathbb{Re}\{f\}^-$$

and similarly for the imaginary part, so that any complex function is the (complex) linear combination of four *nonnegative* functions. Measurability is inherited by all four. Then we want to approximate each nonnegative function with even simpler objects, simple functions.

[2] Note a certain minimal property for these objects: Note that if $f = g - h$ with $g, h \geq 0$ then $f^+ \leq g$ and $f^- \leq h$. This is because $f \leq g$ and $0 \leq g$ clearly implies $\max(\{ f, 0 \}) \leq g$.

Definition 2.24 (Simple function)**.** Let X be a measure space and $s : X \to \mathbb{C}$. If $|\mathrm{im}(s)| < \infty$ then s is called a *simple function*. If in addition, $\mathrm{im}(s) \subseteq [0, \infty)$ then s is called a *nonnegative* simple function. We are *not* including $\pm\infty$ as part of $\mathbb{C}$ so that simple functions, by definition, *cannot* take on the values $\pm\infty$.

Clearly simple functions always take on the form

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$$

for some $n \in \mathbb{N}$, $\alpha_i \in \mathbb{C}$ and $A_i \equiv \{ x \in X \mid s(x) = \alpha_i \}$.

Claim 2.25. A simple function $X \to \mathbb{C}$ of the form $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ is measurable iff $A_i \in \mathrm{Msrbl}(X)$ for $i = 1, \dots, n$.

Proof. (We consider $\mathbb{C}$ w.r.t. the Borel sigma algebra, as usual). By Corollary 2.18 we only need to check that the pre-image of closed sets is msrbl. Hence let $F \subseteq \mathbb{C}$ be closed. If F does not contain any of the points α_i then $s^{-1}(F) = \varnothing \in \mathrm{Msrbl}(X)$. If F contains $\alpha_{i_1}, \dots, \alpha_{i_k}$ then

$$s^{-1}(F) = \bigcup_{j=1}^{k} A_{i_j}$$

and the union of measurable sets is measurable. Conversely, if s is measurable, take the (closed) singleton $\{ \alpha_i \}$ to verify that $A_i \in \mathrm{Msrbl}(X)$. $\square$

Remark 2.26. The product and sum of simple functions is again a simple function. Scalar multiplication also preserves this property. Hence they form an algebra over $\mathbb{C}$. As we shall see, they however are not closed under limits.

Now we want to establish that any nonnegative measurable function may be approximated by simple functions *from below*.

Theorem 2.27 (Approximation by simple functions)**.** *Let $f : X \to [0, \infty]$ be measurable. Then there exist simple measurable functions $s_n : X \to [0, \infty)$ such that*

1. $0 \leq s_1 \leq s_2 \leq \dots \leq f$.
2. $s_n \to f$ *pointwise.*

Proof. (*Thanks to Lydia Boubendir*)

For every $n \in \mathbb{N}$, define

$$\begin{aligned} \varphi_n : [0, \infty] &\to [0, \infty) \\ t &\mapsto \begin{cases} 2^{-n} \lfloor 2^n t \rfloor & 0 \leq t < n \\ n & t \in [n, \infty] \end{cases} \end{aligned}$$

which is depicted, at $n = 3$ in Figure 1. The function φ_n converges to $t \mapsto t$ as $n \to \infty$. It is doing that in two ways simultaneously:

1. The region over which it does *not* resemble the identity function, $[n, \infty]$ keeps shrinking.
2. The region over which it does resemble the identity function, it becomes finer and finer at approximation the identity function there by subdividing $[0, n]$ into roughly 2^n sub-intervals and being saw-toothed there.

First, note that at each fixed $n \in \mathbb{N}$, φ_n is a Borel function. Indeed, it is a simple function that takes on basically 2^n values on *intervals* and as such it is measurable. For monotonicity, want to establish

$$\varphi_n(t) \quad \leq \quad \varphi_{n+1}(t) \qquad (t \in [0, \infty], n \in \mathbb{N}) \,.$$

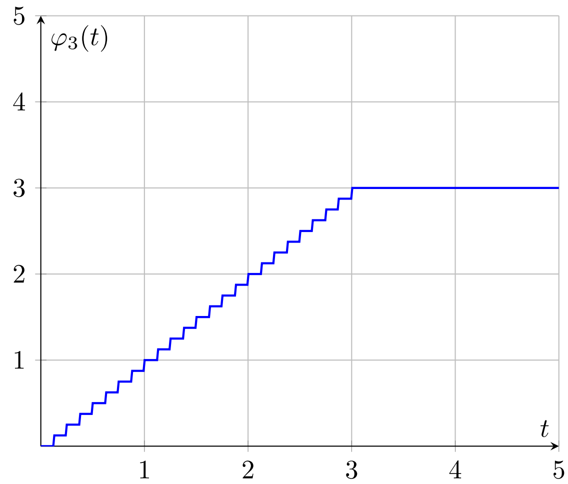

Figure 1: The function φ_3 approximating the identity.

For $t \geq n+1$ this is easy because $n \leq n+1$. For $n \leq t \leq n+1$ as well, since there,

$$\varphi_{n+1}(t) \equiv 2^{-n-1}\left\lfloor 2^{n+1}t\right\rfloor \geq 2^{-n-1}\left\lfloor 2^{n+1}n\right\rfloor = n \equiv \varphi_n(t) \ .$$

Finally, we want to show that for $t \in [0, n]$,

$$\begin{array}{ccc} 2^{-n-1}\left\lfloor 2^{n+1}t\right\rfloor & \overset{?}{\geq} & 2^{-n}\left\lfloor 2^n t\right\rfloor \\ & \updownarrow & \\ \left\lfloor 2^{n+1}t\right\rfloor & \overset{?}{\geq} & 2\left\lfloor 2^n t\right\rfloor \ . \end{array}$$

This last relation is implied by the relation

$$\lfloor x\rfloor \leq \frac{1}{2}\lfloor 2x\rfloor \qquad (x \geq 0)$$

which is always true. Indeed, for all $m \in \mathbb{N}$, if $x \in [m, m+1)$, then $\lfloor x\rfloor = m$ whereas

$$\frac{1}{2}\lfloor 2x\rfloor = m + \frac{1}{2}\chi_{\left[m+\frac{1}{2}, m+1\right)}(x) \ .$$

Now we set

$$s_n := \varphi_n \circ f$$

which automatically fulfills both of our constraints, using the fact that the composition of measurable functions is measurable Claim 2.5. ☐

2.4 Measures

We now come to the notion of *measure* which for us is to be understood as a generalization of volume in $\mathbb{R}^n$ to much more exotic sets (yet they still have to be measurable), or of a *weight* of sets.

Definition 2.28 (Measure)**.** A *complex measure* is a map

$$\mu : \text{Msrbl}\,(X) \to \mathbb{C} \cup \{\,\infty\,\}$$

which is countably additive, i.e.,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\,(A_n) \qquad (A_n \in \text{Msrbl}\,(X) : A_n \cap A_m = \varnothing \forall n \neq m) \tag{2.2}$$

and for which $\exists A : \mu\,(A) < \infty$ (otherwise it is not very interesting). If $\text{im}\,(\mu) \subseteq [0,\infty]$ then we say μ is a *positive measure*.

Note: Despite the "logical" Definition 2.28, when using the term *complex measure* Rudin assumes μ never takes on the value ∞ (unlike when we use the phrase *positive measure*). Following him, so we will really only consider the dichotomy:

- Either μ takes values in $\mathbb{C}$ (complex measure).
- Or μ takes values in $[0,\infty]$ (positive measure).

Theorem 2.29. *Let* $\mu : \text{Msrbl}\,(X) \to [0,\infty]$ *be a positive measure. Then*

1. $\mu\,(\varnothing) = 0$ (so in particular (2.2) holds also for finitely many unions).
2. (*monotonicity*) $A \subseteq B$ implies
$$\mu\,(A) \leq \mu\,(B) \tag{2.3}$$
for all $A, B \in \text{Msrbl}\,(X)$.
3. μ may be approximated from "inside" as follows:
$$\lim_{n\to\infty} \mu\,(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \tag{2.4}$$
for all increasing sequences $A_n \in \text{Msrbl}\,(X)$: $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$.
4. μ may be approximated from "outside" as follows:
$$\lim_{n\to\infty} \mu\,(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right) \tag{2.5}$$
for all decreasing sequences $A_n \in \text{Msrbl}\,(X)$: $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ with $\mu\,(A_1)$ assumed finite.

Proof. By assumption, there exists $B \in \text{Msrbl}\,(X)$ with $\mu\,(B) < \infty$. Define now a sequence $A_1 := B$, $A_j := \varnothing$ for all $j \geq 2$. This sequence obeys the conditions of (2.2) since it is pairwise disjoint. Hence we find

$$\infty > \mu\,(B) = \mu\,(B) + \sum_{j=2}^{\infty} \mu\,(\varnothing)$$

and the only way this equation could hold is if $\mu\,(\varnothing) = 0$. Now that we know $\mu\,(\varnothing) = 0$, we have additivity for finite sequences.

For monotonicity, given $A, B \in \text{Msrbl}\,(X)$ with $A \subseteq B$, let us decompose $B = A \cup (B \setminus A)$ which are now disjoint. Hence additivity implies

$$\mu\,(B) = \mu\,(A) + \mu\,(B \setminus A)$$

and using positivity of the measure, we find this is larger than or equal to $\mu\,(A)$.

Let us now establish the approximation properties. To do so, given any increasing sequence $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, we decompose it into disjoint parts as follows:

$$\begin{aligned} B_1 &:= A_1 \\ B_n &:= A_n \setminus A_{n-1} \qquad (n \geq 2)\ . \end{aligned}$$

Note that $A_n = \bigcup_{j=1}^{n} B_j$. So by (2.2) we find

$$\mu(A_n) = \sum_{j=1}^{n} \mu(B_j)$$

and moreover, since $\bigcup_n A_n = \bigcup_n B_n$, we get

$$\mu\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \, .$$

The result now follows by taking the limit $n \to \infty$ on the penultimate displayed equation.

For approximation from outside, we make the following new variables.

$$C_n := A_1 \setminus A_n \qquad (n \geq 1) \, .$$

This implies $C_1 \subseteq C_2 \subseteq C_3 \subseteq \cdots$ and

$$\mu(C_n) = \mu(A_1) - \mu(A_n) \, .$$

Moreover, $A_1 \setminus (\bigcap A_n) = \bigcup_n C_n$, so now we may invoke the previous statement on the sequence C_n to get

$$\begin{aligned} \mu(A_1) - \mu\left(\bigcap_n A_n\right) &= \mu\left(A_1 \setminus \bigcap_n A_n\right) \\ &= \mu\left(\bigcup_n C_n\right) \\ &= \lim_n \mu(C_n) \\ &= \lim_n (\mu(A_1) - \mu(A_n)) \\ &= \mu(A_1) - \lim_n \mu(A_n) \end{aligned}$$

from which our result follows. □

Our main example for a positive measure will be *the Lebesgue measure* on $\mathbb{R}^n$, but it will be a little while before we can define it.

Example 2.30 (Counting measure)**.** Let $\mathrm{Msrbl}(X) = \mathscr{P}(X)$ and define $c : \mathrm{Msrbl}(X) \to [0, \infty]$ via

$$S \mapsto |S|$$

(the cardinality of a set, ∞ if it is countable or higher). c is called the counting measure. Usually we only define the counting measure if X is countable.

Example 2.31 (Unit mass; "Dirac delta measure")**.** Let $\mathrm{Msrbl}(X) = \{ \varnothing, X, \{ x_0 \}, X \setminus \{ x_0 \} \}$ be a σ-algebra and define $\delta_{x_0} : \mathrm{Msrbl}(X) \to [0, \infty]$ by

$$S \mapsto \begin{cases} 1 & x_0 \in S \\ 0 & x_0 \notin S \end{cases} \equiv \chi_S(x_0) \, .$$

In cryptic symbols,

$$\delta_{x_0} = \chi_{\cdot}(x_0) \, .$$

δ_{x_0} is called the unit mass concentrated at x_0. It is closely related to the *Dirac delta function*. While the latter is *not* actually a function (it is a distribution), the unit mass is a very simple object.

Example 2.32. If we take the counting measure c on $\mathbb{N}$ and set $A_n := \mathbb{N}_{\geq n}$ then $\bigcap_n A_n = \varnothing$ and yet $\mu(A_n) = \infty$. This does not violate the theorem above since the assumption $\mu(A_1) < \infty$ is clearly violated here.

Definition 2.33 (Complete measure)**.** A measure $\mu : \mathrm{Msrbl}(X) \to \mathbb{C}$ is called *complete* iff for any $Z \in \mathrm{Msrbl}(X)$ such that $\mu(Z) = 0$, any subset $A \subseteq Z$ is also measurable $A \in \mathrm{Msrbl}(X)$.

Example 2.34. We will later on see that the Lebesgue measure μ on $\mathbb{R}$ is not complete if we insist its domain is $\mathcal{B}(\mathbb{R})$ since there are Lebesgue measurable subsets which are not Borel.

Theorem 2.35. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Define*

$$\overline{\mathcal{M}} := \{ E \in \mathcal{P}(X) \mid \exists A_E, B_E \in \mathcal{M} : A_E \subseteq E \subseteq B_E \wedge \mu(B_E \setminus A_E) = 0 \}$$

and $\overline{\mu} : \overline{\mathcal{M}} \to \mathbb{C}$ via

$$\overline{\mu}(E) \quad := \quad \mu(A_E) \qquad \left(E \in \overline{\mathcal{M}}\right) .$$

Then $\overline{\mathcal{M}}$ is a σ-algebra in X and $\overline{\mu}$ is a measure.

Proof. TODO ☐

Claim 2.36. Let X, Y be two measurable spaces and $\mu : \text{Msrbl}(X) \to [0, \infty]$ be a *complete* measure. If $f : X \to Y$ is measurable and $g : X \to Y$ equals to f μ-almost-everywhere then g is also measurable.

Proof. Let

$$N \quad := \quad \{ x \in X \mid f(x) \neq g(x) \} .$$

By hypothesis,

$$\mu(N) = 0 .$$

In particular part of the hypothesis is that $N \in \text{Msrbl}(X)$!

Let $A \in \text{Msrbl}(Y)$. We want to show that $g^{-1}(A) \in \text{Msrbl}(X)$. Again, by hypothesis, $f^{-1}(A) \in \text{Msrbl}(X)$.

$$\begin{aligned} g^{-1}(A) &\equiv \{ x \in X \mid g(x) \in A \} \\ &= [\{ x \in X \mid g(x) \in A \} \cap N] \sqcup [\{ x \in X \mid g(x) \in A \} \cap N^c] \\ &= [\{ x \in X \mid g(x) \in A \} \cap N] \sqcup [\{ x \in X \mid f(x) \in A \} \cap N^c] \\ &= [\{ x \in X \mid g(x) \in A \} \cap N] \sqcup \left[f^{-1}(A) \cap N^c\right] \end{aligned}$$

Since μ is complete, $\mu(N) = 0$ and

$$\{ x \in X \mid g(x) \in A \} \cap N \subseteq N ,$$

$$[\{ x \in X \mid g(x) \in A \} \cap N] \in \text{Msrbl}(X) .$$

Hence, $g^{-1}(A)$ is a σ-algebra-closed combination of procedures on measurable sets, and is hence measurable itself. ☐

Definition 2.37 (σ-finite measure)**.** A measure $\mu : \text{Msrbl}(X) \to \mathbb{C}$ is called *σ-finite* iff $\forall A \in \text{Msrbl}(X)$, there is a sequence $\{ E_i \}_{i=1}^{\infty} \subseteq \text{Msrbl}(X)$ such that $A \subseteq \bigcup_{i=1}^{\infty} E_i$ and $\mu(E_i) < \infty$ for all $i \in \mathbb{N}$.

Example 2.38. The counting measure $c : \mathcal{P}(X) \to [0, \infty]$ is *not* σ-finite if X is uncountable.

Example 2.39. We will see that the Lebesgue measure on $\mathbb{R}$ is σ-finite.

2.5 Integrating positive functions

Given a *positive* measure $\mu : \text{Msrbl}(X) \to [0, \infty]$, we now proceed to define the Lebesgue integral associated to μ.

Definition 2.40 (The Lebesgue integral of positive simple measurable functions)**.** Let $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ be a nonnegative measurable simple function. Then we define the integral of s on a set w.r.t. μ as

$$\int_E s\mathrm{d}\mu := \sum_{i=1}^{n} \alpha_i \mu \left(A_i \cap E\right) \qquad \left(E \in \mathrm{Msrbl}\left(X\right)\right) . \tag{2.6}$$

We use the convention $0 \cdot \infty = 0$ in case $\alpha_i = 0$ yet $\mu\left(A_i \cap E\right) = \infty$.

Definition 2.41 (The Lebesgue integral of positive functions)**.** Let $f : X \to [0, \infty]$ be measurable. Then

$$\int_E f\mathrm{d}\mu := \sup_s \int_E s\mathrm{d}\mu$$

where the supremum ranges over all simple measurable functions s which obey $0 \leq s \leq f$. Note if f is simple the two definitions coincide, since then the supremum is attained on f itself.

Example 2.42 (The integral against the counting measure)**.** Recall the counting measure from Example 2.30

$$S \overset{c}{\mapsto} |S| \ .$$

What then is

$$\int_S f\mathrm{d}c?$$

We claim that if S is countable then

$$\int_S f\mathrm{d}c = \sum_{x \in S} f\left(x\right) \ .$$

(If S is not countable then the expression $\sum_{x \in S} f\left(x\right)$ requires a bit more definition) To prove this however we will need a limit theorem (see Claim 2.50 below for the proof). We contend ourselves with just the simple function case for now. Let $f = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ be a simple function. Then

$$\int_S f\mathrm{d}c = \sum_{i=1}^{n} \alpha_i c\left(S \cap A_i\right) = \sum_{i=1}^{n} \alpha_i \left|S \cap A_i\right|$$

since, by definition, $A_i \equiv f^{-1}\left(\{\ \alpha_i\ \}\right) \equiv \{\ x \in X \mid f\left(x\right) = \alpha_i\ \}$, we get the result of the claim.

Example 2.43 (The integral against the delta measure)**.** Recall the delta measure δ_{x_0} from Example 2.31. If $f : X \to Y$ is measurable, then what is $\int_S f\mathrm{d}\delta_{x_0}$? We claim

$$\int_S f\mathrm{d}\delta_{x_0} = \chi_S\left(x_0\right) f\left(x_0\right) \ .$$

Indeed, by definition,

$$\begin{aligned}\int_S f\mathrm{d}\delta_{x_0} &= \sup_{s\text{ simple s.t. } 0\leq s\leq f}\int_S s\mathrm{d}\delta_{x_0} \\ &= \sup_{s\text{ simple s.t. } 0\leq s\leq f}\sum_{i=1}^{n}\alpha_i\delta_{x_0}\left(A_i\cap S\right) \\ &= \sup_{s\text{ simple s.t. } 0\leq s\leq f}\sum_{i=1}^{n}\alpha_i\chi_{A_i\cap S}\left(x_0\right) \\ &= \sup_{s\text{ simple s.t. } 0\leq s\leq f}\chi_S\left(x_0\right)s\left(x_0\right) \\ &= \chi_S\left(x_0\right)\sup_{s\text{ simple s.t. } 0\leq s\leq f}s\left(x_0\right) \\ &= \chi_S\left(x_0\right)f\left(x_0\right)\,.\end{aligned}$$

Proposition 2.44. *In the following statements, all functions are assumed to be measurable from a measure space X into $[0,\infty]$ and all sets are elements of $\mathrm{Msrbl}(X)$:*

1. *If $0\leq f\leq g$ then*
$$\int_E f\mathrm{d}\mu\leq\int_E g\mathrm{d}\mu\,. \tag{2.7}$$
2. *If $A\subseteq B$ and $f\geq 0$ then $\int_A f\mathrm{d}\mu\leq\int_B f\mathrm{d}\mu$.*
3. *If $f\geq 0$ and $c\in[0,\infty)$ then*
$$\int_E cf\mathrm{d}\mu = c\int_E f\mathrm{d}\mu\,. \tag{2.8}$$
4. *If $f=0$ for all $x\in E$ then $\int_E f\mathrm{d}\mu=0$. Note this holds even if $\mu(E)=\infty$.*
5. *If $\mu(E)=0$ then*
$$\int_E f\mathrm{d}\mu = 0 \tag{2.9}$$
, even if f takes on the value ∞ on E.
6. *If $f\geq 0$ then*
$$\int_E f\mathrm{d}\mu = \int_X \chi_E f\mathrm{d}\mu\,. \tag{2.10}$$

Proposition 2.45. *Let $s:X\to[0,\infty)$ be a measurable simple function and $\mu:\mathrm{Msrbl}(X)\to[0,\infty]$ be a positive measure. Define $\varphi:\mathrm{Msrbl}(X)\to[0,\infty]$ via*

$$\varphi(E):=\int_E s\mathrm{d}\mu\qquad(E\in\mathrm{Msrbl}(X))\,.$$

Then φ is also a measure on $\mathrm{Msrbl}(X)$.

Proof. Since μ is a measure, $\mu(\varnothing)=0$ so that $\varphi(\varnothing)=0$ via (2.9) and this satisfies the first condition on a measure having at least one measurable set not have infinite-measure.

Next, we verify the countable additivity. Let $\{A_n\}_n\subseteq\mathrm{Msrbl}(X)$ be a sequence of pairwise disjoint sets and write

$$s=\sum_{j=1}^{J}\alpha_j\chi_{B_j}\,.$$

Then

$$\varphi\left(\bigcup_{n=1}^{\infty} A_n\right) = \int_{\bigcup_{n=1}^{\infty} A_n} s\mathrm{d}\mu$$
$$= \sum_{j=1}^{J} \alpha_j \mu\left(B_j \cap \bigcup_{n=1}^{\infty} A_n\right)$$
$$= \sum_{j=1}^{J} \alpha_j \sum_{n=1}^{\infty} \mu\left(B_j \cap A_n\right)$$
$$= \sum_{n=1}^{\infty}\sum_{j=1}^{J} \alpha_j \mu\left(B_j \cap A_n\right)$$
$$= \sum_{n=1}^{\infty} \int_{A_n} s\mathrm{d}\mu$$
$$= \sum_{n=1}^{\infty} \varphi\left(A_n\right) .$$

□

Proposition 2.46. *(Additivity of integral on simple functions) Let* $s, t : X \to [0, \infty)$ *be two measurable simple functions and* $\mu : \mathrm{Msrbl}(X) \to [0, \infty]$ *be a positive measure. Then*

$$\int_X (s+t)\,\mathrm{d}\mu = \int_X s\mathrm{d}\mu + \int_X t\mathrm{d}\mu \,.$$

Proof. Write $s = \alpha_1\chi_{A_1} + \cdots + \alpha_n\chi_{A_n}$ and $t = \beta_1\chi_{B_1} + \cdots + \beta_m\chi_{B_m}$. Then

$$s+t = \alpha_1\chi_{A_1} + \cdots + \alpha_n\chi_{A_n} + \beta_1\chi_{B_1} + \cdots + \beta_m\chi_{B_m}$$

is *not* necessarily of the form Definition 2.24 since there might be intersections between the A_i's and the B_j's, and on those intersections, the value of $s+t$ is $\alpha_i + \beta_j$. Hence let us write

$$s+t = \gamma_1\chi_{C_1} + \cdots + \gamma_l\chi_{C_l}$$

where the γ's and C's correctly account for the intersections. Then we are allowed to write

$$\int_X (s+t)\,\mathrm{d}\mu \equiv \gamma_1\mu\left(C_1\right) + \cdots + \gamma_l\mu\left(C_l\right) .$$

Now, when $\gamma_i = \alpha_{i_r} + \beta_{i_p}$, that means we are on an intersection, in which case we can write that intersection set C_{ij} as $C_{ij} = A_i \cap B_j$ and then

$$A_i = (A_i \setminus B_j) \sqcup C_{ij}$$

and similarly

$$B_j = (B_j \setminus A_i) \sqcup C_{ij}$$

and hence the result. □

2.6 Limit theorems I

The importance of the following result on sequences of *positive* measurable functions cannot overstated.

Theorem 2.47 (Lebesgue's monotone convergence)**.** *Let $f_n : X \to [0,\infty]$ be a sequence of measurable functions such that*

$$f_n(x) \leq f_{n+1}(x) \qquad (x \in X, n \in \mathbb{N}) . \tag{2.11}$$

Assume further that f_n converges pointwise. Then (as we saw in Corollary 2.23) $\lim_n f_n$ is measurable and

$$\lim_n \left(\int_X f_n \mathrm{d}\mu \right) \quad = \quad \int_X \left(\lim_n f_n \right) \mathrm{d}\mu .$$

Proof. The monotonicity (2.11) implies that

$$\int_X f_n \mathrm{d}\mu \quad \leq \quad \int_X f_{n+1} \mathrm{d}\mu \qquad (n \in \mathbb{N})$$

so that $\int_X f_n \mathrm{d}\mu \subseteq [0,\infty]$ is a monotone increasing sequence of *numbers*, and as such necessarily has a (possibly infinite) limit in $[0,\infty]$. Moreover, we also have

$$f_n \leq \lim_{\tilde{n}} f_{\tilde{n}} \qquad (n \in \mathbb{N})$$

and by Corollary 2.23, $\lim_n f_n$ is measurable too, so

$$\int_X f_n \mathrm{d}\mu \leq \int_X \left(\lim_{\tilde{n}} f_{\tilde{n}} \right) \mathrm{d}\mu \qquad (n \in \mathbb{N})$$

and taking the limit of both sides of this w.r.t. n we obtain

$$\lim_n \int_X f_n \mathrm{d}\mu \leq \int_X \left(\lim_n f_n \right) \mathrm{d}\mu .$$

For the other direction, let s be a simple measurable function such that $0 \leq s \leq \lim_n f_n$ and $c \in (0,1)$. In particular, $0 \leq cs < \lim_n f_n$ when $\lim_n f_n > 0$. Then defining

$$E_n \quad := \quad \{ \, x \in X \mid f_n(x) \geq cs(x) \, \} \qquad (n \in \mathbb{N})$$

which are all measurable, and obey $E_n \subseteq E_{n+1}$ by (2.11). We claim that $X = \bigcup_n E_n$. Indeed, let $x \in X$. Then either $\lim_n f_n(x) = 0$ in which case $s(x) = 0$ so that $x \in E_1$. Otherwise, $\lim_n f_n(x) > 0$, so $cs(x) < \lim_n f_n(x)$ and so there must be some n such that $f_n(x) \geq cs(x)$ and for that n, $x \in E_n$.

Finally,

$$\int_X f_n \mathrm{d}\mu \geq \int_{E_n} f_n \mathrm{d}\mu \geq c \int_{E_n} s \mathrm{d}\mu \qquad (n \in \mathbb{N}) .$$

Taking the limit $n \to \infty$ of both sides we obtain

$$\lim_n \int_X f_n \mathrm{d}\mu \quad \geq \quad c \lim_n \int_{E_n} s \mathrm{d}\mu .$$

By the above, we know that s, μ define a new measure $E \mapsto \int_E s \mathrm{d}\mu$ on X, and applying the monotonicity result (2.4) we obtain

$$\lim_n \int_X f_n \mathrm{d}\mu \quad \geq \quad c \int_X s \mathrm{d}\mu .$$

Now take the limit $c \to 1$ here to get

$$\lim_n \int_X f_n \mathrm{d}\mu \geq \int_X s \mathrm{d}\mu .$$

Now take the supremum over simple functions s obeying $0 \leq s \leq \lim_n f_n$ to get

$$\lim_n \int_X f_n \mathrm{d}\mu \geq \int_X \lim_n f_n \mathrm{d}\mu$$

which is what we were trying to show. □

Theorem 2.48 (Exchanging the sum with the integral). *Let $f, g : X \to [0, \infty]$ be measurable. Then*

$$\int_X (f+g)\,\mathrm{d}\mu = \int_X f \mathrm{d}\mu + \int_X g \mathrm{d}\mu \,.$$

Proof. Let s_n, t_n be sequences of positive measurable functions which approximate f, g respectively, according to Theorem 2.27. Since these approximating sequences are monotone, we have by Theorem 2.47 that

$$\lim_n \int_X s_n \mathrm{d}\mu = \int_X f \mathrm{d}\mu \,.$$

Moreover, we know that $s_n + t_n$ is a sequence of positive simple functions which approximates $f + g$ monotonically from below. Hence again via Theorem 2.47

$$\lim_n \int_X (s_n + t_n)\,\mathrm{d}\mu = \int_X (f+g)\,\mathrm{d}\mu \,.$$

But by Proposition 2.46 we know that

$$\int_X (s_n + t_n)\,\mathrm{d}\mu = \int_X s_n \mathrm{d}\mu + \int_X t_n \mathrm{d}\mu$$

since these are simple functions. □

Theorem 2.49 (Exchanging series summation with the integral). *Let $f_n : X \to [0, \infty]$ be a sequence of measurable functions. Then*

$$\int_X \left(\sum_{n=1}^{\infty} f_n \right) \mathrm{d}\mu = \sum_{n=1}^{\infty} \left(\int_X f_n \mathrm{d}\mu \right) \,.$$

Proof. The sequence of partial sums $\sum_{n=1}^{N} f_n$ itself converges monotonically to f from below, so we apply Theorem 2.47 to it, after applying Theorem 2.48 N times on the partial sum. □

Claim 2.50. If $f : X \to [0, \infty]$ and X is countable, and $c : \mathscr{P}(X) \to [0, \infty]$ is the counting measure, then

$$\int_S f \mathrm{d}c = \sum_{x \in S} f(x) \,.$$

Proof. Since X is countable, let $\eta : \mathbb{N} \to X$ be an enumeration of it. Let us then define, for each $n \in \mathbb{N}$, the simple function

$$s_n(x) := \chi_{\{1, \cdots, n\}} \left(\eta^{-1}(x) \right) f(x) = \sum_{j=1}^{n} f(\eta_j) \chi_{\{\eta_j\}}(x) \qquad (x \in X) \,.$$

Then clearly $s_n \to f$ pointwise, and since $f \geq 0$, $s_n \geq 0$. In fact $s_{n+1} \geq s_n$ for all $n \in \mathbb{N}$ so this sequence obeys the conditions of the monotone convergence theorem. Then

$$\begin{aligned}\int_S f\mathrm{d}c &= \lim_n \int_S s_n \mathrm{d}c \\ &= \lim_n \sum_{j=1}^{n} f(\eta_j)\,|S\cap\{\,\eta_j\,\}| \\ &= \sum_{x\in S} f(x)\,.\end{aligned}$$

□

Remark 2.51. If X is *not* countable we can still make sense of this, however, then we need a definition of

$$\sum_{x\in X} f(x)$$

for X uncountable. One such possible definition which is common is

$$\sum_{x\in X} f(x) := \sup_{F\subseteq X:|F|<\infty} \sum_{x\in F} f(x)\,.$$

It turns out that with this definition the integral against the counting measure is precisely $\sum_{x\in X} f(x)$ but we do not pursue this here.

Yet another corollary of Theorem 2.47 is the fact we can exchange double summation on positive double sequences.

Corollary 2.52. *If* $a:\mathbb{N}^2\to[0,\infty]$ *is a double-sequence then*

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{nm} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{nm}\,.$$

Proof. We set up the problem as $\mathbb{N}$ being our measure space with $\mathrm{Msrbl}(\mathbb{N}) := \mathscr{P}(\mathbb{N})$ and we choose c as the counting measure. Then for any $M\in\mathbb{N}$ we define $b_M:\mathbb{N}\to[0,\infty]$ via

$$b_M(n) := \sum_{m=1}^{M} a_{nm} \qquad (n\in\mathbb{N})\,.$$

This is an increasing positive sequence so Theorem 2.47 applies to it:

$$\lim_{M\to\infty}\int_{\mathbb{N}} b_M \mathrm{d}c = \int_{\mathbb{N}} \lim_{M\to\infty} b_M \mathrm{d}c\,. \tag{2.12}$$

By Claim 2.50 the LHS of (2.12) equals

$$\begin{aligned}\lim_{M\to\infty}\sum_{n\in\mathbb{N}} b_M(n) &= \lim_{M\to\infty}\sum_{n\in\mathbb{N}}\sum_{m=1}^{M} a_{nm} \\ &= \lim_{M\to\infty}\sum_{m=1}^{M}\sum_{n\in\mathbb{N}} a_{nm} \\ &= \sum_{m\in\mathbb{N}}\sum_{n\in\mathbb{N}} a_{nm}\,.\end{aligned}$$

On the other hand the RHS of (2.12) yields

$$\begin{aligned}\int_{\mathbb{N}} \lim_{M\to\infty} b_M \mathrm{d}c &= \sum_{n\in\mathbb{N}} \lim_{M\to\infty} b_M(n) \\ &= \sum_{n\in\mathbb{N}} \lim_{M\to\infty} \sum_{m=1}^{M} a_{nm} \\ &= \sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} a_{nm}\,.\end{aligned}$$

□

Lemma 2.53 (Fatou's)**.** *Let $f_n : X \to [0,\infty]$ be a measurable sequence on a measure space $(X, \mathrm{Msrbl}(X), \mu)$. Then*

$$\int_X \left(\liminf_n f_n\right) \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \mathrm{d}\mu\,.$$

Proof. We use the characterization of lim inf given in (2.1). Then

$$\liminf_{n\to\infty} f_n = \lim_{n\to\infty}\left(\inf_{m\geq n} f_m\right) = \sup_{n\in\mathbb{N}} \inf_{m\geq n} f_m\,.$$

Hence, let us define the sequence $g_n := \inf_{m\geq n} f_m$. Then $g_n \leq f_n$ and so

$$\int_X g_n \mathrm{d}\mu \leq \int_X f_n \mathrm{d}\mu \qquad (n\in\mathbb{N})\,. \tag{2.13}$$

Moreover, g_n is an increasing measurable sequence whose limit is $\liminf_{n\to\infty} f_n$, So applying Theorem 2.47 to this sequence we find

$$\begin{aligned}\lim_{n\to\infty}\int_X g_n \mathrm{d}\mu &= \int_X \lim_{n\to\infty} g_n \mathrm{d}\mu \\ &\updownarrow \\ \lim_{n\to\infty}\int_X g_n \mathrm{d}\mu &= \int_X \left(\liminf_{n\to\infty} f_n\right)\mathrm{d}\mu\,.\end{aligned}$$

Hence taking the lim inf on (2.13) we find

$$\begin{aligned}\liminf_n \int_X f_n \mathrm{d}\mu &\geq \liminf_n \int_X g_n \mathrm{d}\mu \\ &= \lim_n \int_X g_n \mathrm{d}\mu \\ &= \int_X \left(\liminf_{n\to\infty} f_n\right)\mathrm{d}\mu\end{aligned}$$

which is what we were trying to show. □

With the monotone convergence theorem we can also generalize Proposition 2.45 from simple functions to general measurable functions.

Theorem 2.54. *Let $f : X \to [0,\infty]$ be measurable and define $\varphi : \mathrm{Msrbl}(X) \to [0,\infty]$ via*

$$\varphi(E) := \int_E f \mathrm{d}\mu \qquad (E \in \mathrm{Msrbl}(X))\,.$$

Then φ is a positive measure on $\mathrm{Msrbl}(X)$ and

$$\int_X g \mathrm{d}\varphi = \int_X g f \mathrm{d}\mu \qquad (g : X \to [0,\infty] \text{ measurable})\,.$$

Proof. Since μ is a measure, $\mu(\varnothing) = 0$ so that $\varphi(\varnothing) = 0$ via (2.9). Next, we want to verify countable additivity of φ. To that end, let $\{ A_i \}_{i\in\mathbb{N}}$ be a sequence of pairwise disjoint measurable sets. Our goal is to show that $\varphi\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \varphi(A_i)$. Then using (2.10) we have

$$\begin{aligned}\varphi\left(\bigcup_{i\in\mathbb{N}} A_i\right) &= \int_{\bigcup_{i\in\mathbb{N}} A_i} f \mathrm{d}\mu \\ &= \int_X \left(\chi_{\bigcup_{i\in\mathbb{N}} A_i} f\right) \mathrm{d}\mu \\ &= \int_X \left(\sum_{i\in\mathbb{N}} \chi_{A_i} f\right) \mathrm{d}\mu\end{aligned}$$

Now using Theorem 2.49 we get

$$\begin{aligned}\varphi\left(\bigcup_{i\in\mathbb{N}} A_i\right) &= \sum_{i\in\mathbb{N}} \int_X \chi_{A_i} f \mathrm{d}\mu \\ &= \sum_{i\in\mathbb{N}} \int_{A_i} f \mathrm{d}\mu \\ &= \sum_{i\in\mathbb{N}} \varphi(A_i)\end{aligned}$$

which is what we wanted to prove. □

2.7 Integrating complex-valued functions

Here again

$$(X, \mathrm{Msrbl}(X), \mu)$$

is a measure space. As we mentioned above in the beginning of Section 2.3, we shall write a so-called *polarization identity*. For any $f : X \to \mathbb{C}$, we may decompose it as the complex linear combination of four non-negative functions as

$$f = \mathbb{Re}\{f\}^+ - \mathbb{Re}\{f\}^- + \mathrm{i}\,\mathbb{Im}\{f\}^+ - \mathrm{i}\,\mathbb{Im}\{f\}^- . \tag{2.14}$$

However, as it turns out, we don't want to *just* define

$$\int_E f \mathrm{d}\mu := \int_E \mathbb{Re}\{f\}^+ \mathrm{d}\mu - \int_E \mathbb{Re}\{f\}^- \mathrm{d}\mu + \mathrm{i} \int_E \mathbb{Im}\{f\}^+ \mathrm{d}\mu - \mathrm{i} \int_E \mathbb{Im}\{f\}^- \mathrm{d}\mu \tag{2.15}$$

because that might cause some weird algebraic cancelations of the form $\infty - \mathrm{i}\infty$. For that reason, we prefer to first define

Definition 2.55 ($L^1(X,\mu)$ space)**.** Recall that if $f : X \to \mathbb{C}$ is measurable, then so is $|f| : X \to [0,\infty)$ by Corollary 2.23. It is then legitimate to consider the integral

$$\int_X |f| \mathrm{d}\mu$$

and using it we define the space

$$L^1(X,\mu) := \left\{ f : X \to \mathbb{C} \,\middle|\, f \text{ is msrbl. and } \int_X |f| \mathrm{d}\mu < \infty \right\} .$$

We then only define (2.15) only for $f \in L^1(X,\mu)$. We note that since

$$\mathbb{Re}\{f\}^+ \leq |f|$$

etc we have

$$\int_X f \mathrm{d}\mu \leq 4 \int_X |f| \mathrm{d}\mu$$

but actually we will quickly get rid of the factor 4 in Theorem 2.60 below.

Theorem 2.56. *$L^1(X,\mu)$ is a $\mathbb{C}$-vector space.*

Proof. We want to show that if $f,g\in L^1(X,\mu)$ and $\alpha\in\mathbb{C}$ then $\alpha f+g\in L^1(X,\mu)$ too. First, we know that $\alpha f+g$ is measurable by Theorem 2.20. Moreover,

$$|\alpha f+g|\leq|\alpha|\,|f|+|g|$$

so that

$$\int_X|\alpha f+g|\leq|\alpha|\int_X|f|\,\mathrm{d}\mu+\int_X|g|\,\mathrm{d}\mu<\infty\,.$$

□

Theorem 2.57. *The map*

$$\int_X\cdot\mathrm{d}\mu:L^1(X,\mu)\to\mathbb{C}$$

is itself $\mathbb{C}$-linear, so that the integral is a linear functional on the $\mathbb{C}$-vector space $L^1(X,\mu)$.

Proof. Let $f,g\in L^1(X,\mu)$ and $\alpha\in\mathbb{C}$. We want to show that

$$\int_X(\alpha f+g)\,\mathrm{d}\mu=\alpha\int_X f\mathrm{d}\mu+\int_X g\mathrm{d}\mu\,. \tag{2.16}$$

To that end, let $u,v\in L^1(X,\mu)$ be two *real-valued* functions. Set $h:=u+v$ and note that the decomposition into the positive and negative parts obeys

$$\begin{aligned}h^+-h^-&=u^+-u^-+v^+-v^-\\&\updownarrow\\h^++u^-+v^-&=u^++v^++h^-\,.\end{aligned}$$

Each side of this latter equation is non-negative, and so obeys additivity as stipulated by Theorem 2.48, i.e.,

$$\int_X h^+\mathrm{d}\mu+\int_X u^-\mathrm{d}\mu+\int_X v^-\mathrm{d}\mu=\int_X u^+\mathrm{d}\mu+\int_X v^+\mathrm{d}\mu+\int_X h^-\mathrm{d}\mu\,.$$

By the $u,v\in L^1$, each of these integrals is finite, so we may move sides again to get

$$\int_X h^+\mathrm{d}\mu-\int_X h^-\mathrm{d}\mu=\int_X u^+\mathrm{d}\mu-\int_X u^-\mathrm{d}\mu+\int_X v^+\mathrm{d}\mu-\int_X v^-\mathrm{d}\mu\,.$$

Now *by definition* in (2.15) we have

$$\int_X h\mathrm{d}\mu=\int_X h^+\mathrm{d}\mu-\int_X h^-\mathrm{d}\mu$$

and similarly for u,v, so we get

$$\int_X(u+v)\,\mathrm{d}\mu=\int_X u\mathrm{d}\mu+\int_X v\mathrm{d}\mu$$

which is additivity for *real-valued* functions. Also note that if $u=u^+-u^-$ is the decomposition into positive and negative parts of $u:X\to\mathbb{R}$, then $-u=-u^++u^-$ so the positive and negative parts switch and so

$$\begin{aligned}\int_X(-u)\,\mathrm{d}\mu&\equiv\int_X u^-\mathrm{d}\mu-\int_X u^+\mathrm{d}\mu\\&=-\left(\int_X u^+\mathrm{d}\mu-\int_X u^-\mathrm{d}\mu\right)\\&=-\int_X u\mathrm{d}\mu\end{aligned}$$

so that still with $u, v : X \to \mathbb{R}$

$$\int_X (u - v)\,\mathrm{d}\mu = \int_X u\mathrm{d}\mu + \int_X (-v)\,\mathrm{d}\mu = \int_X u\mathrm{d}\mu - \int_X v\mathrm{d}\mu$$

and

$$\begin{aligned}\int_X (u + \mathrm{i}v)\,\mathrm{d}\mu &\equiv \int_X u^+\mathrm{d}\mu - \int_X u^-\mathrm{d}\mu + \mathrm{i}\int_X v^+\mathrm{d}\mu - \mathrm{i}\int_X v^-\mathrm{d}\mu \\ &= \int_E u\mathrm{d}\mu + \mathrm{i}\int_X v\mathrm{d}\mu\,.\end{aligned}$$

Finally, we learn then that if $f, g : X \to \mathbb{C}$ then

$$\begin{aligned}\int_X (f+g)\,\mathrm{d}\mu &= \int_X (\mathbb{Re}\{f+g\} + \mathrm{i}\,\mathbb{Im}\{f+g\})\,\mathrm{d}\mu \\ &= \int_X \mathbb{Re}\{f+g\}\,\mathrm{d}\mu + \mathrm{i}\int_X \mathbb{Im}\{f+g\}\,\mathrm{d}\mu \\ &= \int_X (\mathbb{Re}\{f\} + \mathbb{Re}\{g\})\,\mathrm{d}\mu + \mathrm{i}\int_X (\mathbb{Im}\{f\} + \mathbb{Im}\{g\})\,\mathrm{d}\mu \\ &= \int_X \mathbb{Re}\{f\}\,\mathrm{d}\mu + \int_X \mathbb{Re}\{g\}\,\mathrm{d}\mu + \mathrm{i}\int_X \mathbb{Im}\{f\}\,\mathrm{d}\mu + \mathrm{i}\int_X \mathbb{Im}\{g\}\,\mathrm{d}\mu \\ &= \int_X f\mathrm{d}\mu + \int_X g\mathrm{d}\mu\,.\end{aligned}$$

Now we want to show that $\int \alpha f \mathrm{d}\mu = \alpha \int_X f\mathrm{d}\mu$ for any $\alpha \in \mathbb{C}$ and $f \in L^1$. To that end, we already know from (2.8) that if $\alpha \geq 0$ then

$$\begin{aligned}\int_X \alpha f\mathrm{d}\mu &\equiv \int_E \mathbb{Re}\{\alpha f\}^+\,\mathrm{d}\mu - \int_E \mathbb{Re}\{\alpha f\}^-\,\mathrm{d}\mu + \mathrm{i}\int_E \mathbb{Im}\{\alpha f\}^+\,\mathrm{d}\mu - \mathrm{i}\int_E \mathbb{Im}\{\alpha f\}^-\,\mathrm{d}\mu \\ &= \int_E \alpha\,\mathbb{Re}\{f\}^+\,\mathrm{d}\mu - \int_E \alpha\,\mathbb{Re}\{f\}^-\,\mathrm{d}\mu + \mathrm{i}\int_E \alpha\,\mathbb{Im}\{f\}^+\,\mathrm{d}\mu - \mathrm{i}\int_E \alpha\,\mathbb{Im}\{f\}^-\,\mathrm{d}\mu \\ &= \alpha\int_E \mathbb{Re}\{f\}^+\,\mathrm{d}\mu - \alpha\int_E \mathbb{Re}\{f\}^-\,\mathrm{d}\mu + \mathrm{i}\alpha\int_E \mathbb{Im}\{f\}^+\,\mathrm{d}\mu - \mathrm{i}\alpha\int_E \mathbb{Im}\{f\}^-\,\mathrm{d}\mu \\ &= \alpha\int_X f\mathrm{d}\mu\,.\end{aligned}$$

Clearly if $\alpha = -1$ or $\alpha = \mathrm{i}$ then this just rearranges the quadruplet $\mathbb{Re}\{f\}^+, \mathbb{Re}\{f\}^-, \mathbb{Im}\{f\}^+, \mathbb{Im}\{f\}^-$. ☐

Corollary 2.58. *We may exchange real and imaginary parts with integration.*

Claim 2.59. If $u, v : X \to \mathbb{R}$ are L^1 and $u \leq v$ then $\int_X u\mathrm{d}\mu \leq \int_X v\mathrm{d}\mu$.

Proof. We have

$$\begin{aligned} u^+ - u^- &\leq v^+ - v^- \\ &\updownarrow \\ u^+ + v^- &\leq v^+ + u^-\,. \end{aligned}$$

Now invoking (2.7) we get

$$\int_X \left(u^+ + v^-\right)\mathrm{d}\mu \leq \int_X \left(v^+ + u^-\right)\mathrm{d}\mu\,.$$

Using Theorem 2.57 and re-arranging we obtain the result. ☐

Theorem 2.60 (The triangle inequality)**.** *For all $f \in L^1(\mu)$ we have*

$$\left|\int_X f\mathrm{d}\mu\right| \leq \int_X |f|\,\mathrm{d}\mu\,.$$

Proof. If $\int_X f\mathrm{d}\mu = 0$ we are finished. Otherwise,

$$\begin{aligned} 0 &< \left|\int_X f\mathrm{d}\mu\right| \\ &= \frac{\int_X f\mathrm{d}\mu}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)} \\ &= \int_X \left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right] f\mathrm{d}\mu \end{aligned}$$

where in the last line we simply inserted the scalar into the integral thanks to linearity Theorem 2.57. Taking the real part of the equation

$$\left|\int_X f\mathrm{d}\mu\right| = \int_X \left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right] f\mathrm{d}\mu$$

yields

$$\left|\int_X f\mathrm{d}\mu\right| = \mathbb{Re}\left\{\int_X \left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right] f\mathrm{d}\mu\right\} = \int_X \mathbb{Re}\left\{\left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right] f\right\}\mathrm{d}\mu\,.$$

Next, thanks to Claim 2.59 and $u \leq |u|$ for $u \in L^1$ which is real valued, we have

$$\left|\int_X f\mathrm{d}\mu\right| \leq \int_X \left|\mathbb{Re}\left\{\left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right] f\right\}\right|\mathrm{d}\mu \leq \int_X \left|\left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right] f\right|\mathrm{d}\mu$$

where in the last inequality we used $|\mathbb{Re}\{z\}| \leq |z|$ and then Claim 2.59 once more. But we note that $\left|\left[\frac{1}{\left(\int_X f\mathrm{d}\mu/\left|\int_X f\mathrm{d}\mu\right|\right)}\right]\right| = 1$ so we obtain

$$\left|\int_X f\mathrm{d}\mu\right| \leq \int_X |f|\,\mathrm{d}\mu$$

which is what we were trying to show. ☐

2.8 Limit theorems II

We come to a basic result in Lebesgue integration, one of the most powerful limit theorems.

Theorem 2.61 (Lebesgue dominated convergence)**.** *Let $f_n : X \to \mathbb{C}$ be a sequence of measurable functions which converges pointwise on X. Assume further there is some $g \in L^1(\mu)$ which dominates the entire sequence:*

$$|f_n(x)| \leq g(x) \qquad (x \in X, n \in \mathbb{N})\,. \tag{2.17}$$

(note this inequality automatically implies $\operatorname{im}(g) \subseteq [0,\infty)$*)*
Then $\lim_n f_n \in L^1(\mu)$*,*

$$\lim_n \int_X \left|f_n - \lim_{n'} f_{n'}\right|\mathrm{d}\mu = 0$$

and

$$\lim_n \int_X f_n\mathrm{d}\mu = \int_X \lim_n f_n\mathrm{d}\mu\,.$$

Proof. Recall from Theorem 2.22 that $\lim_n f_n$ is measurable. Taking the limit on (2.17) we obtain $|\lim_n f_n| \leq g$ so that $\lim_n f_n \in L^1(\mu)$ indeed. Moreover, by the triangle inequality we have

$$\left|f_n - \lim_{n'} f_{n'}\right| \leq |f_n| + \left|\lim_{n'} f_{n'}\right| \leq g + g = 2g$$

so that $2g - |f_n - \lim_{n'} f_{n'}| \geq 0$ and hence Fatou's lemma Lemma 2.53 applies to the sequence $\{ 2g - |f_n - \lim_{n'} f_{n'}| \}_n$. It implies

$$\begin{aligned}
\int_X \liminf_n \left[2g - \left|f_n - \lim_{n'} f_{n'}\right|\right] \mathrm{d}\mu &\leq \liminf_n \int_X \left[2g - \left|f_n - \lim_{n'} f_{n'}\right|\right] \mathrm{d}\mu \\
&\Updownarrow \\
\int_X 2g \mathrm{d}\mu &\leq \liminf_n \int_X 2g\mathrm{d}\mu - \limsup_n \int_X \left|f_n - \lim_{n'} f_{n'}\right| \mathrm{d}\mu \\
&\Updownarrow \\
\limsup_n \int_X \left|f_n - \lim_{n'} f_{n'}\right| \mathrm{d}\mu &\leq 0 \\
&\Downarrow \\
\lim_n \int_X \left|f_n - \lim_{n'} f_{n'}\right| \mathrm{d}\mu &= 0\,.
\end{aligned}$$

Next, we have by Theorem 2.60 that

$$\lim_n \int_X \left(f_n - \lim_{n'} f_{n'}\right) \mathrm{d}\mu = 0\,.$$

□

Corollary 2.62 (The bounded convergence theorem)**.** *Let $(X, \mathrm{Msrbl}(X), \mu)$ be a measure space such that $\mu(X) < \infty$ and assume that $f_n : X \to \mathbb{C}$ is a sequence of measurable functions which converges pointwise and such that*

$$\sup_n \|f_n\|_\infty < \infty\,.$$

Then

$$\lim_n \int_X f_n \mathrm{d}\mu = \int_X \lim_n f_n \mathrm{d}\mu\,.$$

Proof. Let $g : X \to \mathbb{C}$ be given by

$$g(x) := \sup_n \|f_n\|_\infty \qquad (x \in X)\,.$$

Then as a constant function g is measurable and it dominates the sequence. Moreover, since $\mu(X) < \infty$, $g \in L^1(\mu)$:

$$\int_X |g| \mathrm{d}\mu = \int_X g\mathrm{d}\mu = \left(\sup_n \|f_n\|_\infty\right) \mu(X) < \infty\,.$$

Hence, Theorem 2.61 implies the result. □

2.9 Construction of non-trivial measures [Folland]

In our journey so far we have encountered only two measures: the counting measure and the Dirac delta measure. To get more interesting measures we outline a construction whereby we define not a measure, but an outer measure, on the entire power set $\mathscr{P}(X)$ and then restrict the domain to get an honest measure; see Figure 2. The definition of the outer measure is a bit easier and follows geometric intuition.

Remark 2.63. Outer measures are NOT measures according to Definition 2.28.

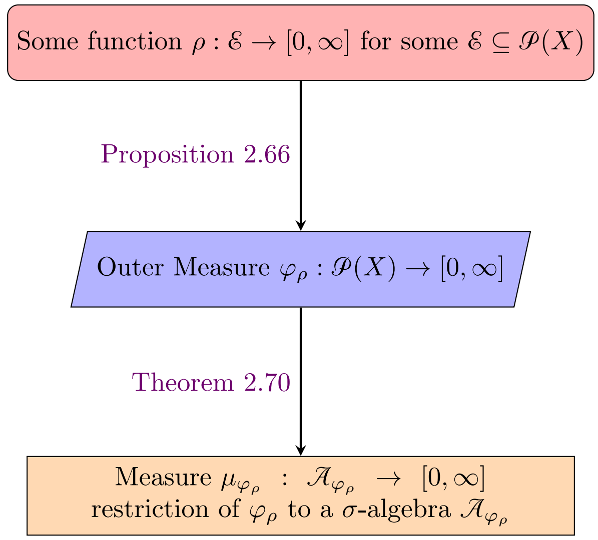

Figure 2: The process of constructing measures. If the initial input to this process ρ happens to be a premeasure defined on an algebra, then Theorem 2.76 guarantees further properties to μ_{φ_ρ}.

2.9.1 Outer measures

An outer measure is a map defined on *more* sets than just measurable sets, in fact, it is defined on the entire power set, but it is required to obey *less* axioms than an actual measure.

Definition 2.64 (Outer measure)**.** Let X be some non-empty set (we don't need to choose $\text{Msrbl}(X)$ on it yet). An outer measure φ on it is a map

$$\varphi : \mathcal{P}(X) \to [0,\infty]$$

such that

1. (*zero on empty set*) $\varphi(\varnothing) = 0$.
2. (*monotonicity*) $\varphi(A) \leq \varphi(B)$ if $A \subseteq B$.
3. (*countable sub-additivity*) $\varphi\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \varphi(A_n)$ for all sequences $A_n \in \mathcal{P}(X)$.

We see that φ obeys *less* than a measure: it is merely countably sub-additive rather than countably additive. But it is defined on the entire $\mathcal{P}(X)$.

We could define outer measures directly, for example

Example 2.65. An outer measure on $\mathbb{R}$ is given by

$$\varphi(A) \quad := \quad \begin{cases} 0 & A = \varnothing \\ 1 & A \neq \varnothing \end{cases} \qquad (A \in \mathcal{P}(\mathbb{R}))\ .$$

One easily verifies the axioms.

A more useful way for us will be to get outer measures out of more primitive functions:

Proposition 2.66. *Let $\mathcal{E} \subseteq \mathcal{P}(X)$ such that $\varnothing, X \in \mathcal{E}$ and $\rho : \mathcal{E} \to [0,\infty]$ be given such that $\rho(\varnothing) = 0$. Define $\varphi_\rho : \mathcal{P}(X) \to [0,\infty]$ via*

$$\varphi_\rho(A) := \inf\left(\left\{ \sum_{n=1}^{\infty} \rho(E_n) \,\middle|\, \{ E_n \}_n \subseteq \mathcal{E} \wedge A \subseteq \bigcup_n E_n \right\}\right) \qquad (A \in \mathcal{P}(X))\ .$$

Then φ_ρ is an outer measure.

Proof. First we want to show that $\varphi_\rho(\varnothing) = 0$. Since $\varnothing \in \mathcal{E}$, we may take the cover $E_n = \varnothing$ for all n. Since $\rho(\varnothing) = 0$, we get that we are taking an infimum over a set of positive numbers which contains zero, and hence the infimum equals zero, so $\varphi_\rho(\varnothing) = 0$.

Next, if $A \subseteq B$ are two subsets of X, then every cover of B is also a cover of A, so necessarily

$$\left\{ \{ E_n \}_n \subseteq \mathcal{E} \;\middle|\; B \subseteq \bigcup_n E_n \right\} \subseteq \left\{ \{ E_n \}_n \subseteq \mathcal{E} \;\middle|\; A \subseteq \bigcup_n E_n \right\}$$

so the infimum over the bigger set will be smaller, and hence $\varphi_\rho(A) \leq \varphi_\rho(B)$ as desired.

Finally, we need to establish countable sub-additivity. Let $\{ A_n \}_n \subseteq \mathcal{P}(X)$ be some sequence, and choose $\varepsilon > 0$ and another sequence $\{ \varepsilon_n \}_n \subseteq (0, \infty)$ such that $\sum_n \varepsilon_n = \varepsilon$. By the approximation property for the infimum, we have for every n,

$$\varphi_\rho(A_n) > \sum_{m=1}^{\infty} \rho(E_{nm}^{\varepsilon_n}) - \varepsilon_n$$

for some sequence $\{ E_{nm}^{\varepsilon_n} \}_m \subseteq \mathcal{E}$ which covers A_n. Moreover, since each such sequence covers A_n for fixed n, taking the union of *all* sequences covers the union of all A_n's. I.e.,

$$\bigcup_n A_n \subseteq \bigcup_n \bigcup_m E_{nm}^{\varepsilon_n} .$$

Since we have established monotonicity of φ_ρ, we invoke it now on this last inclusion to obtain

$$\varphi_\rho\left(\bigcup_n A_n\right) \leq \varphi_\rho\left(\bigcup_n \bigcup_m E_{nm}^{\varepsilon_n}\right) .$$

But now, we can explicitly estimate the term on the right hand side since $E_{nm}^{\varepsilon_n}$ covers itself so the infimum is attained on itself and we obtain

$$\varphi_\rho\left(\bigcup_n A_n\right) \leq \varphi_\rho\left(\bigcup_n \bigcup_m E_{nm}^{\varepsilon_n}\right) \leq \sum_{n,m} \rho(E_{nm}^{\varepsilon_n}) = \sum_n \sum_m \rho(E_{nm}^{\varepsilon_n}) < \sum_n [\varphi_\rho(A_n) + \varepsilon_n] = \varepsilon + \sum_n \varphi_\rho(A_n) .$$

Since $\varepsilon > 0$ was arbitrary we obtain the result. □

Example 2.67. An example that will be actually the raison d'être of this entire construction is to take $X = \mathbb{R}$,

$$\mathcal{E} := \{ [a, b) \mid a < b \in \mathbb{R} \}$$

and

$$\rho([a, b)) := b - a .$$

There are still some pitfalls with this construction.

Claim 2.68. If ρ is not countably additive, φ_ρ could fail to coincide with ρ when restricted to $\mathcal{E}$.

Proof. Consider $X = \mathbb{N}$, $\mathcal{E} := \{ A \subseteq \mathbb{N} \mid |A| < \infty \vee |A^c| < \infty \}$ and define

$$\rho(A) \quad := \quad \begin{cases} 1 & |A^c| < \infty \\ 0 & |A| < \infty \end{cases} .$$

Then one may verify that $\varphi_\rho = 0$ always, and so, does not agree with ρ when restricted to $\mathcal{E}$: (*Chayim Lowen*) It is clear that $\varnothing, X \in \mathcal{E}$ and that $\rho(\varnothing) = 0$, so φ_ρ is well-defined. Note that since every set forms a cover of itself, we have $\varphi_\rho(S) \leq \rho(S)$ for all $S \subseteq \mathcal{E}$.[a] Since φ_ρ is an outer measure, it is countably subadditive. Hence

$$\varphi_\rho(\mathbb{N}) = \varphi_\rho\left(\bigcup_{i=1}^{\infty} \{i\}\right) \leq \sum_{i=1}^{\infty} \varphi_\rho(\{i\}) \leq \sum_{i=1}^{\infty} \rho(\{i\}) = \sum_{i=1}^{\infty} 0 = 0$$

Thus $\varphi_\rho(X) = 0 \neq 1 = \rho(X)$[b]. □

[a]This holds in complete generality.
[b]Since φ_ρ is an outer measure, it will follow that $\varphi_\rho = 0$.

2.9.2 Constructing measures out of outer measures

Now that we have some idea of what an outer measure would be, we want a systematic process to get from an outer measure φ to a measure μ_φ. To do so, we must restrict the domain of the resulting measure μ_φ. Indeed, it is not realistic that the measure we shoot for will have $\mathscr{P}(X)$ as its domain since we know that eventually some sets will need to be *non* measurable. It turns out that the correct criterion for this is as follows

Definition 2.69 (Measurable sets w.r.t. an outer measure)**.** Let X be a non-empty set and $\varphi : \mathscr{P}(X) \to [0,\infty]$ be some outer measure on it. Let

$$\mathscr{A}_\varphi := \{ A \in \mathscr{P}(X) \mid \forall Q \in \mathscr{P}(X), \varphi(Q) = \varphi(Q \cap A) + \varphi(Q \cap (X \setminus A)) \} .$$

We call the elements of $\mathscr{A}_\varphi$ the φ-measurable subsets of X.

Note that since $Q = (Q \cap A) \cup (Q \cap (X \setminus A))$, by subadditivity we *always* have

$$\varphi(Q) \leq \varphi(Q \cap A) + \varphi(Q \cap (X \setminus A))$$

so one could just as well define

$$\mathscr{A}_\varphi := \{ A \in \mathscr{P}(X) \mid \forall Q \in \mathscr{P}(X) : \varphi(Q) < \infty, \varphi(Q) \geq \varphi(Q \cap A) + \varphi(Q \cap (X \setminus A)) \} .$$

Theorem 2.70 (Carathéodory's restriction theorem)**.** *$\mathscr{A}_\varphi$ is a σ-algebra on X and $\mu_\varphi : \mathscr{A}_\varphi \to [0,\infty]$ defined via*

$$A \mapsto \varphi(A)$$

is a measure on X.

Proof. Following Definition 2.1, we show that $X \in \mathscr{A}_\varphi$. If $\varphi(Q) < \infty$, then we want to show that

$$\begin{aligned} \varphi(Q) &\geq \varphi(Q \cap X) + \varphi(Q \cap (X \setminus X)) \\ &= \varphi(Q) + \varphi(Q \cap \varnothing) \\ &= \varphi(Q) + \varphi(\varnothing) = \varphi(Q) \end{aligned}$$

which is true, so $X \in \mathscr{A}_\varphi$.

Next, we want to show closure under complements. Let $A \in \mathscr{A}_\varphi$. Then we want to show if $\varphi(Q) < \infty$,

$$\begin{aligned} \varphi(Q) &\geq \varphi(Q \cap (X \setminus A)) + \varphi(Q \cap (X \setminus (X \setminus A))) \\ &= \varphi(Q \cap (X \setminus A)) + \varphi(Q \cap A) \end{aligned}$$

which is true since $A \in \mathscr{A}_\varphi$.

Finally, we want to show closure under countable unions. Let us first show closure under finite unions. So let $A, B \in \mathscr{A}_\varphi$ and $Q \in \mathscr{P}(X)$ with $\varphi(Q) < \infty$. Then invoking $A \in \mathscr{A}_\varphi$ we get

$$\varphi(Q) \geq \varphi(Q \cap A) + \varphi(Q \cap A^c)$$

and invoking now $B \in \mathscr{A}_\varphi$ on each term on the RHS yields

$$\varphi(Q) \geq \varphi(Q \cap A \cap B) + \varphi(Q \cap A \cap B^c) + \varphi(Q \cap A^c \cap B) + \varphi(Q \cap A^c \cap B^c) .$$

Note that $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, so since φ is subadditive, we get

$$\varphi(Q \cap (A \cup B)) \leq \varphi(Q \cap A \cap B) + \varphi(Q \cap A \cap B^c) + \varphi(Q \cap A^c \cap B)$$

and so all together

$$\varphi(Q) \geq \varphi(Q \cap (A \cup B)) + \varphi(Q \cap A^c \cap B^c) \; .$$

Finally, observe that $A^c \cap B^c = (A \cup B)^c$ which leads to the closure under finite unions we seek.

Now we want to go to closure under countable unions. Note that it will suffice to show this for pairwise disjoint sequences via the construction

$$B_n := A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k \right)$$

given a sequence $\{ A_n \}_n$. So assume WLOG that $\{ A_n \}_n \subseteq \mathscr{A}_\varphi$ is pairwise disjoint. Our goal is to show that $\bigcup_n A_n \in \mathscr{A}_\varphi$. Let $Q \in \mathscr{P}(X)$ with $\varphi(Q) < \infty$. Since we know that finitely many unions are in $\mathscr{A}_\varphi$, for any N we have

$$\varphi(Q) \geq \varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right) \right) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right)^c \right) \; .$$

Let us invoke now $A_N \in \mathscr{A}_\varphi$ on the set $Q \cap \left(\bigcup_{n=1}^{N} A_n \right)$ to get

$$\begin{aligned} \varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right) \right) &\geq \varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right) \cap A_N \right) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right) \cap A_N^c \right) \\ &= \varphi(Q \cap A_N) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{N-1} A_n \right) \right) . \end{aligned}$$

Performing now induction on N shows that

$$\varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right) \right) \geq \sum_{n=1}^{N} \varphi(Q \cap A_n) \; .$$

Hence we find

$$\varphi(Q) \geq \sum_{n=1}^{N} \varphi(Q \cap A_n) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{N} A_n \right)^c \right) .$$

Note that $\bigcup_{n=1}^{N} A_n \subseteq \bigcup_{n=1}^{\infty} A_n$ so $\left(\bigcup_{n=1}^{N} A_n \right)^c \supseteq \left(\bigcup_{n=1}^{\infty} A_n \right)^c$ and hence by monotonicity of φ we get

$$\varphi(Q) \geq \sum_{n=1}^{N} \varphi(Q \cap A_n) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{\infty} A_n \right)^c \right) .$$

Take now the limit $N \to \infty$ to get

$$\varphi(Q) \geq \sum_{n=1}^{\infty} \varphi(Q \cap A_n) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{\infty} A_n \right)^c \right) .$$

By countable subaddivitiy of φ, we have $\sum_{n=1}^{\infty} \varphi(Q \cap A_n) \geq \varphi\left(\bigcup_n (Q \cap A_n)\right)$ so

$$\begin{aligned} \varphi(Q) &\geq \varphi\left(\bigcup_n (Q \cap A_n) \right) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{\infty} A_n \right)^c \right) \\ &= \varphi\left(Q \cap \bigcup_n A_n \right) + \varphi\left(Q \cap \left(\bigcup_{n=1}^{\infty} A_n \right)^c \right) \end{aligned} \tag{2.18}$$

which is what we were trying to show.

Next, we want to show that μ_φ is a measure. To that end, we may take $\varnothing \in \mathscr{A}_\varphi$ as that set for which

$$\mu_\varphi(\varnothing) = 0 < \infty \, .$$

So we are left with showing countable additivity on pairwise disjoint sets. First let us show additivity. Let $A,B\in\mathcal{A}_\varphi$ with $A\cap B=\varnothing$. We already know that $A\cup B\in A_\varphi$, and so

$$\begin{aligned}\mu_\varphi(A\cup B) &\equiv \varphi(A\cup B)\\ &= \varphi((A\cup B)\cap A)+\varphi((A\cup B)\cap A^c)\\ &= \varphi(A)+\varphi(B)\ .\end{aligned}$$

To get countable additivity, we invoke the above demonstration that $\mathcal{A}_\varphi$ is closed under countable unions, in particular, Section 2.9.2, with $Q=\bigcup_n A_n$. This yields

$$\varphi\left(\bigcup_n A_n\right)\geq\sum_{n=1}^\infty\varphi(A_n)$$

and since the other direction of the inequality is true by definition, we get countable additivity on pairwise disjoint sets which belong to $\mathcal{A}_\varphi$, and hence, of μ_φ. □

Claim 2.71. μ_φ as constructed above is complete, in the sense that if $A\in\mathcal{A}_\varphi$ has $\mu_\varphi(A)=0$ and $B\subseteq A$ then $B\in\mathcal{A}_\varphi$ too.

Proof. Let $Q\in\mathcal{P}(X)$ such that $\varphi(Q)<\infty$. Then we want to show that

$$\varphi(Q)\geq\varphi(Q\cap B)+\varphi(Q\cap B^c)\ .$$

Note that $Q\cap B\subseteq Q\cap A\subseteq A$ so $0\leq\varphi(Q\cap B)\leq\varphi(A)=0$. So we only have to show

$$\varphi(Q)\quad\geq\quad\varphi(Q\cap B^c)\ .$$

But this is of course true since $Q\cap B^c\subseteq Q$ and φ is monotone. □

Again there are issues with this construction

$$\rho\to\varphi_\rho\to\mu_{\varphi_\rho}\ .$$

Claim 2.72. There are choices of ρ such that $\mathcal{A}_{\varphi_\rho}$ *does not contain* the σ-algebra generated by $\mathcal{E}$.

Proof. The example presented in the proof of Claim 2.68 will not do, because for that, actually $\sigma(\mathcal{E})=\mathcal{P}(\mathbb{N})=\mathcal{A}_{\varphi_\rho}$. For an actual counter example, consider $X=[0,1]$ with $\mathcal{E}:=\{\,\varnothing,[0,1]\,\}\cup\{\,[0,a)\mid a\in(0,1)\,\}$ and define

$$\rho(\varnothing)\quad:=\quad 0$$

as well as

$$\rho([0,1])\quad:=\quad 1$$

and

$$\rho([0,a))\quad:=\quad 0\qquad(a\in(0,1))\ .$$

One verifies that the resulting outer measure is given by

$$\varphi_\rho(E)\quad:=\quad\begin{cases}0 & \sup(E)<1\\ 1 & \sup(E)=1\end{cases}\ .$$

Moreover, a set $A\subseteq[0,1]$ is φ_ρ-measurable iff $\sup(A)<1$ or $\sup(A^c)<1$. BUT, $\sigma(\mathcal{E})=\mathcal{B}([0,1])$, so that it is *not* true that

$$\mathcal{A}_{\varphi_\rho}\quad\supseteq\quad\sigma(\mathcal{E})\ .$$

□

2.9.3 Constructing outer measures out of premeasures

In principle we are now already prepared to define a new measure out of a given outer measure. For example, this construction applied on Example 2.67 yields the Lebesgue measure. The problem is that stopping now would lead to the problems outlined in Claim 2.68 and Claim 2.72. We need a somewhat more systematic construction to get outer measures compared with Proposition 2.66 which will guarantee all the properties we want. For that reason, we turn our attention to

Definition 2.73 (Premeasures)**.** Let $\mathcal{A}$ be an algebra (the definition is as in Definition 2.1 but replacing closure under countable unions with closure under finite unions). A map $\rho : \mathcal{A} \to [0, \infty]$ is called a *premeasure* iff

- $\rho(\varnothing) = 0$.
- If $\{ A_j \}_{j=1}^{\infty} \subseteq \mathcal{A}$ is a pairwise disjoint sequence such that $\bigcup_{j=1}^{\infty} A_j$ happens to lie in $\mathcal{A}$, then

$$\rho\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \rho(A_j) \, .$$

We see that a premeasure and measure basically obey the same axioms, the main issue is that the domain of a premeasure is merely an algebra and that of a measure is a σ-algebra. In particular, every measure is itself a premeasure.

Remark 2.74. Premeasures are also monotone: If $A \subseteq B$ then $\rho(A) \leq \rho(B)$ for all $A, B \in \mathcal{A}$, for the same reason as is true for measures.

Since a pre-measure obeys $\rho(\varnothing) = 0$, it may well be the input for Proposition 2.66 so as to obtain an outer measure out of it, φ_ρ.

However, since φ_ρ is now coming with the assumption that ρ is a premeasure, we have additionally

Proposition 2.75. *If $\mathcal{A} \subseteq \mathcal{P}(X)$ is an algebra and $\rho : \mathcal{A} \to [0, \infty]$ is a premeasure, with $\varphi_\rho : \mathcal{P}(X) \to [0, \infty]$ the outer measure induced by it in accordance to Proposition 2.66, then $\varphi_\rho|_{\mathcal{A}} = \rho$ and $\mathcal{A} \subseteq \mathcal{A}_{\varphi_\rho}$.*

Proof. (*Thanks to Ary Cheng* and *Joshua Lin*) Let $Q \in \mathcal{A}$. We want to show that

$$\varphi_\rho(Q) = \rho(Q) \, .$$

With the cover $E_1 = Q$ and $E_n = \varnothing$ for all $n \geq 2$ we get

$$\varphi_\rho(Q) \leq \rho(Q) \, .$$

For the reverse inequality, suppose that Q is covered by some sequence $\{ E_n \}_{n=1}^{\infty} \subseteq \mathcal{A}$. In principle $\bigcup_{n=1}^{\infty} E_n$ need not lie in $\mathcal{A}$ since it is merely an algebra and not a σ-algebra, so we may not plug it into ρ. However, $Q = Q \cap \bigcup_{n=1}^{\infty} E_n$, so $\{ E_n \cap Q \}_n$ is a sequence of elements in the algebra whose countable union, Q, happens to lie in the algebra. Then by monotonicity and countable subadditivity of ρ,

$$\rho(Q) = \rho\left(Q \cap \bigcup_{n=1}^{\infty} E_n\right) \leq \sum_n \rho(Q \cap E_n) \leq \sum_n \rho(E_n) \, .$$

Now take infimum over all covers E_n to get

$$\rho(Q) \leq \varphi_\rho(Q) \, .$$

Next, we want to show that $\mathcal{A} \subseteq \mathcal{A}_{\varphi_\rho}$. Let then $A \in \mathcal{A}$. We want to show that for any $Q \in \mathcal{P}(X)$,

$$\varphi_\rho(Q) \geq \varphi_\rho(Q \cap A) + \varphi_\rho(Q \cap A^c) \, .$$

By the approximation property of the infimum, for any $\varepsilon > 0$ there exists a sequence $\{ E_n^\varepsilon \}_n \subseteq \mathcal{A}$ such that

$$\varphi_\rho(Q) \quad > \quad \sum_{n=1}^\infty \rho(E_n^\varepsilon) - \varepsilon\,.$$

Then

$$\begin{aligned}\varphi_\rho(Q) + \varepsilon \quad &> \quad \sum_{n=1}^\infty \rho(E_n^\varepsilon) \\ &= \quad \sum_{n=1}^\infty \rho(E_n^\varepsilon \cap A) + \rho(E_n^\varepsilon \cap A^c) \\ &= \quad \sum_{n=1}^\infty \varphi_\rho(E_n^\varepsilon \cap A) + \varphi_\rho(E_n^\varepsilon \cap A^c)\end{aligned}$$

where in the last line we used the fact that φ_ρ restricts to ρ on $\mathcal{A}$. But now, $Q \cap A \subseteq \bigcup_n (E_n^\varepsilon \cap A)$ so by countable subaddivitiy of φ_ρ we get

$$\varphi_\rho(Q \cap A) \leq \varphi_\rho\left(\bigcup_n (E_n^\varepsilon \cap A)\right) \leq \sum_{n=1}^\infty \varphi_\rho(E_n^\varepsilon \cap A)$$

and same for A^c so all together

$$\varphi_\rho(Q) + \varepsilon \geq \varphi_\rho(Q \cap A) + \varphi_\rho(Q \cap A^c)$$

and since $\varepsilon > 0$ was arbitrary we get the result. □

Theorem 2.76 (Carathéodory's extension theorem)**.** *Let $\mathcal{A} \subseteq \mathcal{P}(X)$ be an algebra and $\rho : \mathcal{A} \to [0, \infty]$ a premeasure. In this scenario we already know that there exists a measure μ_{φ_ρ} induced by ρ via Theorem 2.70. Then, since ρ is a premeasure, we have the following additional properties:*

1. *The σ-algebra generated by $\mathcal{A}$, $\sigma(\mathcal{A})$ is contained within $\mathcal{A}_{\varphi_\rho}$ (defined according to Definition 2.69).*
2. *If $\nu : \sigma(\mathcal{A}) \to [0, \infty]$ is any measure such that $\nu|_{\mathcal{A}} = \rho$ then $\nu(E) \leq \mu_{\varphi_\rho}(E)$ for all $E \in \sigma(\mathcal{A})$ and*

$$\nu(E) = \mu_{\varphi_\rho}(E) \qquad \left(E \in \sigma(\mathcal{A}) : \mu_{\varphi_\rho}(E) < \infty\right)\,.$$

3. *If X is σ-finite w.r.t. ρ, in the sense that there exists some $\{ A_n \}_{n=1}^\infty \subseteq \mathcal{A}$ such that $\rho(A_n) < \infty$ and $X \subseteq \bigcup_{n=1}^\infty A_n$ then μ_{φ_ρ} is the* unique *extension of ρ to $\sigma(\mathcal{A})$.*

Proof. For the first statement, we know that $\mathcal{A} \subseteq \mathcal{A}_{\varphi_\rho}$ by the previous claim and since the latter is a σ-algebra and $\sigma(\mathcal{A})$ is the *smallest* σ-algebra containing $\mathcal{A}$, we get the claim.

For the second statement, let $E \in \sigma(\mathcal{A})$ and pick some cover $\{ E_n \}_n \subseteq \mathcal{A}$ such that $\bigcup_n E_n \supseteq E$. Then

$$\nu(E) \leq \sum_{n=1}^\infty \nu(E_n) = \sum_{n=1}^\infty \rho(E_n)\,.$$

Taking now infimum over the covers we get

$$\nu(E) \quad \leq \quad \mu_{\varphi_\rho}(E)\,.$$

Moreover, by (2.4) we have

$$\nu\left(\bigcup_n E_n\right) = \lim_{n\to\infty} \nu\left(\bigcup_{j=1}^n E_j\right) = \lim_{n\to\infty} \mu_{\varphi_\rho}\left(\bigcup_{j=1}^n E_j\right) = \mu_{\varphi_\rho}\left(\bigcup_n E_n\right)\,.$$

Now, if $\mu_{\varphi_\rho}(E) < \infty$, by the approximation property of the infimum let us choose for any $\varepsilon > 0$ the cover so that

$$\mu_{\varphi_\rho}\left(\bigcup_{j=1}^{\infty} E_j\right) < \mu_{\varphi_\rho}(E) + \varepsilon$$

which implies

$$\mu_{\varphi_\rho}\left(\left(\bigcup_{j=1}^{\infty} E_j\right) \setminus E\right) < \varepsilon$$

so that

$$\begin{aligned}
\mu_{\varphi_\rho}(E) &\leq \mu_{\varphi_\rho}\left(\bigcup_{j=1}^{\infty} E_j\right) \\
&= \nu\left(\bigcup_{j=1}^{\infty} E_j\right) \\
&= \nu\left(\left(\bigcup_{j=1}^{\infty} E_j\right) \cap E\right) + \nu\left(\left(\bigcup_{j=1}^{\infty} E_j\right) \cap E^c\right) \\
&= \nu(E) + \nu\left(\left(\bigcup_{j=1}^{\infty} E_j\right) \setminus E\right) \\
&\leq \nu(E) + \mu_{\varphi_\rho}\left(\left(\bigcup_{j=1}^{\infty} E_j\right) \setminus E\right) \\
&\leq \nu(E) + \varepsilon
\end{aligned}$$

but since $\varepsilon > 0$ was arbitrary we get equality.

Lastly, if X is σ-finite w.r.t. ρ, i.e., $X = \bigcup_{j=1}^{\infty} A_j$ with $A_j \in \mathcal{A}$ and $\rho(A_j) < \infty$ then (WLOG assuming A_j's are disjoint) we get for any $E \in \sigma(\mathcal{A})$,

$$\mu_{\varphi_\rho}(E) = \sum_{j=1}^{\infty} \mu_{\varphi_\rho}(E \cap A_j) = \sum_{j=1}^{\infty} \nu(E \cap A_j) = \nu(E)$$

so really $\nu = \mu_{\varphi_\rho}$. □

Remark 2.77. As we presented the theory so far, it may well happen that $\sigma(\mathcal{A}) \subsetneq \mathcal{A}_{\varphi_\rho}$. For an example: let $X := \mathbb{R}$ and pick ρ so we get the Lebesgue measure on $\mathbb{R}$. Then $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$. Now pick any subset of $\mathbb{R}$ which has measure zero, for example the Cantor set C. Thanks to Claim 2.71 the Lebesgue measure μ is complete in the sense of Definition 2.33, i.e., any subset of a zero measure set is Lebesgue measurable. That means that since $\mu(C) = 0$ then any subset of C is in Lebesgue measure, i.e., in $\mathcal{A}_{\varphi_\rho}$. But there are certainly non-Borel subsets of C. Indeed, by Theorem 2.15, the cardinality of $\mathcal{B}(C)$ is $2^{\aleph_0}$ (since the Borel sigma algebra on C can be generated by a countable subset, even though C itself is uncountable, just like the Borel sigma algebra of $\mathbb{R}$ may be generated by a countable set though $\mathbb{R}$ is uncountable; e.g. take the collection $C \cap (a, b)$ where a, b have rational endpoints) yet the power set equals $\mathcal{P}(C) \equiv 2^C$ so its cardinality is $|\mathcal{P}(C)| = 2^{2^{\aleph_0}} > 2^{\aleph_0}$, so there *must* be way more subsets of C than there are Borel subsets of C.

Actually, one can also go in the reverse direction:

Theorem 2.78. *Let a measurable space $(X, \mathcal{M})$ and a measure on it $\mu : \mathcal{M} \to [0, \infty]$ be given, such that μ is σ-finite as in Definition 2.37. Then there exists an outer measure $\varphi_\mu : \mathcal{P}(X) \to [0, \infty]$ such that when we apply the Caratheodory restriction Theorem 2.70 to it to get μ_{φ_μ} we get back $\overline{\mu}$ and $\mathcal{A}_{\varphi_\mu} \supseteq \overline{\mathcal{M}}$.*

Proof. Let us define

$$\varphi_\mu(A) := \inf\left(\left\{ \sum_{n=1}^{\infty}\mu(M_n) \,\middle|\, \{M_n\}_{n\in\mathbb{N}} \subseteq \mathcal{M} \wedge \bigcup_{n=1}^{\infty} M_n \supseteq A \right\}\right) \qquad (A \in \mathcal{P}(X)) .$$

Then using the very same proof Proposition 2.66 we get that φ_μ is an outer measure. In doing so, we treat μ as an "arbitrary" function such that $\mu(\varnothing) = 0$, we don't need any of its additional structure to show that φ_μ defined so is an outer measure. The main issue is rather to show $\mu_{\varphi_\mu} = \overline{\mu}$. Since every measure is itself a premeasure, Theorem 2.76 applies to μ_{φ_μ} to get a complete measure on $\mathcal{A}_{\varphi_\mu}$ (completeness thanks to Claim 3.10). The assumption that X is σ-finite in particular implies that μ_{φ_μ} is the unique extension of μ to $\sigma(\mathcal{M}) = \mathcal{M}$. ☐

2.9.4 The Kakutani-Markov-Riesz representation theorem [extra, Folland and Rudin]

Thanks to Olivia Kwon for contributing this section about the KMR theorem.

So far we have seen one way to construct new measures:

$$\rho \to \varphi_\rho \to \mu_{\varphi_\rho}$$

where ρ is a *premeasure*. This general strategy uses the Caratheodory extension theorem.

It turns out that there is yet another way to construct measures. It would yield yet another way to construct the Lebesgue measure. First, we start with a

Definition 2.79 (*Radon measure*). *A radon measure* is a positive Borel measure such that:

1. It is finite on every compact set
2. (Outer Regularity) It is outer regular on all Borel sets
3. (Inner Regularity) It is inner regular on all open sets.

Definition 2.80. We say a set E in a measure space is *σ-finite* if there exists $\{E_j\}_{j\in\mathbb{N}}$ such that $E = \bigcup_{j\in\mathbb{N}} E_j$ with $\mu(E_j) < \infty$ for all $j \in \mathbb{N}$.

Proposition 2.81. *(Folland 7.5) Every Radon measure μ is inner regular on all of its σ-finite sets.*

Proof. Suppose E is σ-finite. We first consider the case in which $\mu(E) < \infty$. Then, given $\epsilon > 0$, find an open set U containing E such that $\mu(U) < \mu(E) + \epsilon/2$ and a compact set F such that $\mu(F) > \mu(U) - \epsilon/2$ using the definition of Radon measure. Then, since $\mu(U - E) < \epsilon/2$, we can choose an open set V containing $U - E$ such that $\mu(V) < \epsilon/2$ as well. Define the compact set $K = F - V$. Notice that $K \subset E$ and that

$$\mu(K) = \mu(F) - \mu(F \cap V) > (\mu(E) - \epsilon/2) - \mu(V) > \mu(E) - \epsilon.$$

Therefore, we have that E is inner regular.

Now consider the case when $\mu(E) = \infty$. By σ-finiteness, we can find $\{E_j\}_{j\in\mathbb{N}}$ such that $E = \bigcup_{j\in\mathbb{N}} E_j$ with $\mu(E_j) < \infty$ for all $j \in N$. We know that for every $N \in \mathbb{N}$, there exists N_j such that $N < \mu(\bigcup_{i=1}^{N_j} E_j) < \infty$. By argument from the above paragraph, find compact set $K_N \subset \bigcup_{i=1}^{N_j} E_j$ such that $\mu(K_N) > N$ as well. Because N is arbitrary, we have that E is inner regular in this case as well. ☐

Corollary 2.82. *(Folland 7.6) Every σ-finite Radon measure is regular. If X is σ-compact, every Radon measure on X is regular.*

Definition 2.83. We say that a linear function $\Lambda : C_c(X) \to \mathbb{R}$ is *positive* if $\Lambda(f) \geq 0$ for all $f \geq 0$.

Theorem 2.84 (Kakutani-Markov-Riesz)**.** *Let X be a locally compact Hausdorff space and Λ a positive linear functional on $C_c(X \to \mathbb{C})$. Then, there exists a unique positive measure μ such that it satisfies*

1. The equality:
$$\Lambda(f) = \int_X f \, \mathrm{d}\mu \quad (f \in C_c(X \to \mathbb{C}))\,.$$
2. $\mu(K) < \infty$ for every K compact.
3. (Outer regularity)
$$\mu_\Lambda(E) = \inf\left(\{\mu_\Lambda(U) \mid E \subseteq U \wedge U \in \text{Open}(X)\}\right) \quad (E \in \mathcal{B}(X))$$
4. (Inner regularity)
$$\mu_\Lambda(E) = \sup\left(\{\mu_\Lambda(K) \mid E \supseteq K \wedge K \in \text{Compact}(X)\}\right) \quad (E \in \mathcal{B}(X) : \mu(E) < \infty)$$
5. μ is complete.

In addition, μ satisfies:

7. $\mu(U) = \sup\{\Lambda(f) : f \in C_c(X), f \prec U\}$ for all open $U \subset X$.
8. $\mu(K) = \inf\{\Lambda(f) : f \in C_c(X), K \prec f\}$ for all compact $K \subset X$.

Proof. Notice that by Proposition Proposition 2.81, it is enough to show that μ is a complete measure satisfying the first condition such that when restricted to the borel sets, it is a Radon measure, as well as properties 7 and 8. We prove the theorem in 5 steps. The moral of the proof is that we define an outer measure μ and restrict it to a sigma-algebra that satisfies the desired property, using Carathodory's construction.

Step 1 (Uniqueness) This is the easiest part of the proof. Assuming the existence of such μ, we show that it must be unique.

Note that because μ is a Radon measure, it is determined by its values on compact sets of X. This is because by inner regulariy, the measure of open sets is determined by that of compact sets, and by outer regularity, the measure of every Borel set is determined by that of open sets.

Therefore, given two measures μ_1, μ_2 satisfying the above properties, it is enough to show that they agree on compact sets to prove the uniqueness.

Given two measures μ_1, μ_2 satisfying the above properties, fix arbitrary compact set $K \subset X$ and $\epsilon > 0$. By outer regularity property, find V containing K such that $\mu_2(V) < \mu_2(K) + \epsilon$. Using Theorem E.6, find $f \in C_c(X)$ such that $K \prec f \prec V$. Then we have,
$$\mu_1(K) = \int_X \chi_K d\mu_1 \le \int_X f d\mu_1 = \Lambda(f) = \int_X f d\mu_2 \le \int_X \chi_V d\mu_2 = \mu_2(V) < \mu_2(K) + \epsilon.$$
Taking $\epsilon \to 0$, we have that $\mu_1(K) \le \mu_2(K)$. By symmetric argument, we have that $\mu_2(K) \le \mu_1(K)$ as well, giving as the desired conclusion. □

Step 2 (Defining φ and Proving that it is an Outer Measure) We first define φ on $P(X)$. First, for all open sets, define
$$\varphi(U) = \sup\{\Lambda(f) : f \in C_c(X), f \prec U\}.$$
Then for any arbitrary set $E \subset X$, define
$$\varphi(E) = \inf\{\varphi(U) : E \subset U, U \text{ open}\}.$$
We claim that φ is an outer measure.

First, notice that $\varphi(\emptyset) = 0$ because the only function satisfying $f \prec \emptyset$ is $f \equiv 0$.

Secondly, φ is monotone. This property follows because if $A \subset B$, then for all U open containing B also contains A and therefore $\mu(A) \le \mu(B)$ by construction.

Lastly, φ satisfies countable subadditivity. To prove this, we first prove finite subadditivity for open sets and then generalize to the desired claim. Given V_1, V_2 open in X, find $g \in C_c(X)$ such that $g \prec V_1 \cup V_2$. Then, by the Corollary E.7 on $K = \text{supp}(g)$, V_1, and V_2, we can find h_1 and h_2 in C_c such that $h_1(x) + h_2(x) = 1$ on $\text{supp}(g)$ and $h_i \prec V_i$ for $1 \le i \le 2$. Therefore, we have that $g = h_1 g + h_2 g$ and hence
$$\Lambda g = \Lambda h_1 g + \Lambda h_2 g \le \mu(V_1) + \mu(V_2).$$

Thus by the definition of $\mu(V_1 \cup V_2)$, we get that $\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2)$.

Now given $\{E_i\} \subset X$, we assume for all i, $\mu(E_i) < \infty$ for if there exists some i such that $\mu(E_i) = \infty$, then the inequality becomes immediate. Given arbitrary $\epsilon > 0$, using the definition of $\mu(E_i)$, for all E_i, find V_i open such that $\mu(V_i) < \mu(E_i) + \frac{\epsilon}{2^i}$. Denote $V = \bigcup_i V_i$. Given arbitrary f such that $f \prec V$, by compactness of $\text{supp}(f)$, we can take finite subcover of $\text{supp}(f)$ such that $\text{supp}(f) \subset V_{n_1} \cup \cdots \cup V_{n_N}$. Therefore,

$$\Lambda f \leq \mu(V_{n_1} \cup \cdots \cup V_{n_N}) \leq \mu(V_{n_1}) + \cdots + \mu(V_{n_N}) \leq \sum_{i \in \mathbb{N}} \mu(V_i) < \sum_{i \in \mathbb{N}} \mu(E_i) + \epsilon.$$

Hence, for $f_n \in C_c$ such that $\Lambda f_n \to \mu(V)$, we get that:

$$\begin{aligned} \Lambda f_n &< \sum_{i \in \mathbb{N}} \mu(E_i) + \epsilon \\ \mu(V) &\leq \sum_{i \in \mathbb{N}} \mu(E_i) + \epsilon && (\because \Lambda f_n \to \mu(V)) \\ \mu(V) &\leq \sum_{i \in \mathbb{N}} \mu(E_i) && (\because \text{Take } \epsilon \text{ to } 0.) \\ \mu(\bigcup_i E_i) &\leq \sum_{i \in \mathbb{N}} \mu(E_i) && (\because \mu \text{ is monotone.}) \end{aligned}$$

as desired. □

Step 3 (Proving that every open set is φ-measurable.) To show the claim, it is enough to show that if U is open and E is any subset of X with $\varphi(E) < \infty$, we have that $\varphi(E) \geq \varphi(E \cap U) + \varphi(E \cap U^c)$. We first prove it for the case when E is open, then generalize it to prove the claim.

Suppose E is open. Then, given $\epsilon > 0$, because $E \cap U$ is open, we can find $f \in C_c(X)$ such that $f \prec E \cap U$ and $\Lambda(f) > \varphi(E \cap U) - \epsilon$. Also, $E - \text{supp}(f)$ is open, so we can find $g \in C_c(X)$ such that $g \prec E - \text{supp}(f)$ and $\Lambda(g) > \varphi(E - \text{supp}(f)) - \epsilon$. Notice that by construction, $f + g \prec E$. So,

$$\begin{aligned} \varphi(E) &\geq \Lambda(f) + \Lambda(g) \\ &> \varphi(E \cap U) + \varphi(E - \text{supp}(f)) - 2\epsilon \\ &\geq \varphi(E \cap U) + \varphi(E \cap U^c) - 2\epsilon && (\because \varphi \text{ is monotone and } (E \cap U^c) \subset (E - \text{supp}(f))) \end{aligned}$$

Thus letting $\epsilon \to 0$, we have the desired inequality.

Now suppose E is an arbitrary set with $\varphi(E) < \infty$. By definition of φ, find open set V containing E such that $\varphi(V) < \varphi(E) + \epsilon$. Then, we get that:

$$\begin{aligned} \varphi(E) + \epsilon &> \varphi(V) \\ &\geq \varphi(V \cap U) + \varphi(V \cap U^c) && (\because \text{By the case when } E \text{ open applied on } V) \\ &\geq \varphi(E \cap U) + \varphi(E \cap U^c) && (\because \varphi \text{ is monotone}) \end{aligned}$$

Letting $\epsilon \to 0$, we have the desired inequality. □

Now define μ to be the measure generated by this outer measure via Caratheodory's construction μ_φ. μ's completeness follows automatically by this Caratheodory construction. Step 3 tells us that $\mathcal{M}_\mu$ contains all the Borel sets and therefore $\varphi|_{\mathcal{B}(X)}$ is a Borel measure. By definition of φ, it is immediate that it is outer regular for all sets and satisfies (1) of the property we want. Therefore, it remains for us to show the other two properties of Radon measure, the fact that $\Lambda(f) = \int_X f \, \mathrm{d}\mu \quad (f \in C_c(X \to \mathbb{C}))$ for all $f \in C_c(X)$, and that μ satisfies property (2).

Step 4 (μ satisfies (2), i.e. $\mu(K) = \inf\{\Lambda(f) : f \in C_c(X), K \prec f\}$ for all compact $K \subset X$.) Suppose that K is compact and $f \in C_c(X)$ such that $K \prec f$. For every $0 < \epsilon < 1$, define $U_\epsilon = \{x : f(x) > 1 - \epsilon\}$. Because f is continuous, we see that U_ϵ is open. Given any $g_\epsilon \prec U_\epsilon$, we have that $(1-\epsilon)^{-1} f - g_\epsilon \geq 0$ because for all $x \in \text{supp}(g_\epsilon) \subset U_\epsilon$, we must have that $(1-\epsilon)^{-1} f(x) > 1$. This means that because Λ is positive linear functional, we get that $\Lambda(g_\epsilon) \leq (1-\epsilon)^{-1} \Lambda(f)$. This means that $\mu(U_\epsilon) \leq (1-\epsilon)^{-1} \Lambda(f)$. Hence, $\mu(K) \leq \mu(U_\epsilon) \leq (1-\epsilon)^{-1} \Lambda(f)$. Letting $\epsilon \to 0$, we then get that $\mu(K) \leq \Lambda(f)$. Taking infimum over all such f's, we get that

$$\mu(K) \leq \inf\{\Lambda(f) : f \in C_c(X), K \prec f\}.$$

On the other hand, given $\epsilon > 0$, we can find open set V_ϵ containing K with $\mu(K) > \mu(V_\epsilon) - \epsilon$. By Theorem E.6 find $f_\epsilon \in C_c(X)$ such that $K \prec f_\epsilon \prec V_\epsilon$. This means $\Lambda(f_\epsilon) \le \mu(V_\epsilon)$ by definition of φ. So we get that $\Lambda(f_\epsilon) < \mu(K) + \epsilon$. Because Λ is monotone i.e. $f \ge g$ implies $\Lambda(f) \ge \Lambda(g)$, we have that taking $\epsilon \to 0$, the left hand side goes to $\inf\{\Lambda(f) : f \in C_c(X), K \prec f\}$ while the right hand side goes to $\mu(K)$. Therefore, we have that

$$\inf\{\Lambda(f) : f \in C_c(X), K \prec f\} \le \mu(K)$$

and hence the desired equality follows. □

Step 4 tells us that μ is finite on compact sets because $\inf\{\Lambda(f) : f \in C_c(X), K \prec f\} < \infty$ for all compact set K. It also tells us that μ is inner regular on open sets: If U is open, given $\alpha < \mu(U)$, we can choose $f_\alpha \in C_c(X)$ such that $f_\alpha \prec U$ and $\lambda(f_\alpha) > \alpha$ (by the definition of φ via supremum). Let $K_\alpha = \text{supp}(f_\alpha)$. Then, given $g \in C_c(X)$ such that $K_\alpha \prec g$, we have that $g - f_\alpha \ge 0$ by construction and hence $\Lambda(g) \ge \Lambda(f_\alpha) > \alpha$. This means that $\mu(K_\alpha) > \alpha$ by Step 4. This means that by the definition of infimum, μ is inner regular on U. Hence we have also shown that μ restricts to a Radon measure when restricted to its Borel sets. Now it remains us to show the last property.

Step 5 ($\Lambda(f) = \int_X f \, \mathrm{d}\mu$ for all $f \in C_c(X \to \mathbb{C})$.) It is enough to show the claim for $f \in C_c(X, [0,1])$, i.e. compactly supported continuous functions with range [0,1]. This is because $C_c(X)$ is linear span of $f \in C_c(X, [0,1])$. Given $N \in \mathbb{N}$, for $1 \le j \le N$, define $K_j = \{x : f(x) \ge \frac{j}{N}\}$ and let $K_0 = \text{supp}(f)$. Note then $K_N \subset K_{N-1} \subset \cdots \subset K_1 \subset K_0$. Also, define $f_1, \cdots, f_N \in C_c(X)$ by

$$f_j(x) = \begin{cases} 0 & \text{if } x \notin K_{j-1} \\ f(x) - \frac{j-1}{N} & \text{if } x \in K_{j-1} - K_j \\ \frac{1}{N} & \text{if } x \in K_j. \end{cases} = \min\left(\max\left(f - \frac{j-1}{N}, 0\right), \frac{1}{N}\right).$$

Then, $N^{-1}\chi_{K_j} \le f_j \le N^{-1}\chi_{K_{j-1}}$. Hence by monotonicity of integrals,

$$\frac{1}{N}\mu(K_j) \le \int_X f_j d\mu \le \frac{1}{N}\mu(K_{j-1}).$$

Also, if U is an open set containing K_{j-1} we have $Nf_j \prec U$ and so $\Lambda(f_j) \le N^{-1}\mu(U)$ by the definition of φ. Hence, by outer regularity, we get that $\Lambda(f_j) \le \frac{1}{N}\mu(K_{j-1})$. Moreover, by Step 4, we know that $\frac{1}{N}\mu(K_j) = \frac{1}{N}\inf\{\Lambda(f) : f \in C_c(X), K_j \prec f\}$ and thus $\frac{1}{N}\mu(K_j) \le \Lambda(f_j)$. Putting them together we get:

$$\frac{1}{N}\mu(K_j) \le \Lambda(f_j) \le \frac{1}{N}\mu(K_{j-1}).$$

Observe now that $f = \sum_{j=1}^N f_j$. Hence, summing above two equations for $1 \le j \le N$,

$$\frac{1}{N}\sum_{j=1}^{N}\mu(K_j) \le \int_X f d\mu \le \frac{1}{N}\sum_{j=0}^{N-1}\mu(K_j)$$
$$\frac{1}{N}\sum_{j=1}^{N}\mu(K_j) \le \Lambda(f) \le \frac{1}{N}\sum_{j=0}^{N-1}\mu(K_j)$$

Therefore, it follows that:

$$\left|\Lambda(f) - \int f d\mu\right| \le \frac{1}{N}\left(\sum_{j=0}^{N-1}\mu(K_j) - \sum_{j=1}^{N}\mu(K_j)\right) = \frac{\mu(K_0) - \mu(K_N)}{N} \le \frac{\mu(\text{supp}(f))}{N}$$

Since $\text{supp}(f)$ is compact, $\mu(\text{supp}(f)) < \infty$. Since N is arbitrary, we can let $N \to \infty$. Then we have that $\frac{\mu(\text{supp}(f))}{N} \to 0$ and thus we get that

$$\Lambda(f) = \int f d\mu$$

as desired.□

With this, we have proved the Kakutami-Markov-Riesz Representation theorem. □

3 Borel measures on topological spaces

In this chapter we want to explore the special properties of *Borel measures*. These are measures defined on the Borel σ-algebra of a topological space X: given a topological space X, we saw in Definition 2.16 that there is a natural σ-algebra induced by the topology of X, namely the Borel σ-algebra $\mathcal{B}(X)$. It is the smallest σ-algebra containing all of $\text{Open}(X)$. It turns out that the topological structure of X implies some regularity properties on Borel measure $\mu : \mathcal{B}(X) \to \mathbb{C}$. Loosely speaking, the measure of any Borel set may be approximated by open sets containing it or compact sets contained within it. To establish this regularity one needs to make additional assumptions on X as a topological space.

Let us make precise the regularity properties we seek to establish on our Borel measures:

Definition 3.1 (Regular measures and μ-regular sets)**.** Let a Borel measure $\mu : \mathcal{B}(X) \to [0, \infty]$ be given. A set $A \in \mathcal{B}(X)$ is called *μ-outer regular* iff

$$\mu(A) = \inf(\{ \mu(U) \mid A \subseteq U \in \text{Open}(X) \}) .$$

A set $A \in \mathcal{B}(X)$ is called *μ-inner regular* iff

$$\mu(A) = \sup(\{ \mu(K) \mid \text{Compact}(X) \ni K \subseteq A \}) .$$

Note that it is not a-priori clear that compact sets are measurable, but we shall only invoke this definition on Hausdorff spaces where compact implies closed and hence Borel measurable, as we see right below.

If all Borel sets are μ-outer regular, then μ is called outer regular. For inner regularity, some authors differ. Rudin defines the measure μ to be inner regular only if either all open sets and all Borel sets with finite μ measure are inner regular. Others ask that *all* Borel sets be μ-inner regular.

If μ is both outer regular and inner regular, it is called *regular*. Some authors also use the name *Radon* for measures which are both inner regular and locally finite, which, depending on the topological properties of X, may imply outer regularity.

3.1 Some topological notions

Let us present the topological definitions we will need to make on X.

In general we are interested in *separation axioms*. These are axioms that allow to separate elements, or sets, of X by open neighborhoods. These axioms are usually denoted by the label $T\sharp$ where $\sharp$ is an increasing nonnegative rational number: the higher the number the stronger the axiom, and *generally* if X is $T\sharp$ then it is also $T\left(\widetilde{\sharp}\right)$ for all $\widetilde{\sharp} \leq \sharp$ (but not always).

Definition 3.2 (Hausdorff topological space, $T2$)**.** A topological space X is called *Hausdorff* or $T2$ iff for any $x, y \in X : x \neq y$, there exist $U_{xy}, U_{yx} \in \text{Open}(X)$ such that $x \in U_{xy}$, $y \in U_{yx}$ and $U_{xy} \cap U_{yx} = \varnothing$.

Example 3.3. $\mathbb{R}$ is Hausdorff, since given any $x, y \in \mathbb{R}$ with $x \neq y$ and, WLOG, $x < y$, we may choose, say, $\varepsilon := \frac{y-x}{3}$ whence $(x - \varepsilon, x + \varepsilon)$ and $(y - \varepsilon, y + \varepsilon)$ are two disjoint neighborhoods of x, y respectively. In fact any metric space is Hausdorff.

Example 3.4. Take $X = \{ 1, 2 \}$ and $\text{Open}(X) = \{ X, \varnothing \}$. Then X is not Hausdorff.

Example 3.5. $\mathbb{R}$ with the cofinite topology is *not* Hausdorff (but it is $T1$ in fact). The cofinite topology $\mathcal{T}$ is given by

$$\mathcal{T} := \{ A \subseteq \mathbb{R} \mid A = \varnothing \vee |A^c| < \infty \} .$$

Claim 3.6. In a Hausdorff topological space, every compact set is closed.

Proof. Let $K \in \text{Compact}(X)$ and let $x \in K^c$. By the Hausdorff property, then, $\forall y \in K$, $\exists U_{xy}, U_{yx} \in \text{Open}(X)$ such that $x \in U_{xy}$, $y \in U_{yx}$ and $U_{xy} \cap U_{yx} = \varnothing$. Then

$$K \subseteq \bigcup_{y \in K} U_{yx}$$

is an open cover which by compactness has a finite subcover by some $\{ y_1, \cdots, y_n \} \subseteq K$:

$$K \subseteq \bigcup_{j=1}^{n} U_{y_j x} .$$

Define now

$$U := \bigcap_{j=1}^{n} U_{x y_j}$$

which is open (as a finite intersection of open) and also contains x, since each of the sets in the intersection contain x. We claim $U \cap K = \varnothing$. Assume otherwise. Then $\exists z \in U \cap K$. Then $z \in \bigcup_{j=1}^{n} U_{y_j x}$ and so $\exists j_z \in \{ 1, \cdots, n \}$ such that $z \in U_{y_{j_z} x}$. But we also have $z \in U$ and hence $z \in U_{x y_{j_z}}$ which is a contradiction since we have that

$$U_{y_{j_z} x} \cap U_{x y_{j_z}} = \varnothing .$$

We have thus established that $K^c \in \text{Open}(X)$, i.e., $K \in \text{Closed}(X)$. □

Remark 3.7. Note that this does *not* mean that every compact is bounded. We have no notion of bounded for general topological spaces: we need at least a topological vector space for that [Sha23b]. More commonly, we need a metric space, which further has what is known as the *Heine-Borel property*.

Corollary 3.8. *In a Hausdorff topological space, every compact set is Borel.*

Definition 3.9 (Locally finite measure)**.** A Borel measure $\mu : \mathcal{B}(X) \to \mathbb{C}$ is called *locally finite* iff $\forall x \in X$ there is some $U \in \text{Open}(X)$ such that $x \in U$ and $\mu(U) < \infty$.

Claim 3.10. Let X be a Hausdorff topological space. If a Borel measure $\mu : \mathcal{B}(X) \to \mathbb{C}$ is locally finite then $\mu(K) < \infty$ for any compact subset K.

Proof. Consider the open cover of K as

$$K \subseteq \bigcup_{x \in K} U_x$$

where U_x is the open neighborhood of any $x \in K$ which is guaranteed to have finite measure by the locally finite property of μ. By the fact that K is compact there are $x_1, \cdots, x_n$ such that

$$K \subseteq \bigcup_{j=1}^{n} U_{x_j} .$$

Then

$$\mu(K) \leq \sum_{j=1}^{n} \mu\left(U_{x_j}\right) \leq n \max_{j=1,\cdots,n} \mu\left(U_{x_j}\right) < \infty .$$

□

Definition 3.11 (σ-compact topological space)**.** A topological space X is called σ*-compact* or countably-compact iff $X = \bigcup_{n=1}^{\infty} K_n$ and each K_n is a compact subset of X.

Note that without loss of generality, since the finite union of compact is compact, we may define $\tilde{K}_n := \bigcup_{j=1}^{n} K_j$ so that $\bigcup_n K_n = \bigcup_n \tilde{K}_n$ and $\tilde{K}_n \subseteq \tilde{K}_{n+1}$ is an increasing sequence.

Example 3.12. $\mathbb{R}$ is σ-compact, since we may write $\mathbb{R} = \bigcup_{n=1}^{\infty} [n-1, n] \cup [-n, -n+1]$.

Example 3.13. The product space $\mathbb{R}^{\mathbb{N}}$ (countable Cartesian product of $\mathbb{R}$ with the product topology) is *not* σ-compact.

Proof. $\mathbb{R}^{\mathbb{N}}$ is the space of all sequences $a : \mathbb{N} \to \mathbb{R}$. Recall according to the product topology, we define $\text{Open}\left(\mathbb{R}^{\mathbb{N}}\right)$ as the coarsest topology such that the projections

$$\begin{aligned} \pi_n : \mathbb{R}^{\mathbb{N}} &\to \mathbb{R} \\ a &\mapsto a(n) \end{aligned}$$

are continuous. First we note that if $K \subseteq \mathbb{R}^{\mathbb{N}}$ is compact then it is coordinate-wise bounded. Indeed, since π_n are *by definition* continuous, and the continuous image of a compact set is compact, we have that $\pi_n(K) \in \text{Compact}(\mathbb{R})$ for all $n \in \mathbb{N}$. Then by Heine-Borel on $\mathbb{R}$, that means that $\pi_n(K)$ is closed and bounded, so there must exist some $M_n < \infty$ such that

$$\sup_{a \in K} |\pi_n(a)| \leq M_n\,.$$

Now if $\mathbb{R}^{\mathbb{N}}$ were σ-compact, we would have

$$\mathbb{R}^{\mathbb{N}} = \bigcup_{m \in \mathbb{N}} K_m$$

for some sequence of compacts K_m. Hence for all $n, m \in \mathbb{N}$ there exists $M_{nm} < \infty$ such that

$$\sup_{a \in K_m} |\pi_n(a)| \leq M_{nm} \qquad (n, m \in \mathbb{N})\,.$$

Define now a new sequence $b \in \mathbb{R}^{\mathbb{N}}$ via

$$b_n := \max\left(\{ M_{n1}, \cdots, M_{nn} \}\right) + 1\,.$$

Now by construction $b \notin K_n$ for all n since

$$b_n \quad > \quad M_{nn}\,.$$

□

Definition 3.14 (Locally compact topological space)**.** A topological space X is called *locally compact* iff $\forall x \in X \exists U \in \text{Open}(X) : x \in \overline{U} \in \text{Compact}(X)$. In words: every point has a compact neighborhood.

Example 3.15. $\mathbb{R}$ is locally compact, because for any $x \in \mathbb{R}$, the interval $(x - \varepsilon, x + \varepsilon)$ has a closure $[x - \varepsilon, x + \varepsilon]$ which is indeed compact (by, say, Heine-Borel); this holds for any $\varepsilon > 0$.

Example 3.16. $\mathbb{Q}$ (with the subspace topology from $\mathbb{R}$) is not locally compact (prove this).

Definition 3.17 (Normal topological space, $T4$)**.** A topological space X is termed *normal* iff any two disjoint closed sets have disjoint open neighborhoods: $\forall F_1, F_2 \in \text{Closed}(X)$ such that $F_1 \cap F_2 = \varnothing$ $\exists U_1, U_2 \in \text{Open}(X)$ such that $F_i \subseteq U_i$ for $i = 1, 2$ and $U_1 \cap U_2 = \varnothing$. It is termed $T4$ iff it is both normal and Hausdorff.

Definition 3.18 (Perfectly normal topological space, $T6$)**.** A topological space X is termed *perfectly normal* iff any closed set is G_δ, i.e., if for any $F \in \text{Closed}(X)$ there exists some $\{ U_n \}_{n \in \mathbb{N}} \subseteq \text{Open}(X)$ such that $F = \bigcap_{n \in \mathbb{N}} U_n$. We say that X is $T6$ iff it is perfectly normal and Hausdorff.

Clearly we can make the sequence U_n nested and decreasing by defining

$$V_n := \bigcap_{j=1}^{n} U_j$$

and noting that $\bigcap_n V_n = \bigcap U_n$, that all the V_n's are open, and $V_n \supseteq V_{n+1}$.

Example 3.19. Clearly every metric space is $T6$. The product of uncountably many non-compact metric spaces is *not* normal. Consider the space

$$X := [0,1] \times \{ 0, 1 \}$$

with

$$(x,a) < (y,b) \iff (x < y \vee [x = y \wedge a < b]) \ .$$

This makes X totally ordered and every such totally ordered set has a natural topology on it. Let $\text{Open}(X)$ be given the *order topology,* i.e., the topology generated by the basis of "open intervals" $(\alpha, \beta) \equiv \{ x \in X \mid \alpha < x < \beta \}$ together with $\{ x \in X \mid a < x \}$ and $\{ x \in X \mid x < b \}$. Then X with this topology is a Hausdorff compact normal topological space which is not perfectly normal.

3.2 Establishing regularity properties of measures from topological properties of X [Folland]

Thanks to Olivia Kwon for contributing this section about deriving Borel regularity via the KMR theorem.

In this section, we discuss further properties of Radon measures, assuming Theorem 2.84.

Proposition 3.20 (Folland 7.7)**.** *Suppose that μ is σ-finite Radon measure on X and E is a Borel set in X. Then*

1. *For every $\epsilon > 0$, there exists an open set U and a closed set F with $F \subset E \subset U$ and $\mu(U - F) < \epsilon$.*
2. *There exists $A \in F_\sigma$ and $B \in G_\delta$ such that $A \subset E \subset B$ and $\mu(B - A) = 0$.*

Proof. 1. Write $E = \bigcup_{j\in\mathbb{N}} E_j$ where E_j are pairwise disjoint and satisfies $\mu(E_j) < \infty$. Given $\epsilon > 0$, for every j, because μ is a Radon measure, we can find U_j open such that $E_j \subset U_j$ and

$$\mu(U_j) < \mu(E_j) + 2^{-(j+1)}\epsilon.$$

Let us moreover define $U = \bigcup_{j\in\mathbb{N}} U_j$. Then U is open and contains E by construction, Moreover,

$$\mu(U - E) \leq \sum_{j\in\mathbb{N}} \mu(U_j - E_j) < \frac{\epsilon}{2} \sum_{j\in\mathbb{N}} \frac{1}{2^j} = \epsilon/2.$$

Similarly, find an open set V that contains E^c and satisfies $\mu(V - E^c) < \epsilon/2$. Now, define $F = V^c$. By construction F is closed and satisfies $F \subset E \subset U$. Note also that $E - F = V - E^c$. Therefore, we have that:

$$\mu(U - F) = \mu(U - E) + \mu(E - F) = \mu(U - E) + \mu(V - E^c) < \epsilon.$$

Hence we have the first statement.

2. Now, using the first statement, we prove the second statement. For every $n \in \mathbb{N}$, we find U_n open and F_n compact such that $F_n \subset E \subset U_n$ and $\mu(U_n - F_n) < \frac{1}{n}$. Define $A = \bigcap_{n\in\mathbb{N}} U_n$ and $B = \bigcup_{n\in\mathbb{N}} F_n$. Then, we see that $B \subset E \subset A$ by construction. What's more, we have that

$$A - B = \left(\bigcap_{n\in\mathbb{N}} U_n\right) \bigcap \left(\bigcup_{n\in\mathbb{N}} F_n\right)^c = \bigcap_{n\in\mathbb{N}} (U_n \cap F_n^c) = \bigcap_{n\in\mathbb{N}} (U_n - F_n).$$

Therefore, by (2.5), we have that

$$\mu(A - B) = \lim_{n\to\infty} \mu(U_n - F_n) \leq \lim_{n\to\infty} \frac{1}{n} = 0.$$

We are done. □

Theorem 3.21. *Let X be a locally compact Hausdorff space in which every open set is σ-compact. Then, every Borel measure on X that is finite on compact sets is regular and hence Radon.*

Proof. If μ is Borel measure that is finite on compact sets, then we have that then $C_c(X) \subset L^1(\mu)$. Therefore, the map $\Lambda : C_c(X) \to \mathbb{R}$ defined by

$$\Lambda(f) = \int_X f d\mu$$

is a positive linear functional. Let ν be the restrictions of the associated unique measure given by Theorem 2.84 to the Borel sets.

We first show that μ and ν agree on open sets. Given open set $U \subset X$, using the $\sigma-$compact properties, write $U = \bigcup_{j \in \mathbb{N}} K_j$ where each K_j compact. Using Theorem E.6, find $f_1 \in C_c(X)$ such that

$$K_1 \prec f_1 \prec U.$$

Recursively, for all $n \geq 2$, find $f_n \in C_c(X)$ satisfying

$$\left(\bigcup_{j=1}^{n} K_j \right) \bigcup \left(\bigcup_{j=1}^{n-1} \operatorname{supp}(f_j) \right) \prec f_n \prec U,$$

using Theorem E.6. Then, by construction, f_n pointwise to χ_U as $n \to \infty$. Therefore, by the Theorem 2.47 twice, we get that

$$\mu(U) = \int_X \chi_U d\mu = \lim_{n\to\infty} \int_X f_n d\mu = \lim_{n\to\infty} \Lambda(f_n) = \lim \int_X f_n d\nu = \int_X \chi_U d\nu = \nu(U).$$

Now, we show that μ is regular. Given $\epsilon > 0$ and E an arbitrary Borel measure, by Proposition Proposition 2.81, we can find open set V and compact set F such that $\nu(V - F) < \epsilon$ and satisfies $F \subset E \subset U$. Note that $V - F$ is in particular open, and hence $\mu(V - F) = \nu(V - F)$.

By monotonicity,

$$\mu(V) < \mu(F) + \epsilon \leq \mu(E) + \epsilon,$$

proving that μ is outer regular.

Moreover, by motonocity again,

$$\mu(F) > \mu(E) - \epsilon \geq \mu(E) - \epsilon.$$

Since F is σ compact (as X is), we can find compact sets $\{K_j\}$ such that $\bigcup_{j\in\mathbb{N}} K_j = F$ and hence $\mu(K_j) \to \mu(F)$. Thus, we can find N big enough such that $\mu(K_N) + \epsilon \geq \mu(F)$. Thus, we have that

$$\mu(K_n) \geq \mu(E) - 2\epsilon,$$

proving that μ too is inner regular thus Radon. □

Remark 3.22. By the uniqueness of ν, we in fact have that $\mu = \nu$.

Corollary 3.23. *Let X be a locally compact Hausdorff space in which every open set is σ-compact. If Radon measures μ_1, μ_2 agree on all open sets then they are equal.*

4 The Lebesgue measure on $\mathbb{R}$

We focus our attention now to the special case of the measurable space $X := \mathbb{R}$ with the choice $\operatorname{Msrbl}(X) := \mathcal{B}(\mathbb{R})$, i.e., the Borel sigma-algebra generated by open sets on $\mathbb{R}$. Our goal is to define a positive measure $\lambda : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ which has all the robust abstract properties discussed in the previous chapter (in particular that we could invoke the limit theorems on its associated integral) and on the other hand, that it extends the notion of *length* when applied on intervals. It turns out that to zero in on a *unique* such object we need an additional property on this measure:

Definition 4.1 (translation invariance)**.** Let X be a measurable space which also has the structure of a vector space. A positive measure $\mu : \text{Msrbl}(X) \to [0,\infty]$ is *translation invariant* iff

$$\mu(S+x) \quad = \quad \mu(S) \qquad (S \in \text{Msrbl}(X), x \in X)$$

where by $S+x$ we mean the translation of the set S by x, which is a new set defined by

$$S + x \equiv \{ y + x \mid y \in S \} .$$

Our main and most immediate goal in this chapter is to prove

Theorem 4.2 (Existence and uniqueness of the Lebesgue measure on $\mathbb{R}$)**.** *There exists a* unique *positive, translation invariant measure* $\lambda : \mathcal{B}(\mathbb{R}) \to [0,\infty]$ *such that* $\lambda([0,1]) = 1$.

We call the measure λ guaranteed by the above theorem *the Lebesgue measure on* $\mathbb{R}$ and the associated integral

$$(f : \mathbb{R} \to \mathbb{C} \text{ Borel msrbl.}) \mapsto \int_{\mathbb{R}} f \mathrm{d}\lambda \in \mathbb{C}$$

the Lebesgue integral on $\mathbb{R}$.

To prove this theorem we employ the machinery to actually *construct* measures out of more primitive objects, the premeasures, which we studied in Section 2.9.

4.1 The premeasure which generates the Lebesgue measure

Claim 4.3. Let

$$\mathcal{A}_0 := \{ \varnothing \} \cup \{ (a,b] \mid a \in [-\infty,\infty), b \in \mathbb{R}, a < b \} \cup \{ (a,\infty) \mid a \in [-\infty,\infty) \} .$$

Then $\mathcal{A}_0$ is an *elementary family* in the sense of Definition B.1 below.

Proof. By definition we have $\varnothing \in \mathcal{A}_0$. Next, we want to show closure under intersection. This is clear if we take an intersection of anything with $\varnothing$. If we have

$$(a,b] \cap (a',b'] = \begin{cases} \varnothing & b < a' \vee b' < a \\ (a',b] & a < a' < b < b' \\ \text{etc.} \end{cases}$$

we see that in all cases we obtain a set of one of the forms in $\mathcal{A}_0$. Finally, if we take complements we get a finite disjoint union of elements in $\mathcal{A}_0$. Indeed,

$$\varnothing^c = (-\infty,\infty) \in \mathcal{A}_0 .$$

$$(a,b]^c = (-\infty,a] \cup (b,\infty)$$

each of which lies in $\mathcal{A}_0$ etc. □

Hence by Claim B.2 below, the set $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{A}_0$ is an algebra.

Claim 4.4. The σ-algebra generated by $\mathcal{A}$, $\sigma(\mathcal{A})$, equals $\mathcal{B}(\mathbb{R})$.

Proof. We can write

$$(a,b] = \bigcap_{j=1}^{\infty} \left(a, b + \frac{1}{j}\right)$$

so certainly $\sigma(\mathcal{A}) \subseteq \mathcal{B}(\mathbb{R})$. Conversely, every open set in $\mathbb{R}$ is the countable union of open intervals, and we may write any open interval as an element in $\sigma(\mathcal{A})$. For example,

$$(a,b) = \bigcup_{j=1}^{\infty} \left(a, b - \frac{1}{j}\right] \in \sigma(\mathcal{A}) .$$

□

Theorem 4.5. *Define $\rho : \mathcal{A} \to [0,\infty]$ via: If $\bigcup_{j=1}^{n}(a_j, b_j]$ is a disjoint union of* finite *intervals,*

$$\rho\left(\bigcup_{j=1}^{n}(a_j,b_j]\right) \quad := \quad \sum_{j=1}^{n} b_j - a_j\,, \tag{4.1}$$

$\rho\left(\bigcup_{j=1}^{n}(a_j,b_j]\right) = \infty$ if the disjoint union contains an infinite interval, and of course,

$$\rho(\varnothing) \quad := \quad 0\,.$$

Then ρ is a pre-measure.

Proof. First let us verify that ρ is well-defined, since the representation $\bigcup_{j=1}^{n}(a_j,b_j]$ for elements of $\mathcal{A}$ is *not* unique. For example

$$(0,1] = \left(0,\frac{1}{2}\right] \cup \left(\frac{1}{2},1\right] = \left(0,\frac{1}{3}\right] \cup \left(\frac{1}{3},1\right]$$

and so on. But clearly, the sum in Theorem 4.5 telescopes so this does not a problem for us. We leave the remaining cases as an exercise to the reader.

Once we know that ρ is well-defined, we need to verify the axioms in Definition 2.73. See HW3Q5 for a complete description. □

Definition 4.6 (The Lebesgue measure)**.** The Lebesgue measure on $\mathbb{R}$ is the measure obtained by $\rho \to \varphi_\rho \to \mu_{\varphi_\rho}$ according to Theorem 2.76, with the choice of ρ as in Theorem 4.5.

Theorem 2.76 yields a σ-algebra of all Lebesgue measurable subsets, named there $\mathcal{A}_{\varphi_\rho}$. We call such sets *Lebesgue measurable*. One of the conclusions of the theorem, combined with Claim 4.4, is that all Borel subsets are Lebesgue measurable. Note that the reverse inclusion is *not* true: there are Lebesgue measurable sets which are not Borel measurable.

Since $\mathbb{R}$ is σ-finite, another conclusion of the theorem is that the Lebesgue measure is the unique extension of ρ to $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$.

4.2 Proof of the uniqueness theorem

We are now ready for the

Proof of Theorem 4.2 . For existence, we take the Lebesgue measure $\mu_{\varphi_\rho} : \mathcal{A}_{\varphi_\rho} \to [0,\infty]$ from above and restrict it to

$$\lambda : \mathcal{B}(\mathbb{R}) \quad \to \quad [0,\infty]\;.$$

Clearly, since λ is an extension of ρ and $[0,1] \in \mathcal{A}$, we have $\lambda((0,1]) = \rho((0,1]) = 1$ as desired. In a minute we shall see that singletons do not matter for λ so that will imply that $\lambda([0,1]) = 1$. Moreover, λ is translation invariant because ρ is, since it is defined via differences of the endpoints of half intervals.

We finally get to uniqueness. Let $\tilde{\lambda} : \mathcal{B}(\mathbb{R}) \to [0,\infty]$ be some other translation invariant measure such that $\tilde{\lambda}([0,1]) = 1$. We want to show that $\lambda = \tilde{\lambda}$. By translation invariance, we immediately have

$$\lambda([t,t+1]) = \tilde{\lambda}([t,t+1]) \qquad (t \in \mathbb{R})\;.$$

By **??** it is sufficient to show λ and $\tilde{\lambda}$ agree on open subsets. Now, let $U \in \operatorname{Open}(\mathbb{R})$. We know that U is the countable disjoint union of open intervals, so by countable-additivity of both measures we only need to show

$$\begin{aligned} \lambda((a,b)) &= \tilde{\lambda}((a,b)) \qquad (a < b \in \mathbb{R}) \\ &\updownarrow \text{ (transl. invar.)} \\ \lambda((0,b)) &= \tilde{\lambda}((0,b)) \qquad (b > 0)\;. \end{aligned}$$

to show $\lambda(U) = \tilde{\lambda}(U)$. Let us show that translation invariant measures with $\tilde{\lambda}([0,1]) = 1$ have a scaling property. Before we start, let us also show that singletons cannot matter for $\tilde{\lambda}$, i.e., that $\tilde{\lambda}(\{x\}) = 0$ for any $x \in \mathbb{R}$ and any normalized translation invariant Borel measure $\tilde{\lambda}$. By additivity of $\tilde{\lambda}$, let $\{x_j\}_{j=1}^N$ be N distinct points within $[0,1]$. Then by translation invariance

$$\begin{aligned} N\tilde{\lambda}(\{0\}) &= \sum_{j=1}^{N} \tilde{\lambda}(\{x_j\}) \\ &= \tilde{\lambda}\left(\bigcup_{j=1}^{N} \{x_j\}\right) \\ &\leq \tilde{\lambda}([0,1]) = 1\,. \end{aligned}$$

But we can certainly pick N as large as we want, so pick it so that $N\tilde{\lambda}(\{0\}) \leq 1$ to get a contradiction with $\tilde{\lambda}(\{0\}) > 0$; thus $\tilde{\lambda}(\{0\}) = 0$. Hence for both $\tilde{\lambda}$ and λ we may ignore singletons [a].

Let us now start with showing that

$$\tilde{\lambda}\left(\left[0, \frac{1}{n}\right]\right) = \frac{1}{n}\,.$$

To that end, write the "almost" disjoint union $[0,1] = \bigcup_{k=0}^{n-1} \left[\frac{k}{n}, \frac{k+1}{n}\right]$. Using additivity and translation invariance we get

$$\begin{aligned} 1 &= \tilde{\lambda}([0,1]) \\ &= \tilde{\lambda}\left(\bigcup_{k=0}^{n-1} \left[\frac{k}{n}, \frac{k+1}{n}\right]\right) \\ &\overset{\star}{=} \sum_{k=0}^{n-1} \tilde{\lambda}\left(\left[\frac{k}{n}, \frac{k+1}{n}\right]\right) \\ &= \sum_{k=0}^{n-1} \tilde{\lambda}\left(\left[0, \frac{1}{n}\right)\right) \\ &= n\tilde{\lambda}\left(\left[0, \frac{1}{n}\right)\right) = n\tilde{\lambda}\left(\left[0, \frac{1}{n}\right]\right) \end{aligned}$$

so $\tilde{\lambda}\left(\left[0, \frac{1}{n}\right]\right) = \frac{1}{n}$ indeed. In $\star$ we used

$$\tilde{\lambda}(A \cup B) = \tilde{\lambda}(A \setminus B) + \tilde{\lambda}(B \setminus A) + \tilde{\lambda}(A \cap B)$$

to add and remove singletons of measure zero as necessary. In a similar fashion we can extract out of this the scaling property for rational end points

$$\tilde{\lambda}\left(\left[0, \frac{m}{n}\right]\right) = \frac{m}{n}\,.$$

Write $[0,m] = \bigcup_{k=0}^{n-1} m\left[\frac{k}{n}, \frac{(k+1)}{n}\right]$ so

$$\begin{aligned} m &= \tilde{\lambda}([0,m]) \\ &= \tilde{\lambda}\left(\bigcup_{k=0}^{n-1} \left[\frac{mk}{n}, \frac{m(k+1)}{n}\right]\right) \\ &= \sum_{k=0}^{n-1} \tilde{\lambda}\left(\left[\frac{mk}{n}, \frac{m(k+1)}{n}\right]\right) \\ &= n\tilde{\lambda}\left(\left[0, \frac{m}{n}\right]\right) \end{aligned}$$

yielding the result $\tilde{\lambda}\left(\left[0,\frac{m}{n}\right]\right)=\frac{m}{n}$. We then extend this to all end points $b\in\mathbb{R}$ by monotone approximation and the monotone property:

$$\tilde{\lambda}\left(\left[0,r_j\right]\right)\leq\tilde{\lambda}\left(\left[0,b\right]\right)\leq\tilde{\lambda}\left(\left[0,q_j\right]\right)$$

for all rational sequences $r_j\to b$ from below and $q_j\to b$ from above. Taking the limit $j\to\infty$ on both sides of the inequality yields

$$b=\lim_j r_j\leq\tilde{\lambda}\left(\left[0,b\right]\right)\leq\lim_j q_j=b\,.$$

We conclude $\tilde{\lambda}=\lambda$ on $\mathcal{B}\left(X\right)$ which is what we were trying to prove. □

[a]Note that for λ we should rather use $[0,1)$ rather than $[0,1]$ since we don't a-priori know yet that $\lambda\left([0,1]\right)=1$. This, however, doesn't change the argument at all since we know that λ is also translation invariant and certainly $[0,1)$ contains as many points as we need. Thanks to Kevin Xu for pointing out this discrepancy.

4.3 The Lebesgue measure on $\mathbb{R}^n$

To define the Lebesgue measure on $\mathbb{R}^n$, we appeal to the construction of a *product measure* which appears below in Section 5.1.1.

Definition 4.7 (Lebesgue measure on $\mathbb{R}^n$)**.** We *define* the Lebesgue measure on $\mathbb{R}^n$ to be the result of the product measure of n copies of the Lebesgue measure on the n-fold Cartesian product of $\mathbb{R}$.

Remark 4.8. In our convention the Lebesgue measure on $\mathbb{R}^n$ is indeed complete as it should be, because in our convention product measures are always complete, being the result of the Caratheodory construction.

Remark 4.9. Clearly we *could* have defined the Lebesgue measure on $\mathbb{R}^n$ directly using a similar premeasure as in Section 4.1, defining volumes of boxes instead of lengths of intervals. The two constructions yield the same measure by a uniqueness theorem identical to the one we proved in the one dimensional case.

Remark 4.10. What about $\mathbb{C}$ or $\mathbb{C}^n$? One also defines the Lebesgue measure on these, and the complex structure here plays no role. For the purposes of both topology (and hence measure theory, since we are using the Borel σ-algebra to construct everything) $\mathbb{C}^n\cong\mathbb{R}^{2n}$.

Theorem 4.11. *The Lebesgue measure on $\mathbb{R}^n$ is invariant not only under translations, but also under reflections and rotations. Under dilations it has a simple transformation formula.*

Proof. The invariance under translations is immediate from the fact that each constituent in the product measure is invariant w.r.t. translations and we can break an arbitrary translation in $\mathbb{R}^n$ into composition of translations in each axis. The other properties will be a direct consequence of Theorem 6.18. □

We also phrase an analog of Theorem 4.2:

Theorem 4.12. *There exists a* unique *positive, translation invariant measure $\lambda:\mathcal{B}\left(\mathbb{R}^n\right)\to[0,\infty]$ such that* $\lambda\left([0,1]^n\right)=1$.

The proof follows a similar pattern to the one-dimensional proof, with the fact that we only need to work on boxes, and on those, we may show a scaling property to get that any translation invariant Borel measure $\mu:\mathcal{B}\left(\mathbb{R}^n\right)\to[0,\infty]$ must satisfy

$$\mu\left([0,b_1]\times\cdots\times[0,b_n]\right)=b_1\cdots b_n\,.$$

4.4 Exotic phenomena

In this section we want to explore sets which are Lebesgue measurable but not Borel, or sets which are not even Lebesgue measurable. We first encountered such questions in Remark 2.77 where we saw that since $\lambda\left(C\right)=0$ for C the Cantor set, any subset of it must be Lebesgue measurable by completeness of λ. *But,* since the cardinality of $\mathcal{B}\left(C\right)$ is $\mathfrak{c}=2^{\aleph_0}$, and the cardinality of $\mathcal{P}\left(C\right)$ is $2^{2^{\aleph_0}}$, there must be Lebesgue measurable but non-Borel sets! Can we study these sets directly? What do they "look" like?

In HW4Q8 and HW4Q9 we explore some of these explicit constructions.

4.5 The relation between the Riemann and the Lebesgue integral

Now that we have defined the Lebesgue integral with respect to the Lebesgue measure on $\mathbb{R}$ and on $\mathbb{R}^n$ it is natural to study the relationship between the preexisting Riemann integral.

Recall that we only want to Riemann integrate *bounded* functions $f : [a,b] \to \mathbb{R}$ for some $a < b \in \mathbb{R}$ and that according to Lebesgue's Theorem 1.3, f is Riemann integrable iff it is continuous λ-almost-everywhere, i.e.,

$$\lambda\left(\left\{\, x \in [a,b] \mid f \text{ is not continuous at } x \,\right\}\right) \quad = \quad 0\,.$$

The first step in our study is to verify that any Riemann integrable function is at all Lebesgue measurable:

Claim 4.13. If $f : [a,b] \to \mathbb{R}$ is Riemann integrable then it is measurable w.r.t. $\mathcal{L}([a,b])$ (the Lebesgue measurable sets, on which λ is complete) on its domain and $\mathcal{B}(\mathbb{R})$ (the Borel subsets of $\mathbb{R}$) on its codomain.

Cf. with Claim 2.36.

Proof. By Theorem 1.3, f is continuous outside a set $N \subseteq [a,b]$ of measure zero. Hence by Corollary 2.18, $f|_{N^c} : N^c \to \mathbb{R}$ is continuous and hence measurable w.r.t. $(\mathcal{B}(N^c), \mathcal{B}(\mathbb{R}))$. Let $A \in \mathcal{B}(\mathbb{R})$. Then

$$f^{-1}(A) \quad = \quad \left[f^{-1}(A) \cap N\right] \sqcup \left[f^{-1}(A) \cap N^c\right]\,.$$

Since λ is complete and $\lambda(N) = 0$, $f^{-1}(A) \cap N$ is $\mathcal{L}([a,b])$-measurable. Moreover,

$$\left(f|_{N^c}\right)^{-1}(A) = f^{-1}(A) \cap N^c$$

so

$$\begin{aligned} f^{-1}(A) \cap N^c \quad &\in \quad \mathcal{B}(N^c) \\ &= \quad \sigma(\mathrm{Open}(N^c)) \\ &= \quad \sigma(\{\, U \cap N^c \mid U \in \mathrm{Open}([a,b]) \,\}) \\ &= \quad \{\, B \cap N^c \mid B \in \mathcal{B}([a,b]) \,\} \\ &\subseteq \quad \{\, B \cap N^c \mid B \in \mathcal{L}([a,b]) \,\} \end{aligned}$$

so that $f^{-1}(A) \cap N^c \in \mathcal{L}([a,b])$ as desired. □

Now, if $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then it is measurable. Since it is bounded, we necessarily we have

$$|f| \leq M$$

with $M := \sup_{x\in[a,b]} |f(x)| < \infty$ so that $f \in L^1([a,b] \to \mathbb{R}, \lambda)$ and

$$\|f\|_{L^1} \leq M(b-a)\,.$$

Hence we can also Lebesgue-integrate f. Do the two integrals always agree?

Theorem 4.14. *Let $f : [a,b] \to \mathbb{R}$ be Riemann integrable. Then its Riemann integral and its Lebesgue integral agree.*

Proof. We use the characterization of Riemann integrable as in Definition 1.1 to get that the two limits in the following equation exist and agree,

$$\lim_N L_N(f) = \lim_N U_N(f)$$

and they are both equal to *the* Riemann integral of f, where

$$L_N(f) \quad \equiv \quad \frac{b-a}{N} \sum_{n=0}^{N-1} \inf\left(\left\{\, f(x) \,\middle|\, x \in \left(a + [n, n+1]\,\frac{b-a}{N}\right) \right\}\right)\,.$$

Now we define the simple function $l_n(f) : [a,b] \to \mathbb{R}$ via

$$l_N(f)(x) = \sum_{n=0}^{N-1} \chi_{I_n}(x)\, i_n(f) \qquad (x \in [a,b])$$

with

$$I_n := a + [n, n+1]\,\frac{b-a}{N}$$

and

$$i_n(f) := \inf(\{\, f(x) \mid x \in I_n \,\})\ .$$

It doesn't quite matter that $I_n \cap I_m \neq \varnothing$ for $n \neq m$ since the overlap has Lebesgue measure zero (this could be avoided with uglier notation). Then, observe that

$$\int_{[a,b]} l_N(f)\,\mathrm{d}\lambda = L_N(f)$$

by definition of an integral of a simple function. However, it is also clear that

$$\lim_{N\to\infty} l_N(f) = f$$

pointwise wherever f is continuous (a set of full measure by Theorem 1.3), and that

$$|l_N(f)| \leq |f|\ .$$

Hence with $|f|$ a dominating L^1 function, we have

$$\begin{aligned}
\text{Riemann integral of } f &= \lim_{N\to\infty} L_N(f) \\
&= \lim_{N\to\infty} \int_{[a,b]} l_N(f)\,\mathrm{d}\lambda \\
&\overset{\text{DCT}}{=} \int_{[a,b]} f\mathrm{d}\lambda\,.
\end{aligned}$$

□

This deals with functions which are bounded on a bounded interval. We know, using the notion of an improper Riemann integral, that we can also Riemann integrate functions on unbounded intervals, or unbounded functions, via limits outside of the integral.

Example 4.15. Going back to Example 1.4, we study

$$f : (0,1) \to \mathbb{R}$$

given by

$$x \mapsto \frac{1}{\sqrt{x}}$$

and ask whether this function is Lebesgue integrable, since it is unbounded, so we may not ask whether it is Riemann integrable (and indeed only the improper Riemann integral of f exists). Since it is continuous on $(0,1)$ it is clearly Lebesgue measurable there, and as it is positive, we may well calculate its integral (though it may be infinite). Consider the sequence of positive measurable functions

$$f_n := \chi_{\left[\frac{1}{n},1\right]} f$$

which converge to f pointwise from below monotonically. As such, using the monotone convergence Theorem 2.47 we find

$$\begin{aligned}
\int_{(0,1)} f\mathrm{d}\lambda &= \int_{(0,1)} \lim_n f_n\mathrm{d}\lambda \\
&= \lim_n \int_{(0,1)} f_n\mathrm{d}\lambda\,.
\end{aligned}$$

Once we are dealing with $\int_{(0,1)} f_n\mathrm{d}\lambda$, f_n is a bounded Riemann integrable function and so using the previous theorem we can replace its Lebesgue integral with its Riemann integral $2 - \frac{2}{\sqrt{n}}$ to get the result 2.

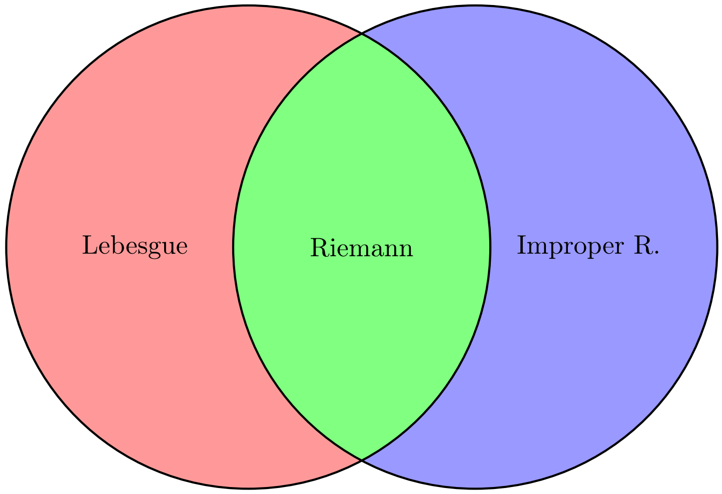

Figure 3: A Venn diagram comparing the Riemann and the Lebesgue integrals.

Be that as it may, one has to be careful because sometimes functions are improperly Riemann integrable only due to oscillations, which *cannot* help a function being Lebesgue integrable (since it always deals with absolute integrability):

Example 4.16. Consider the function $f : (0, \infty) \to \mathbb{R}$ given by $x \mapsto \frac{\sin(x)}{x}$. We can show (e.g. using contour integrals, see e.g. Example 6.40 in [Sha23a]) that the (improper) Riemann integral yields

$$\int_0^\infty \frac{\sin(x)}{x} \mathrm{d}x = \frac{\pi}{2}$$

whereas $(0, \infty) \ni x \mapsto \frac{|\sin(x)|}{x}$ is Lebesgue measurable, but does not decay to infinity quickly enough to be integrable.

Another possible (possibly simpler) example is $\mathbb{N} \ni n \mapsto \frac{(-1)^n}{n}$ w.r.t. the counting measure on $\mathbb{N}$.

Of course, there are *many* Lebesgue measurable functions which are L^1 and yet not at all Riemann integrable. An obvious example is $\chi_{[0,1]\cap\mathbb{Q}} : [0, 1] \to [0, 1]$.

5 More abstract measure theory

In this chapter we want to continue with the abstract theory that is not necessarily linked to X being a topological space. We start with the notion of product spaces. These have already been explored in HW1Q6 and in fact above we used the product structure to define the Lebesgue measure on $\mathbb{R}^n$. For the sake of completeness let us present this again in full detail.

5.1 Products [Folland]

5.1.1 Products of measurable spaces

Let $\{ X_\alpha \}_{\alpha\in A}$ be an indexed collection of non-empty sets (indexed by some set A, not necessarily countable) and let

$$\prod_{\alpha\in A} X_\alpha \equiv \left\{ f : A \to \bigcup_\alpha X_\alpha \,\middle|\, f(\beta) \in X_\beta \forall \beta \in A \right\}$$

be the Cartesian product of this collection of sets. For any $\beta \in A$, let

$$\pi_\beta : \prod_{\alpha\in A} X_\alpha \quad \to \quad X_\beta$$

be the canonical projections, i.e.,

$$\pi_\beta(f) \equiv f(\beta) \,.$$

Definition 5.1 (The product σ-algebra)**.** If we furnish each X_α with a σ-algebra $\mathcal{M}_\alpha$ then we define σ-algebra $\otimes_{\alpha\in A}\mathcal{M}_\alpha$ on the product via

$$\otimes_{\alpha\in A}\mathcal{M}_\alpha := \sigma\left(\left\{\, \pi_\alpha^{-1}(E_\alpha) \,\middle|\, E_\alpha \in \mathcal{M}_\alpha \wedge \alpha \in A \,\right\}\right) .$$

I.e., this is the smallest σ-algebra on the Cartesian product so that all projections π_α are measurable.

Claim 5.2. If A is countable then actually

$$\otimes_{\alpha\in A}\mathcal{M}_\alpha = \sigma\left(\left\{ \prod_{\alpha\in A} E_\alpha \,\middle|\, E_\alpha \in \mathcal{M}_\alpha \forall \alpha \in A \right\}\right) .$$

Proof. See HW2Q6. ☐

5.1.2 The product measure

We may continue the product construction also at the level of measures (for now only for finite products): $|A| < \infty$ and assume $\mu_\alpha : \mathcal{M}_\alpha \to [0,\infty]$ is a measure for any $\alpha \in A$. We seek to define a measure μ on $\otimes_{\alpha\in A}\mathcal{M}_\alpha$ as defined above.

Definition 5.3 (Rectangular sets)**.** Any subset $A \subseteq \prod_{\alpha\in A} X_\alpha$ of the form

$$A = \prod_{\alpha\in A} E_\alpha$$

where $E_\alpha \subseteq X_\alpha$ for all $\alpha \in A$ is called *a rectangular set*. We denote all rectangular sets of measurable sets by the symbol $\mathcal{A}_0$.

Claim 5.4. $\mathcal{A}_0$ is an elementary family in the sense of Definition B.1.

Proof. Indeed, $\varnothing \in \mathcal{A}_0$, the intersection of two rectangular sets is again rectangular, and the complement of a rectangular set is a finite disjoint union of such rectangular sets (since $|A| < \infty$). ☐

The collection of finite disjoint unions of elements in $\mathcal{A}_0$, the rectangular sets, forms an algebra $\mathcal{A}$ by Claim B.2. Then $\sigma(\mathcal{A}) = \otimes_{\alpha\in A}\mathcal{M}_\alpha$, as we saw in Claim 5.2. Hence let us define a *premeasure* $\rho : \mathcal{A} \to [0,\infty]$ given by

$$\rho\left(\sqcup_{j=1}^{n} E_{1,j} \times \cdots \times E_{|A|,j}\right) := \sum_{j=1}^{n} \prod_{\alpha\in A} \mu_\alpha\left(E_{\alpha j}\right) . \tag{5.1}$$

Claim 5.5. ρ is indeed a premeasure.

Proof. We need to verify the axioms of Definition 2.73. Clearly we have $\rho(\varnothing) = 0$. From the definition of ρ it is clear that it is *finitely* additive. Now, let $\{\, A_j \,\}_{j=1}^{\infty} \subseteq \mathcal{A}$. We assume all A_j's are pairwise disjoint, and that $\bigcup_{j=1}^{\infty} A_j$ happens to lie within $\mathcal{A}$. But $\mathcal{A}$ is itself finite disjoint unions of rectangles. So somehow the countable union $\bigcup_{j=1}^{\infty} A_j$ of rectangles happens to be a *finite* union of rectangles.

We write

$$A_j := \sqcup_{i=1}^{n_j} E_{1,i}^{j} \times \cdots \times E_{|A|,i}^{j}$$

so

$$\rho(A_j) \equiv \sum_{i=1}^{n_j} \prod_{\alpha\in A} \mu_\alpha\left(E_{\alpha,i}^{j}\right) .$$

Now by hypothesis, $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$, so

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} \sqcup_{i=1}^{n_j} E_{1,i}^j \times \cdots \times E_{|A|,i}^j$$

is a finite disjoint union of rectangles, i.e.,

$$\bigcup_{j=1}^{\infty} \sqcup_{i=1}^{n_j} E_{1,i}^j \times \cdots \times E_{|A|,i}^j = \sqcup_{k=1}^{N} F_{1,k} \times \cdots \times F_{|A|,k} .$$

Hence

$$\begin{aligned} \rho\left(\bigcup_{j=1}^{\infty} A_j\right) &= \rho\left(\sqcup_{k=1}^{N} F_{1,k} \times \cdots \times F_{|A|,k}\right) \\ &\equiv \sum_{k=1}^{N} \prod_{\alpha\in A} \mu_\alpha\left(F_{\alpha,k}\right) . \end{aligned}$$

Now,

$$\sum_{k=1}^{N} \prod_{\alpha\in A} \chi_{F_{\alpha,k}} = \sum_{k=1}^{N} \chi_{\times_{\alpha\in A} F_{\alpha,k}} = \sum_{j=1}^{\infty}\sum_{i=1}^{n_j} \chi_{\times_{\alpha\in A} E_{\alpha,i}^j} = \sum_{j=1}^{\infty}\sum_{i=1}^{n_j}\prod_{\alpha\in A} \chi_{E_{\alpha,i}^j} .$$

Integrate both sides of this equation on X_α w.r.t μ_α we get

$$\sum_{k=1}^{N} \prod_{\alpha\in A} \mu_\alpha\left(F_{\alpha,k}\right) = \sum_{j=1}^{\infty}\sum_{i=1}^{n_j}\prod_{\alpha\in A} \mu_\alpha\left(E_{\alpha,i}^j\right)$$

where on the RHS we have used the monotone convergence theorem Theorem 2.47 to exchange the j series with the integrals. As a result,

$$\rho\left(\bigcup_j A_j\right) = \sum_{j=1}^{\infty} \rho\left(A_j\right)$$

as desired. □

Definition 5.6 (The product measure)**.** By Caratheodory's procedure Theorem 2.76, $\rho \to \varphi_\rho \to \mu_{\varphi_\rho}$ yields a *complete* measure on a σ-algebra $\overline{\otimes_{\alpha\in A}\mathcal{M}_\alpha}$ which contains $\otimes_{\alpha\in A}\mathcal{M}_\alpha$. That resulting measure is defined as *the* product measure $\prod_{\alpha\in A}\mu_\alpha$ (for $|A| < \aleph_0$).

Note that by definition, on rectangular sets where each factor in the product is measurable, we have

$$\left(\prod_{\alpha\in A}\mu_\alpha\right)\left(\prod_{\alpha\in A} E_\alpha\right) \equiv \prod_{\alpha\in A}\mu_\alpha\left(E_\alpha\right) .$$

Remark 5.7. We caution the reader that there may be a strict gap between $\otimes_{\alpha\in A}\mathcal{M}_\alpha$, the product σ-algebra, and the σ-algebra $\overline{\otimes_{\alpha\in A}\mathcal{M}_\alpha}$ given by the Caratheodory construction starting from the premeasure (5.1). Indeed, the latter contains the former and there are cases when the inclusion is strict.
Thus, according to our convention the σ-algebra on which the product measure acts is automatically complete since it is the result of the Caratheodory construction. This is at odds with some authors, e.g., Folland, who let the product measure act on the product sigma-algebra, which may be incomplete. They then consider the completion of this measure, which is the same thing.

Example 5.8. Consider $\mathcal{L}$ the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}$. Then the product σ-algebra of $\mathcal{L}\otimes\mathcal{L}$ is strictly smaller than the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}^2$ (which is *defined* using the Caratheodory construction as $\overline{\mathcal{L}\otimes\mathcal{L}}$). Indeed, the latter is complete whereas the former may fail to be. There are measure zero subsets of $\mathbb{R}^2$(in particular they are Lebesgue measurable) which are not Lebesgue measurable in $\mathcal{L} \otimes \mathcal{L}$. Indeed, let $A \subseteq \mathbb{R}$ be any non-Lebesgue measurable subset (e.g. a Vitali set) and consider the set $A \times \{ 0 \}$ which is Lebesgue measurable (verify this) and of measure zero.

5.1.3 The Fubini-Tonelli Theorem

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces and $f : X \times Y \to \mathbb{C}$ be measurable w.r.t. $\mathcal{M} \otimes \mathcal{N}$. In this subsection we shall use the two projections

$$\pi_1 : X \times Y \to X$$

and

$$\pi_2 : X \times Y \to Y .$$

We also have two induced functions: for each fixed $x \in X$,

$$\begin{aligned} f_x : Y &\to \mathbb{C} \\ y &\mapsto f(x, y) \end{aligned}$$

and for each fixed $y \in Y$,

$$\begin{aligned} f_y : X &\to \mathbb{C} \\ x &\mapsto f(x, y) . \end{aligned}$$

Claim 5.9. For any $\mathcal{M} \otimes \mathcal{N}$-measurable $f : X \times Y \to \mathbb{C}$, $x \in X$ and $y \in Y$, $f_x : Y \to \mathbb{C}$ and $f_y : X \to \mathbb{C}$ are both measurable. Moreover, for any $A \in \mathcal{M} \otimes \mathcal{N}$ and fixed $x \in X$, $y \in Y$, the sets

$$A_2(x) := \pi_2((\{ x \} \times Y) \cap A) = \{ y \in Y \mid (x, y) \in A \} \subseteq Y$$

and

$$A_1(y) := \pi_1((X \times \{ y \}) \cap A) = \{ x \in X \mid (x, y) \in A \} \subseteq X$$

are both measurable. The sets A_1, A_2 are called the *sections* of A.

Proof. Let

$$\mathcal{R} := \{ E \subseteq X \times Y \mid E_1(y) \in \mathcal{M} \forall y \in Y \wedge E_2(x) \in \mathcal{N} \forall x \in X \} .$$

Clearly, if E is rectangular, i.e., if $E = U \times V$ with $U \subseteq X$ and $V \subseteq Y$ then its sections are $E_1(y) = \begin{cases} \varnothing & y \notin V \\ U & y \in V \end{cases}$ and $E_2(x) = \begin{cases} \varnothing & x \notin U \\ V & x \in U \end{cases}$. Hence $\mathcal{R}$ contains all rectangular sets. In fact $\mathcal{R}$ is a σ-algebra. Indeed, $X \times Y \in \mathcal{R}$. Moreover, $\mathcal{R}$ is closed under complements. Assume $E \in \mathcal{R}$. Let $x \in X$. We want to show that $(E^c)_2(x) \in \mathcal{N}$. We have

$$(E^c)_2(x) = \{ y \in Y \mid (x, y) \in E^c \} .$$

We claim that $(E^c)_2(x) = (E_2(x))^c$. Indeed,

$$\begin{aligned} y \in (E^c)_2(x) &\iff (x, y) \in E^c \\ &\iff (x, y) \notin E \\ &\iff y \notin E_2(x) \\ &\iff y \in (E_2(x))^c . \end{aligned}$$

So we see that really $\mathcal{R}$ is closed under complements. Actually also under countable unions, using the identity

$$\left(\bigcup_{j=1}^{\infty} E_j \right)_2 (x) = \bigcup_{j=1}^{\infty} (E_j)_2(x)$$

which may be proven similarly. But $\mathcal{M} \otimes \mathcal{N}$ is the smallest σ-algebra containing the rectangular sets.

The first statement follows from

$$(f_x)^{-1}(B) = \left(f^{-1}(B) \right)_2 (x)$$

and similarly for the other function. ◻

Definition 5.10 (Monotone class)**.** Let X be a non-empty set. A monotone class $\mathcal{C}$ on X is a subset of $\mathcal{P}(X)$ which is closed under countable increasing unions and countable decreasing intersections.

Claim 5.11. Every σ-algebra is a monotone class, and the intersection of any family of monotone classes is a monotone class. Hence for any $\mathcal{E} \subseteq \mathcal{P}(X)$, there is a unique *smallest monotone class generated by* $\mathcal{E}$, $\mathcal{C}(\mathcal{E})$.

Proof. TODO ◻

Lemma 5.12 (Monotone class lemma)**.** *If $\mathcal{A}$ is an algebra of subsets of X then*

$$\mathcal{C}(\mathcal{A}) = \sigma(\mathcal{A}) \ .$$

Proof. By the above claim, any σ-algebra is itself a monotone class, so by definition

$$\mathcal{C}(\mathcal{A}) \subseteq \sigma(\mathcal{A}) \ .$$

We will show that $\mathcal{C}(\mathcal{A})$ is actually a σ-algebra (due to the fact $\mathcal{A}$ is not just any set, but an algebra) which will finish the proof. To this end, for any $E \in \mathcal{C}(\mathcal{A})$, let

$$\mathcal{D}_E(\mathcal{A}) \quad := \quad \{\, F \in \mathcal{C}(\mathcal{A}) : E \setminus F, F \setminus E, E \cap F \in \mathcal{C}(\mathcal{A}) \,\} \ .$$

Clearly we have $\varnothing, E \in \mathcal{D}_E(\mathcal{A})$, and

$$F \in \mathcal{D}_E(\mathcal{A}) \Longleftrightarrow E \in \mathcal{D}_F(\mathcal{A}) \ .$$

Actually $\mathcal{D}_E(\mathcal{A})$ is itself also a monotone class. In fact, if $E \in \mathcal{A}$ then $F \in \mathcal{D}_E(\mathcal{A})$ for all $F \in \mathcal{A}$ as $\mathcal{A}$ is an algebra, so

$$\mathcal{A} \subseteq \mathcal{D}_E(\mathcal{A})$$

and hence

$$\mathcal{C}(\mathcal{A}) \subseteq \mathcal{D}_E(\mathcal{A}) \ .$$

Hence if $F \in \mathcal{C}(\mathcal{A})$, $F \in \mathcal{D}_E(\mathcal{A})$ for all $E \in \mathcal{A}$. But that means that $E \in \mathcal{D}_F(\mathcal{A})$ for all $E \in \mathcal{A}$, so

$$\mathcal{A} \subseteq \mathcal{D}_F(\mathcal{A})$$

and hence

$$\mathcal{C}(\mathcal{A}) \subseteq \mathcal{D}_F(\mathcal{A}) \ .$$

Hence, if $E, F \in \mathcal{C}(\mathcal{A})$, then $E \setminus F$, $F \setminus E$ and $E \cap F$ are all in $\mathcal{C}(\mathcal{A})$. Now $X \in \mathcal{A} \subseteq \mathcal{C}(\mathcal{A})$, so $\mathcal{C}(\mathcal{A})$ is an algebra. Now if $\{\, E_j \,\}_{j=1}^{\infty} \subseteq \mathcal{C}(\mathcal{A})$ then

$$\bigcup_{j=1}^{n} E_j \in \mathcal{C}(\mathcal{A})$$

as it is an algebra. But moreover, $\mathcal{C}(\mathcal{A})$ is closed under countable increasing unions such $\left\{\, \bigcup_{j=1}^{n} E_j \,\right\}_n$, i.e.,

$$\bigcup_{j=1}^{\infty} E_j = \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{n} E_j \,,$$

so

$$\bigcup_{j=1}^{\infty} E_j \in \mathcal{C}(\mathcal{A})$$

and we are done. ◻

Theorem 5.13 (Relate product measure to integral on sections)**.** *Let $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ be two σ-finite measure spaces. If $E\in\mathcal{M}\otimes\mathcal{N}$ then*

$$x\mapsto\nu\left(E_2\left(x\right)\right),y\mapsto\mu\left(E_1\left(y\right)\right)$$

are measurable on X and Y respectively, and

$$\left(\mu\times\nu\right)\left(E\right)=\int_X\nu\left(E_2\right)\mathrm{d}\mu=\int_Y\mu\left(E_1\right)\mathrm{d}\nu\,.$$

Proof. Case 1: μ and ν are finite.

Let $\mathcal{C}$ be the set of all subsets $E\in\mathcal{M}\otimes\mathcal{N}$ for which the statement of the theorem hold. Clearly, if E is rectangular of the form $E=A\times B$ then

$$\nu\left(E_2\left(x\right)\right)=\chi_A\left(x\right)\nu\left(B\right)\,,\qquad\mu\left(E_1\left(y\right)\right)=\mu\left(A\right)\chi_B\left(y\right)$$

and so $E\in\mathcal{C}$. By additivity the same is true for finite disjoint unions of rectangles, and we know that those form an algebra, so by Lemma 5.12 it is sufficient to show that $\mathcal{C}$ is a monotone class generated by rectangular measurable sets. Let $\left\{\,E_n\,\right\}_n$ be an increasing sequence in $\mathcal{C}$. Want to show that $E:=\bigcup_n E_n\in\mathcal{C}$. Consider

$$f_n\left(y\right):=\mu\left(\left(E_n\right)_1\left(y\right)\right)\qquad\left(y\in Y\right)\,.$$

This forms a sequence of measurable functions which increase pointwise to

$$f\left(y\right)=\mu\left(E_1\left(y\right)\right)\,.$$

So f is measurable and by the monotone convergence theorem Theorem 2.47,

$$\int\mu\left(E_1\right)\mathrm{d}\nu=\lim_n\int\mu\left(\left(E_n\right)_1\right)\mathrm{d}\nu=\lim_n\left(\mu\times\nu\right)\left(E_n\right)=\left(\mu\times\nu\right)\left(E\right)$$

where in the last step we have used (2.4). Similarly for the other integral.

Next, if $\left\{\,E_n\,\right\}_n$ is a decreasing sequence in $\mathcal{C}$, we want $E:=\bigcap_n E_n\in\mathcal{C}$. The function

$$y\mapsto\mu\left(E_1\left(y\right)\right)$$

is in $L^1\left(\nu\right)$ since

$$\mu\left(E_1\left(y\right)\right)\leq\mu\left(X\right)<\infty$$

and $\nu\left(Y\right)<\infty$ by assumption. Hence we may invoke the bounded convergence theorem Corollary 2.62 may be applied to show $E\in\mathcal{C}$.

Case 2: If μ,ν are σ-finite, we write $X\times Y$ as the union of an increasing sequence of rectangles $\left\{\,X_j\times Y_j\,\right\}_j$ each of which has finite (product) measure. Now for any $E\in\mathcal{M}\otimes\mathcal{N}$, Case 1 applies to $E\cap\left(X_j\times Y_j\right)$ for each j to give

$$\left(\mu\times\nu\right)\left(E\cap\left(X_j\times Y_j\right)\right)=\int_{X_j}\nu\left(E_2\left(x\right)\cap Y_j\right)\mathrm{d}\mu$$

and now we apply the monotone convergence theorem again to get the result. $\square$

Theorem 5.14 (Tonelli)**.** *Let $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ be two σ-finite measure spaces (as in Definition 2.37) and $f:X\times Y\to[0,\infty]$ be measurable w.r.t. $\mathcal{M}\otimes\mathcal{N}$. Then*

$$\int_{X\times Y}f\mathrm{d}\mu\times\nu=\int_X\left(x\mapsto\int_Y f_x\mathrm{d}\nu\right)\mathrm{d}\mu=\int_Y\left(y\mapsto\int_X f_y\mathrm{d}\mu\right)\mathrm{d}\nu\,.$$

Proof. If f is a characteristic function onto a measurable set then we are finished by Theorem 5.13. By linearity of the integral it therefore holds for nonnegative simple functions. For the general case, let $\left\{\,f_n\,\right\}_n$ be a sequence of simple functions that increase pointwise to f as in Theorem 2.27. By Theorem 2.47,

$$x \mapsto \int_Y (f_n)_x \, \mathrm{d}\nu$$

is an increasing sequence to

$$x \mapsto \int_Y f_x \mathrm{d}\nu$$

and similarly for f_y, so that these limits are measurable. It also implies that

$$\begin{aligned} \int_X \left(x \mapsto \int_Y f_x \mathrm{d}\nu \right) \mathrm{d}\mu &= \lim_n \int_X \left(x \mapsto \int_Y (f_n)_x \, \mathrm{d}\nu \right) \mathrm{d}\mu \\ &= \lim_n \int f_n \mathrm{d}(\mu \times \nu) \\ &= \int f \mathrm{d}(\mu \times \nu) \end{aligned}$$

and similarly for f_y. □

Theorem 5.15 (Fubini)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces and $f : X \times Y \to \mathbb{C}$ be measurable w.r.t. $\mathcal{M} \otimes \mathcal{N}$ such that*

$$f \in L^1 (X \times Y, \mu \times \nu) \, .$$

Then

$$\int_{X \times Y} f \mathrm{d}\mu \times \nu = \int_X \left(x \mapsto \int_Y f_x \mathrm{d}\nu \right) \mathrm{d}\mu = \int_Y \left(y \mapsto \int_X f_y \mathrm{d}\mu \right) \mathrm{d}\nu \, .$$

Proof. Apply the above to the positive and negative parts of the real and imaginary parts of f separately. Moreover, $\left(x \mapsto \int_Y f_x \mathrm{d}\nu\right)$ is finite ν-almost-everywhere, i.e., $f_x \in L^1(\nu)$ for almost every x and similarly for f_y. □

Note that if $f : X \times Y \to \mathbb{C}$ is measurable w.r.t. $\mathcal{M} \otimes \mathcal{N}$ then it is automatically measurable w.r.t. $\overline{\mathcal{M} \otimes \mathcal{N}}$ since

$$\mathcal{M} \otimes \mathcal{N} \subseteq \overline{\mathcal{M} \otimes \mathcal{N}} \, .$$

The reverse is of course false, and sometimes we may want to consider functions measurable w.r.t. $\overline{\mathcal{M} \otimes \mathcal{N}}$.

Theorem 5.16 (Fubini-Tonelli for complete products)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite complete measure spaces and $f : X \times Y \to \mathbb{C}$ be measurable w.r.t. $\overline{\mathcal{M} \otimes \mathcal{N}}$ such that either*

$$f \in L^1 (X \times Y, \mu \times \nu)$$

or $f \geq 0$. Then $f_x : Y \to \mathbb{C}$, $f_y : X \to \mathbb{C}$ are measurable for almost-all x, y respectively,

$$x \mapsto \int_Y f_x \mathrm{d}\nu, \qquad y \mapsto \int_X f_y \mathrm{d}\mu$$

are measurable and, if $f \in L^1$, they are also integrable, and

$$\int_{X \times Y} f \mathrm{d}\mu \times \nu = \int_X \left(x \mapsto \int_Y f_x \mathrm{d}\nu \right) \mathrm{d}\mu = \int_Y \left(y \mapsto \int_X f_y \mathrm{d}\mu \right) \mathrm{d}\nu \, .$$

Proof. TODO □

Example 5.17. Let $R := [0, 2] \times [0, 1] \subseteq \mathbb{R}^2$ and define $F : R \to \mathbb{R}$ via

$$F(x, y) := x \mathrm{e}^y \, .$$

We want to calculate

$$\int_R F \mathrm{d}\lambda\,.$$

Since F is continuous, it is measurable. Since $F \geq 0$ and $\lambda(R) < \infty$, Theorem 5.14 applies and we may thus calculate this integral iteratively:

$$\begin{aligned}\int_R F\mathrm{d}\lambda &= \int_{[0,2]} x \mapsto \left[\int_{[0,1]} (y \mapsto x\mathrm{e}^y)\,\mathrm{d}\lambda\right]\mathrm{d}\lambda \\ &= \int_{[0,2]} x \mapsto x\left(\mathrm{e}^1 - \mathrm{e}^0\right)\mathrm{d}\lambda \\ &= (\mathrm{e}-1)\,\frac{1}{2}\,(4-0) \\ &= 2\,(\mathrm{e}-1)\,.\end{aligned}$$

Example 5.18 (Counter-example)**.** Consider the product space

$$\mathbb{N}\times\mathbb{N}$$

with the counting measure c on it. Now consider the function

$$f(x,y) := \begin{cases} 1 & x = y \\ -1 & x = y+1 \\ 0 & \text{else}\end{cases} = \delta_{xy} - \delta_{x,y+1}\,.$$

This function is measurable since everything is measurable w.r.t. $\mathscr{P}(\mathbb{N})$. It is *not* L^1 since

$$|f|(x,y) := \begin{cases} 1 & x = y \\ 1 & x = y+1 \\ 0 & \text{else}\end{cases} = \delta_{xy} + \delta_{x,y+1}$$

and so by Tonelli's theorem,

$$\begin{aligned}\int_{\mathbb{N}\times\mathbb{N}} |f|\,\mathrm{d}(c\times c) &= \int_{y\in\mathbb{N}}\left[\int_{x\in\mathbb{N}} |f|(x,y)\,\mathrm{d}c(x)\right]\mathrm{d}c(y) \\ &= \int_{y\in\mathbb{N}}\left[\int_{x\in\mathbb{N}} (\delta_{xy}+\delta_{x,y+1})\,\mathrm{d}c(x)\right]\mathrm{d}c(y) \\ &= \sum_{y=1}^{\infty}\sum_{x=1}^{\infty}(\delta_{xy}+\delta_{x,y+1}) \\ &= \sum_{y=1}^{\infty} 2 = \infty\,.\end{aligned}$$

As a result, Fubini's theorem is *not* applicable. And indeed we see that the iterated integrals do *not* agree:

$$\sum_{y=1}^{\infty}\sum_{x=1}^{\infty}(\delta_{xy}-\delta_{x,y+1}) = \sum_{y=1}^{\infty} 0 = 0 \neq 1 = \sum_{y=1}^{\infty}\delta_{y,1} = \sum_{x=1}^{\infty}\sum_{y=1}^{\infty}(\delta_{xy}-\delta_{x,y+1})\,.$$

Example 5.19 (Counter-example)**.** Let $X = Y = [0,1]$ both with $\mathscr{B}([0,1])$. On X choose the Lebesgue measure but on Y choose the counting measure. Then one may verify that $X\times Y$ is *not* σ-finite. Indeed, if $D := \{\,(x,y)\in X\times Y \mid x = y\,\}$ then

$$\left|\left\{\int_{X\times Y}\chi_D \mathrm{d}\mu\times\nu,\int_X\left(\int_Y \chi_D\mathrm{d}\nu\right)\mathrm{d}\mu,\int_Y\left(\int_X\chi_D\mathrm{d}\mu\right)\mathrm{d}\nu\right\}\right| = 3\,.$$

5.2 Push forward and pull back measures [Not sure about the source]

Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{N})$ be a measurable space and $\varphi : X \to Y$ be measurable.

Definition 5.20 (The push-forward measure)**.** Define a new measure $\mu_\varphi : \mathcal{N} \to \mathbb{C}$ on Y via

$$\mu_\varphi (A) := \mu \left(\varphi^{-1} (A)\right) \qquad (A \in \mathcal{N}) \ .$$

μ_φ is called the *push forward* of μ by φ. In terms of maps,

$$\mu_\varphi := \mu \circ \varphi^{-1}$$

which makes sense since φ^{-1}, while not being a function $Y \to X$, *is* a function $\mathscr{P}(Y) \to \mathscr{P}(X)$; of course $\varphi^{-1}(A)$ is the *preimage* of A under φ: we are *not* assuming φ is invertible.

Claim 5.21. The push-forward measure is a well-defined measure.

Proof. Clearly as φ is measurable, $\varphi^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{N}$, so the formula for μ_φ makes sense on its domain. It obeys the axioms of a measure Definition 2.28. Indeed, first we show that μ_φ is not infinite on *all* sets. Take

$$\mu_\varphi (\varnothing) \equiv \mu \left(\varphi^{-1} (\varnothing)\right) = \mu (\varnothing) = 0 < \infty$$

since μ is a measure. For countable additivity, let $\{ A_j \}_j \subseteq \mathcal{N}$ be a disjoint sequence. Then

$$\begin{aligned} \mu_f \left(\bigcup_j A_j\right) &\equiv \mu \left(f^{-1} \left(\bigcup_j A_j\right)\right) \\ &= \mu \left(\bigcup_j f^{-1} (A_j)\right) \\ &= \sum_j \mu_f (A_j) \end{aligned}$$

where we have used the fact that the preimage preserves disjointness: If $A \cap B = \varnothing$ then $f^{-1}(A) \cap f^{-1}(B) = \varnothing$. Indeed,

$$f^{-1}(A) \cap f^{-1}(B) = f^{-1}(A \cap B) = f^{-1}(\varnothing) = \varnothing \, .$$

□

The following result is the weakest form of change of variable formula, which relates the integral w.r.t. the push forward measure μ_φ to integrals w.r.t. μ. It is so general that it doesn't even need any type of invertibility for φ.

Theorem 5.22 (Abstract measure-theoretic change of variables formula)**.** *Let* $(X, \mathcal{M}, \mu)$ *be a measure space,* $(Y, \mathcal{N})$ *be a measurable space and* $\varphi : X \to Y$ *be measurable. Then for any* $f : Y \to \mathbb{C}$ *in* $L^1(Y, \mu_\varphi)$*,*

$$f \circ \varphi \in L^1 (X, \mu)$$

and

$$\int_X f \circ \varphi \mathrm{d}\mu = \int_Y f \mathrm{d}\mu_\varphi \, . \tag{5.2}$$

Proof. Step 1: Assume that f is a simple nonnegative measurable function of the form:

$$f = \sum_{i=1}^{n} \alpha_i \chi_{A_i} \, .$$

Then

$$\int_Y f \mathrm{d}\mu_\varphi \equiv \sum_{i=1}^{n} \alpha_i \mu_\varphi (A_i) \qquad \text{(def of int on simple func)}$$

$$= \sum_{i=1}^{n} \alpha_i \mu \left(\varphi^{-1}(A_i)\right) \qquad \text{(def of msr)}$$

$$= \sum_{i=1}^{n} \alpha_i \int_X \chi_{\varphi^{-1}(A_i)} \mathrm{d}\mu \qquad \text{(int of char func is msr)}$$

$$= \int_X \sum_{i=1}^{n} \alpha_i \chi_{\varphi^{-1}(A_i)} \mathrm{d}\mu \qquad \text{(linearity)}$$

$$= \int_X f \circ \varphi \mathrm{d}\mu \,.$$

In the last line we have used the fact that

$$\chi_{\varphi^{-1}(A)} = \chi_A \circ \varphi \,.$$

Indeed,

$$\begin{aligned} \chi_{\varphi^{-1}(A)}(x) = 1 \quad &\Longleftrightarrow \quad x \in \varphi^{-1}(A) \\ &\Longleftrightarrow \quad \varphi(x) \in A \\ &\Longleftrightarrow \quad \chi_A(\varphi(x)) = 1 \,. \end{aligned}$$

So we learn that (5.2) holds for nonnegative simple measurable functions.

Step 2: Assume $f : Y \to [0,\infty]$ is measurable. Then by Theorem 2.27 there is a sequence $\{ f_n \}_n$ of simple nonnegative measurable functions which converges monotonically from below, pointwise, to f:

$$0 \le f_n \le f_{n+1} \le f \,.$$

Then by the monotone convergence Theorem 2.47 we have

$$\begin{aligned} \int_Y f \mathrm{d}\mu_\varphi &= \lim_{n\to\infty} \int_Y f_n \mathrm{d}\mu_\varphi \\ &= \lim_{n\to\infty} \int_X f_n \circ \varphi \mathrm{d}\mu \,. \end{aligned}$$

But now, $f_n \circ \varphi$ is a monotone sequence that converges to $f \circ \varphi$, by construction. Thus, invoking again the monotone convergence theorem we get

$$\lim_{n\to\infty} \int_X f_n \circ \varphi \mathrm{d}\mu = \int_X f \circ \varphi \mathrm{d}\mu$$

and hence (5.2) for $f : Y \to [0,\infty]$ measurable.

Step 3: Assume $f : Y \to \mathbb{C}$ is measurable and $L^1(Y, \mu_\varphi)$. Then by (2.14) we have

$$\begin{aligned} \int_Y f \mathrm{d}\mu_\varphi &\equiv \int_Y \mathbb{Re}\{f\}^+ \mathrm{d}\mu_\varphi - \int_Y \mathbb{Re}\{f\}^- \mathrm{d}\mu_\varphi + \mathrm{i}\int_Y \mathbb{Im}\{f\}^+ \mathrm{d}\mu_\varphi - \mathrm{i}\int_Y \mathbb{Im}\{f\}^- \mathrm{d}\mu_\varphi \\ &= \int_X \mathbb{Re}\{f\}^+ \circ \varphi \mathrm{d}\mu - \int_X \mathbb{Re}\{f\}^- \circ \varphi \mathrm{d}\mu + \mathrm{i}\int_X \mathbb{Im}\{f\}^+ \circ \varphi \mathrm{d}\mu - \mathrm{i}\int_X \mathbb{Im}\{f\}^- \circ \varphi \mathrm{d}\mu \\ &= \int_X \mathbb{Re}\{f \circ \varphi\}^+ \mathrm{d}\mu - \int_X \mathbb{Re}\{f \circ \varphi\}^- \mathrm{d}\mu + \mathrm{i}\int_X \mathbb{Im}\{f \circ \varphi\}^+ \mathrm{d}\mu - \mathrm{i}\int_X \mathbb{Im}\{f \circ \varphi\}^- \mathrm{d}\mu \\ &\equiv \int_X f \circ \varphi \mathrm{d}\mu \,. \end{aligned}$$

□

If, however, we *do* assume that φ is at least somewhat invertible (i.e. it is injective but we don't even assume that the left inverse is measurable!) then we may *localize* the integrals to subsets.

Corollary 5.23 (Change of variables formula with injective map)**.** *Let $(X,\mathcal{M},\mu)$ be a measure space, $(Y,\mathcal{N})$ be a measurable space and $\varphi : X \to Y$ be measurable and injective. Then for any $f : Y \to \mathbb{C}$ in $L^1(Y,\mu_\varphi)$,*

$$f \circ \varphi \in L^1(X,\mu)$$

and

$$\int_A f \circ \varphi \mathrm{d}\mu = \int_{\varphi(A)} f \mathrm{d}\mu_\varphi \qquad (A \in \mathcal{M}) \,. \tag{5.3}$$

Proof. We have by the above

$$\begin{aligned}\int_{\varphi(A)} f \mathrm{d}\mu_\varphi &= \int_Y \chi_{\varphi(A)} f \mathrm{d}\mu_\varphi \\ &= \int_X \left(\chi_{\varphi(A)} \circ \varphi\right) (f \circ \varphi) \mathrm{d}\mu \,.\end{aligned}$$

But we claim that if φ is injective then $\chi_{\varphi(A)} \circ \varphi = \chi_A$. Indeed,

$$\begin{aligned}\left(\chi_{\varphi(A)} \circ \varphi\right)(x) = 1 \quad &\Longleftrightarrow \quad \varphi(x) \in \varphi(A) \\ &\Longleftrightarrow \quad \varphi(x) = \varphi(a) \, \exists a \in A \,.\end{aligned}$$

Now if $x \in A$ then take $a = x$. If $x \notin A$ then there cannot exist $a \in A$ with $\varphi(x) = \varphi(a)$ because that would imply $x = a$ which would imply $x \in A$. We conclude $\varphi(x) = \varphi(a) \, \exists a \in A$ if and only if $x \in A$. ☐

We note that we could have localized without assuming φ is injective also previously but we then would've been stuck with

$$\int_{\varphi(A)} f \mathrm{d}\mu_\varphi = \int_X \left(\chi_{\varphi(A)} \circ \varphi\right) (f \circ \varphi) \mathrm{d}\mu \,.$$

If φ is not injective then

$$\chi_{\varphi(A)} \circ \varphi \geq \chi_A$$

but the two could fail to be equal.

Example 5.24. Consider $\varphi(t) = t^2$ for $t \in \mathbb{R}$ which is not injective and $A = [-1,2]$. Then $\varphi(A) = [0,4]$ and then

$$\chi_{[0,4]}\left(t^2\right) = \chi_{[0,2]}(|t|) = \chi_{[-2,2]}(t) > \chi_{[-1,2]}(t) \,.$$

5.3 Important inequalities

Theorem 5.25 (Jensen)**.** *Let $(X,\mathfrak{M},\mu)$ be a measure space with $\mu : \mathfrak{M} \to [0,\infty)$ a measure such that $\mu(X) = 1$. Let $f \in L^1(X \to (a,b),\mu)$ for some $a < b \in \overline{\mathbb{R}}$ and $\varphi : (a,b) \to \mathbb{R}$ be convex. Then*

$$\varphi\left(\int_X f \mathrm{d}\mu\right) \leq \int_X \varphi \circ f \mathrm{d}\mu \,.$$

Note we do not *require $\varphi \circ f \in L^1(\mu)$. It may well happen the RHS is ∞.*

Proof. This is in HW5Q6. ☐

Definition 5.26 (Conjugate pairs)**.** Let $p, q \in [1,\infty]$. If $\frac{1}{p} + \frac{1}{q} = 1$ then we say that p and q are *conjugate pairs.*

Theorem 5.27 (Hölder's inequality)**.** *Let $p, q \geq 1$ be a conjugate pair with $p \in (1,\infty)$ and $(X,\mathfrak{M},\mu)$ a measure space. Let $f, g : X \to [0,\infty]$ be two measurable functions. Then*

$$\int_X fg \mathrm{d}\mu \leq \left(\int_X f^p \mathrm{d}\mu\right)^{\frac{1}{p}} \left(\int_X g^q \mathrm{d}\mu\right)^{\frac{1}{q}} \,.$$

Proof. This is in HW5Q8. ☐

Theorem 5.28 (Minkowski)**.** *Let $p \in (1,\infty)$ and $(X,\mathfrak{M},\mu)$ a measure space. Let $f,g : X \to [0,\infty]$ be two measurable functions. Then*

$$\left(\int_X (f+g)^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} \leq \left(\int_X f^p \mathrm{d}\mu\right)^{\frac{1}{p}} + \left(\int_X g^p \mathrm{d}\mu\right)^{\frac{1}{p}} .$$

Proof. This is in HW5Q7. ☐

5.4 The L^2 structure of a measure space

We have seen that given a measure space $(X,\mathfrak{M},\mu)$ there is a space of integrable functions

$$L^1(\mu) \equiv \left\{ f : X \to \mathbb{C} \,\middle|\, f \text{ is msrbl. and } \int_X |f| \,\mathrm{d}\mu < \infty \right\} .$$

We saw that $L^1(\mu)$ is a $\mathbb{C}$-vector space. Actually one has to be aware of the fact that this vector space is *infinite dimensional* and it is in this sense that it is way richer than the space $\mathbb{C}^n$, which is basically entirely determined by its dimension. For infinite dimensional vector spaces, the *topology* becomes much important. One convenient way to deal with topological questions in Hilbert space is via a norm. In fact $L^1(\mu)$ *is* a normed vector space with the norm

$$\|f\|_{L^1(\mu)} \equiv \|f\|_1 := \int_X |f| \,\mathrm{d}\mu . \tag{5.4}$$

Claim 5.29. The formula in (5.4) yields a norm.

Proof. We follow the axioms of Definition C.1 below: Let $\alpha \in \mathbb{C}$ and $f \in L^1(\mu)$. Then

$$\|\alpha f\|_1 \equiv \int_X |\alpha f| \,\mathrm{d}\mu = \int_X |\alpha| \, |f| \,\mathrm{d}\mu \overset{\star}{=} |\alpha| \int_X |f| \,\mathrm{d}\mu \equiv |\alpha| \, \|f\|_1$$

where in $\star$ we have used the linearity of the integral Theorem 2.57. For the triangle inequality, let $f,g \in L^1(\mu)$. Then

$$\|f+g\|_1 \equiv \int_X |f+g| \,\mathrm{d}\mu .$$

Now we invoke the triangle inequality at the level of the complex plane: for any $x \in X$,

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| .$$

Plug this in and use the linearity of the integral to find

$$\int_X |f+g| \,\mathrm{d}\mu \leq \int_X |f| \,\mathrm{d}\mu + \int_X |g| \,\mathrm{d}\mu \equiv \|f\|_1 + \|g\|_1 .$$

Finally, we want to show that if $f \in L^1(\mu)$ has $\|f\|_1 = 0$ then $f = 0$. This is actually *false*. Consider for instance the function

$$f = \chi_C$$

where C is the Cantor set, $X = \mathbb{R}$ and $\mu = \lambda$. Then

$$\|f\|_1 \equiv \lambda(C) = 0$$

yet f is clearly *not* the zero function. ☐

Let's try again. We define an equivalence relation on $L^1(\mu)$:

$$f \sim g \iff \mu(\{ x \in X \mid f(x) \neq g(x) \}) = 0 \, .$$

Before we begin we must show that the set $\{ x \in X \mid f(x) \neq g(x) \}$ is measurable. We have proven in HW1Q10 that this is indeed so for functions whose domain is $\overline{\mathbb{R}}$. Here it is even easier:

$$\begin{aligned} \{ x \in X \mid f(x) \neq g(x) \} &= \{ x \in X \mid f(x) - g(x) \neq 0 \} \\ &= (f-g)^{-1}(\{ 0 \}^c) \end{aligned}$$

and we are done since $\{ 0 \}^c$ is measurable in $\mathcal{B}(\mathbb{C})$.

$\sim$ is indeed an equivalence relation:

1. Reflexive: $f \sim f$ since
$$\mu(\{ x \in X \mid f(x) \neq f(x) \}) = \mu(\varnothing) = 0 \, .$$
2. Symmetric: $f \sim g \iff g \sim f$ since $f(x) \neq g(x)$ is symmetric.
3. Transitive: $f \sim g \wedge g \sim h \Longrightarrow f \sim h$. First we note that
$$f(x) \neq g(x) \iff |f(x) - g(x)| > 0 \, .$$
Next, write
$$|f(x) - h(x)| \quad \leq \quad |f(x) - g(x)| + |g(x) - h(x)|$$
so if $x \in X$ is such that $0 < |f(x) - h(x)|$ then one of $|f(x) - g(x)|$ or $|g(x) - h(x)|$ *must* be nonzero. This means that
$$\{ x \in X \mid f(x) \neq h(x) \} \subseteq \{ x \in X \mid f(x) \neq g(x) \} \cup \{ x \in X \mid g(x) \neq h(x) \}$$
and so taking μ of this equation we find
$$\begin{aligned} \mu(\{ x \in X \mid f(x) \neq h(x) \}) &\leq \mu(\{ x \in X \mid f(x) \neq g(x) \} \cup \{ x \in X \mid g(x) \neq h(x) \}) \\ &\leq \mu(\{ x \in X \mid f(x) \neq g(x) \}) + \mu(\{ x \in X \mid g(x) \neq h(x) \}) \\ &= 0 + 0 = 0 \, . \end{aligned}$$
Hence $f \sim h$.

The result is that $\sim$ is an equivalence relation on $L^1(\mu)$. We denote the equivalence classes with

$$[f]_{L^1(\mu)} := \left\{ g \in L^1(\mu) \;\middle|\; f \sim g \right\}$$

and now define

$$\widetilde{L_1(\mu)} := \left\{ [f]_{L^1(\mu)} \;\middle|\; f \in L^1(\mu) \right\} .$$

One easily verifies that is *also* a $\mathbb{C}$-vector space with

$$\begin{aligned} [f]_{L^1(\mu)} + [g]_{L^1(\mu)} &:= [f+g]_{L^1(\mu)} \\ \alpha [f]_{L^1(\mu)} &:= [\alpha f]_{L^1(\mu)} \end{aligned}$$

and the same norm (5.4) now is an honest norm:

$$\left\| [f]_{L^1(\mu)} \right\|_{L^1(\mu)} := \int_X |f| \, \mathrm{d}\mu \, .$$

Proof. First, it is clear this formula is a well-defined function at the level of the equivalence classes. Indeed, if $f \sim g$ then since the two functions differ on a set of measure zero, the integral of their absolute values will agree. Same goes with the proof of homogeneity and the triangle inequality. So we are left with showing that

$$\left\| [f]_{L^1(\mu)} \right\|_{L^1(\mu)} = 0 \Longrightarrow [f]_{L^1(\mu)} = 0$$

i.e., that

$$\int_X |f| \, \mathrm{d}\mu = 0 \Longrightarrow \mu\left(\underbrace{\{ x \in X \mid |f(x)| > 0 \}}_{=:N} \right) = 0 \, .$$

To see this, let us define

$$N_n := \left\{ x \in X \;\middle|\; |f(x)| > \frac{1}{n} \right\} \qquad (n \in \mathbb{N}) .$$

Then

$$\frac{1}{n}\mu(N_n) = \int_{N_n} \frac{1}{n} \mathrm{d}\mu \le \int_{N_n} |f| \,\mathrm{d}\mu \le \int_N |f| \,\mathrm{d}\mu = 0 .$$

Hence $\mu(N_n) = 0$ for all $n \in \mathbb{N}$. But

$$N = \bigcup_{n \in \mathbb{N}} N_n$$

so the claim follows. ☐

It is time to lighten up the notation a little. First, a piece of terminology:

Definition 5.30 (Almost-everywhere)**.** If for two measurable functions $f, g : X \to \mathbb{C}$ we have

$$\mu(\{ x \in X \mid f(x) \neq g(x) \}) = 0$$

we say that $f = g$ μ-almost-everywhere, abbreviated as μ-a.e.. In the context of probability one says μ-almost-surely or just almost surely.

Thus, even though $\|\cdot\|_{L^1}$ is not honestly a norm on $L^1(\mu)$ and strictly speaking one should work with

$$\left\| [\cdot]_{L^1(\mu)} \right\|_{L^1(\mu)}, \qquad \widetilde{L_1(\mu)}$$

with abuse of notation, we avoid this notation and shall use the previous notation, even though whenever we appeal to this normed vector space structure we really mean to talk about *equivalence classes* of L^1 functions which only differ on sets of measure zero.

Another important fact is that the norm (5.4) makes $L^1(\mu)$ *complete*: any Cauchy sequence w.r.t. the norm converges.

Proposition 5.31. *The norm (5.4) is complete.*

Proof. Let $\{ f_n \}_n \subseteq L^1(\mu)$ such that for any $\varepsilon > 0$ there exists some $N_\varepsilon \in \mathbb{N}$ such that if $n, m \in \mathbb{N}$ are such that $n, m \ge N_\varepsilon$ then

$$\|f_n - f_m\|_1 < \varepsilon .$$

We want to show that implies there exists some $f \in L^1(\mu)$ such that $f_n \to f$ in the $L^1(\mu)$ norm. From this Cauchy condition, for any $j \in \mathbb{N}$, if $n, m \ge N_{2^{-j}}$ then

$$\|f_n - f_m\|_1 < 2^{-j} .$$

This allows us to find a strictly increasing sequence $\{ n_j \}_{j \in \mathbb{N}} \subseteq \mathbb{N}$ such that

$$\left\| f_{n_{j+1}} - f_{n_j} \right\| < 2^{-j} \qquad (j \in \mathbb{N}) .$$

Define now, for any $x \in X$ for which it makes sense,

$$f(x) := f_{n_1}(x) + \sum_{j=1}^{\infty} \left[f_{n_{j+1}}(x) - f_{n_j}(x) \right] . \tag{5.5}$$

We claim that this series converges absolutely for almost-every $x \in X$. Indeed, set

$$g_k := \sum_{j=1}^{k} \left| f_{n_{j+1}} - f_{n_j} \right|, \qquad g := \lim_k g_k .$$

By the Minkowski inequality Theorem 5.28,

$$\|g_k\|_1 \leq \sum_{j=1}^{k} 2^{-j} < 1\,.$$

Hence by Fatou's lemma Lemma 2.53,

$$\int_X \left(\liminf_k g_k\right) \mathrm{d}\mu \leq \liminf_k \int_X g_k \mathrm{d}\mu < \liminf_k 1 = 1$$

so that

$$\|g\|_1 \leq 1\,.$$

This implies that $g(x) < \infty$ for μ-almost-every x and we indeed get absolute convergence of (5.5) μ-almost-everywhere. Set $f = 0$ on the measure-zero complement of this set. Since the sum in the definition telescopes, it is clear that wherever the sum does converge,

$$f(x) = \lim_{j\to\infty} f_{n_j}(x)\,.$$

We now want to boost this almost-everywhere pointwise convergence to $L^1(\mu)$ convergence. By Fatou's lemma, for any $m \geq N_\varepsilon$ we have

$$\int_X |f - f_m|\,\mathrm{d}\mu \leq \liminf_{j\to\infty} \int_X \left|f_{n_j} - f_m\right| \mathrm{d}\mu \leq \varepsilon$$

so that $f \in L^1(\mu)$ and $f_m \to f$ in $L^1(\mu)$. □

A complete normed vector space is called a *Banach* space. We have thus exhibited $L^1(\mu)$ as a Banach space.

A basic question one may pose is: does this norm arise from an inner product?

The answer is that this is so if and only if the norm obeys the parallelogram rule.

Claim 5.32. If a norm satisfies the parallelogram law:

$$\|\psi+\varphi\|^2 + \|\psi-\varphi\|^2 \leq 2\|\psi\|^2 + 2\|\varphi\|^2 \qquad (\forall \varphi, \psi \text{ in the normed vector space})$$

then

$$\langle\psi,\varphi\rangle := \frac{1}{4}\left[\|\psi+\varphi\|^2 - \|\psi-\varphi\|^2 + \mathrm{i}\|\mathrm{i}\psi-\varphi\|^2 - \mathrm{i}\|\mathrm{i}\psi+\varphi\|^2\right]$$

defines an inner product whose associated norm is $\|\cdot\| \equiv \sqrt{\langle\cdot,\cdot\rangle}$. Conversely if the parallelogram law is violated then *no* inner-product may be defined compatible with that norm.

Proof. Left as an exercise to the reader. □

Claim 5.33. The L^1 norm does *not* in general satisfy the parallelogram law.

Proof. TODO □

If we seek to work with an inner product then there *is* a space we can work with: the L^2 space. We define

$$L^2(\mu) \equiv \left\{ f : X \to \mathbb{C} \,\middle|\, f \text{ is msrbl. and } \int_X |f|^2 \mathrm{d}\mu < \infty \right\}$$

with the associated norm

$$\|f\|_{L^2(\mu)} := \sqrt{\int_X |f|^2 \mathrm{d}\mu}\,. \tag{5.6}$$

Claim 5.34. The formula in (5.6) induces a complete norm (again with the song and dance about equivalence classes of functions which differ on sets of measure zero) which *does* satisfy the parallelogram law. The associated inner product is

$$\langle f,g\rangle_{L^2(\mu)} := \int_X \overline{f} g \mathrm{d}\mu\,.$$

This makes $L^2(\mu)$ into a *Hilbert space*: an inner product space whose associated norm is *complete*.

Proof. The same completeness proof presented above for $L^1(\mu)$ works for any $p \in [1,\infty)$. ☐

5.5 The Lebesgue decomposition theorem [Rudin]

In this chapter we want to perform the opposite operation as in Theorem 2.54. Recall from there that if $f : X \to [0,\infty]$ is measurable then given a measure $\mu : \mathrm{Msrbl}(X) \to [0,\infty]$ we may induce a new one $\varphi_{\mu,f} : \mathrm{Msrbl}(X) \to [0,\infty]$ via

$$\varphi_{\mu,f}(E) \equiv \int_E f \mathrm{d}\mu \qquad (E \in \mathrm{Msrbl}(X))\,. \tag{5.7}$$

Question: Can we do the opposite? Given two measures μ,φ, does there exist a function f so that (5.7) holds? This is what we want to explore here.

Let $(X,\mathcal{M})$ be a measurable space and $\mu : \mathcal{M} \to [0,\infty]\,, \nu : \mathcal{M} \to \mathbb{C}$ be two measures on it.

Definition 5.35 (Absolute continuity)**.** We say that ν is absolutely continuous w.r.t. μ, and write

$$\nu \blacktriangleleft \mu\,,$$

iff for any $E \in \mathcal{M}$,

$$\mu(E) = 0 \Longrightarrow \nu(E) = 0\,.$$

Said differently,

$$\mu^{-1}(\{\,0\,\}) \subseteq \nu^{-1}(\{\,0\,\})\,.$$

Example 5.36. Let $f : X \to [0,\infty]$ be measurable and $\mu : \mathrm{Msrbl}(X) \to [0,\infty]$ be a measure. Define $\varphi_{\mu,f}$ as in (5.7). Then

$$\varphi_{\mu,f} \blacktriangleleft \mu\,.$$

Proof. Assume that $E \in \mathrm{Msrbl}(X)$ is such that $\mu(E) = 0$. Then we want to show that

$$0 \overset{?}{=} \varphi_{\mu,f}(E) \equiv \int_E f\mathrm{d}\mu = \sup_{s \text{ simple s.t. } 0\leq s\leq f} \int_E s\mathrm{d}\mu = \sup_{s \text{ simple s.t. } 0\leq s\leq f} \sum_i \alpha_i \mu(A_i \cap E) = 0\,.$$

☐

Example 5.37. Let $(X,\mathfrak{M},\mu)$ be a measure space with $\mu : \mathrm{Msrbl}(X) \to [0,\infty]$ and let $\varphi : X \to X$ be a measure-preserving map: $\mu(A) = \mu\left(\varphi^{-1}(A)\right)$ for all $A \in \mathfrak{M}$. Then the push-forward measure μ_φ given in Definition 5.20 has

$$\mu_\varphi \blacktriangleleft \mu\,.$$

Definition 5.38 (Concentration)**.** Let $\nu : \mathfrak{M} \to \mathbb{C}$ be a measure. If $\exists A \in \mathcal{M}$ such that

$$\nu(E) = \nu(A \cap E) \qquad (E \in \mathcal{M})$$

then we say that ν is *concentrated on A*. This condition is equivalent to

$$E \cap A = \varnothing \quad\Longrightarrow\quad \nu(E) = 0 \qquad (E \in \mathcal{M})\,.$$

Definition 5.39 (Mutually singular measures)**.** Let μ, ν be two measures on $(X, \mathcal{M})$ and let $A, B \in \mathcal{M}$ be such that $A \cap B = \varnothing$ and such that μ is concentrated on A and ν is concentrated on B. Then we say that μ and ν are mutually singular and write

$$\mu \perp \nu\,.$$

Proposition 5.40. *Let $(X, \mathcal{M})$ be a measurable space and $\mu, \lambda, \lambda_1, \lambda_2$ be measures on $\mathcal{M}$. Assume further that μ is a positive measure. Then*

1. *If $\lambda_1 \blacktriangleleft \mu$ and $\lambda_2 \perp \mu$ then $\lambda_1 \perp \lambda_2$.*
2. *If $\lambda \blacktriangleleft \mu$ and $\lambda \perp \mu$ then $\lambda = 0$.*

Proof. For the first claim, since $\lambda_2 \perp \mu$, there is some $A \in \mathcal{M}$ such that $\mu(A) = 0$ and λ_2 is concentrated on A. But $\lambda_1 \blacktriangleleft \mu$ so, $\lambda_1(E) = 0$ for all $E \subseteq A$ so that λ_1 is concentrated on A^c.

For the second claim, using the first one we have $\lambda \perp \lambda$ which forces $\lambda = 0$. ☐

Lemma 5.41. *If μ is a positive σ-finite measure on a σ-algebra $\mathcal{M}$ in a set X, then there is a function $w : X \to (0, 1)$ such that $w \in L^1(\mu)$.*

Proof. Since μ is σ-finite, $\exists \{ E_n \}_{n\in\mathbb{N}} \subseteq \mathcal{M}$ such that $X = \bigcup_{n\in\mathbb{N}} E_n$ and $\mu(E_n) < \infty$. Define

$$w_n(x) := \begin{cases} 0 & x \notin E_n \\ \frac{1}{2^n(1+\mu(E_n))} & x \in E_n \end{cases} \qquad (x \in X)$$

and $w := \sum_{n\in\mathbb{N}} w_n$. Then w has the required properties. Indeed, we clearly see that $w \geq 0$, so $|w| = w$. Then,

$$\begin{aligned} \int_X w \mathrm{d}\mu &= \int_X \sum_{n\in\mathbb{N}} w_n \mathrm{d}\mu \\ &= \sum_{n\in\mathbb{N}} \int_X w_n \mathrm{d}\mu && \text{(MCT)} \\ &= \sum_{n\in\mathbb{N}} \frac{\mu(E_n)}{2^n(1+\mu(E_n))} \\ &\leq \sum_{n\in\mathbb{N}} \frac{1}{2^n} \\ &= 1 < \infty\,. \end{aligned}$$

We conclude that $w \in L^1$. Moreover, Since $X = \bigcup_{n\in\mathbb{N}} E_n$, given $x \in X$, there exists some $n_x \in \mathbb{N}$ such that $x \in E_{n_x}$. For such x, we have

$$w(x) \quad \geq \quad w_{n_x}(x) = \frac{1}{2^n(1+\mu(E_n))} > 0\,.$$

Hence $w > 0$. The same argument also shows that $w < 1$. ☐

The existence of this w allows us to construct a new, finite measure out of μ as follows. Using Theorem 2.54 we may define a new measure $\tilde{\mu} : \mathcal{M} \to [0, \infty]$ via

$$\tilde{\mu}(E) := \int_E w \mathrm{d}\mu\,.$$

Clearly we have $\tilde{\mu}(X) < \infty$ and for all $N \in \mathcal{M}$, $\tilde{\mu}(N) = 0$ iff $\mu(N) = 0$, since $w > 0$.

Theorem 5.42 (Lebesgue-Radon-Nikodym decomposition theorem)**.** *Let μ be a positive σ-finite measure on a σ-algebra $\mathcal{M}$ in a set X, and let λ be a (*finite*) complex measure on $\mathcal{M}$. Then*

1. *(*Lebesgue decomposition*) There is a* unique *pair of measures $\lambda_a, \lambda_s : \mathcal{M} \to \mathbb{C}$ such that*

$$\lambda = \lambda_a + \lambda_s, \qquad \lambda_a \blacktriangleleft \mu, \qquad \lambda_s \perp \mu\,.$$

   Moreover, if λ is positive and finite, so are λ_a, λ_s; here by $\lambda_a + \lambda_s$ we mean the pointwise sum of measures, which is a new measure.

2. *(*Radon-Nikodym derivative*) There is a* unique *element of $L^1(\mu)$, denoted as $\frac{\mathrm{d}\lambda_a}{\mathrm{d}\mu}$, such that*

$$\lambda_a(E) \quad = \quad \int_E \frac{\mathrm{d}\lambda_a}{\mathrm{d}\mu}\mathrm{d}\mu \qquad (E \in \mathcal{M})\,.$$

Proof due to von Neumann. *Case 1*: Assume that λ is a positive bounded measure on $\mathcal{M}$ and let w be associated with μ as in Lemma 5.41. Define now a new measure φ via

$$\varphi(E) := \lambda(E) + \underbrace{\int_E w\mathrm{d}\mu}_{=:\tilde{\mu}(E)} \qquad (E \in \mathcal{M})\,.$$

Then φ is also a positive bounded measure on $\mathcal{M}$, and

$$\int_X f\mathrm{d}\varphi = \int_X f\mathrm{d}\lambda + \int_X fw\mathrm{d}\mu$$

holds for any measurable function $f : X \to \mathbb{C}$ of the form $f = \chi_E$ by definition, hence for any simple f and hence by the monotone convergence Theorem 2.47 for any nonnegative measurable f.

Moreover, we have

$$\begin{aligned} \left|\int_X f\mathrm{d}\lambda\right| &\leq \int_X |f|\,\mathrm{d}\lambda \\ &= \int_X |f|\,\mathrm{d}\varphi - \int_X |f|\,w\mathrm{d}\mu \\ &\leq \int_X |f|\,\mathrm{d}\varphi \\ &\leq \sqrt{\int_X |f|^2\,\mathrm{d}\varphi}\sqrt{\varphi(X)} \end{aligned}$$

where in the last inequality we invoked the Cauchy-Schwarz inequality Claim D.6. But we have $\varphi(X) < \infty$ so

$$f \mapsto \int_X f\mathrm{d}\lambda$$

is a bounded linear functional on the Hilbert space $L^2(X \to \mathbb{C}, \varphi)$. Hence by the Riesz representation theorem Theorem D.10 there exists a *unique* $g \in L^2(X \to \mathbb{C}, \varphi)$ such that

$$\int_X f\mathrm{d}\lambda = \langle g, f\rangle_{L^2(X\to\mathbb{C},\varphi)} \equiv \int_X f\overline{g}\mathrm{d}\varphi \qquad \left(f \in L^2(X \to \mathbb{C}, \varphi)\right)\,. \tag{5.8}$$

Note that elements of $L^2(X \to \mathbb{C}, \varphi)$ are only defined up to a set of φ measure zero, so g can only be determined up to that equivalence class. Now let $E \in \mathfrak{M}$ be such that $\varphi(E) > 0$ (there must be such a set or else φ is the zero measure; since both $\lambda, \tilde{\mu} \geq 0$, that implies both of these measures are the zero measure, in which case we are anyway finished). Then plugging in $f = \chi_E$ into (5.8) we get

$$\lambda(E) = \int_E \overline{g} \mathrm{d}\varphi \,.$$

But $0 \leq \lambda \leq \varphi$ as measures, or equivalently, $0 \leq \frac{\lambda}{\varphi} \leq 1$ and hence

$$\begin{aligned} 0 &\leq \frac{\lambda(E)}{\varphi(E)} \leq 1 \\ &\updownarrow \\ 0 &\leq \frac{\int_E \overline{g} \mathrm{d}\varphi}{\varphi(E)} \leq 1 \,. \end{aligned}$$

Right below in Lemma 5.43 we show this implies $\overline{g} \in [0,1]$ φ-almost-everywhere, so we may re-define $\overline{g}$ for this to hold for *every* x while still keeping (5.8) (and thus we can drop the bar $\overline{g} \mapsto g$ since g is anyway real-valued) and hence we rewrite (5.8) as

$$\int_X fg \mathrm{d}\varphi = \int_X fg \mathrm{d}\lambda + \int_X fgw \mathrm{d}\mu$$

$$\updownarrow$$

$$\int_X f \mathrm{d}\lambda = \int_X fg \mathrm{d}\lambda + \int_X fgw \mathrm{d}\mu$$

$$\updownarrow$$

$$\int_X (1-g) f \mathrm{d}\lambda = \int_X fgw \mathrm{d}\mu \,. \tag{5.9}$$

Define now

$$A := g^{-1}([0,1)) \,, \qquad B := g^{-1}(\{1\}) \,,$$

These are measurable sets as g is measurable as a member of L^2. Define two new measures

$$\lambda_a(E) := \lambda(A \cap E) \,, \qquad \lambda_s(E) := \lambda(B \cap E) \qquad (E \in \mathfrak{M}) \,.$$

Insert $f = \chi_B$ into (5.9) to get

$$0 = \int_B w \mathrm{d}\mu \,.$$

But since $w > 0$ for all $x \in X$, we conclude $\mu(B) = 0$. Hence $\lambda_s \perp \mu$.

We may moreover replace $f = \chi_E \sum_{j=0}^n g^j$ for $n \in \mathbb{N}$ and $E \in \mathfrak{M}$ into (5.9); here by g^j we mean the jth power of g. We get then

$$\begin{aligned} \int_E gw \sum_{j=1}^n g^j \mathrm{d}\mu &= \int_X (1-g) \chi_E \sum_{j=0}^n g^j \mathrm{d}\lambda \\ &= \sum_{j=0}^n \int_E (1-g) g^j \mathrm{d}\lambda \\ &= \int_E \sum_{j=0}^n \left(g^j - g^{j+1}\right) \mathrm{d}\lambda \\ &\overset{\text{telescoping}}{=} \int_E \left(1 - g^{n+1}\right) \mathrm{d}\lambda \,. \end{aligned}$$

On B, $g = 1$ so $1 - g^{n+1} = 0$. On A, $g^{n+1} \to 0$ monotonically. Hence the RHS of the above converges to $\lambda(A \cap E) = \lambda_a(E)$ as $n \to \infty$. On the LHS,

$$\left\{ gw \sum_{j=1}^{n} g^j \right\}_n$$

increases monotonically to a non-negative measurable limit, call it h, so by the monotone convergence Theorem 2.47, the RHS converges to $\int_E h \mathrm{d}\mu$ as $n \to \infty$. Hence we have proven

$$\lambda_a(E) \quad = \quad \int_E h \mathrm{d}\mu \qquad (E \in \mathfrak{M}) \,.$$

Taking $E = X$ we find that $h \in L^1(\mu)$ since we assume $\lambda_a(X) \leq \lambda(X) < \infty$. Moreover, this equation shows that $\lambda_a \blacktriangleleft \mu$ and the proof is complete for positive λ.

If λ is a complex measure, write $\lambda = \lambda_1 + \mathrm{i}\lambda_2$ for λ_1, λ_2 real and apply the preceding case to the positive and negative total variations of λ_1 and λ_2 respectively (TODO: cross-ref below).

We proceed to the uniqueness claims: Let $\tilde{\lambda_a}, \tilde{\lambda_s}$ be another pair which satisfies $\lambda = \tilde{\lambda_a} + \tilde{\lambda_s}$. Then

$$\begin{aligned} \tilde{\lambda_a} + \tilde{\lambda_s} &= \lambda_a + \lambda_s \\ &\updownarrow \\ \tilde{\lambda_a} - \lambda_a &= \lambda_s - \tilde{\lambda_s} \,. \end{aligned}$$

But we also know that $\tilde{\lambda_a} - \lambda_a \blacktriangleleft \mu$ and $\lambda_s - \tilde{\lambda_s} \perp \mu$. So it must be that both sides of this equation are zero, via Proposition 5.40. For the uniqueness of h we employ Lemma 5.44 right below. ☐

Proof. If we now relax λ to have range in $[0, \infty]$ instead of $\mathbb{C}$, and be σ-finite, most of the theorem is still true, since we can write $X = \bigcup_n X_n$ with $\mu(X_n) < \infty$ and $\lambda(X_n) < \infty$ and then decompose each

$$\lambda(\cdot \cap X_n)$$

w.r.t. μ. However, it is not longer true that $h \in L^1(\mu)$, although it is “locally in L^1” in the sense that $\int_{X_n} h \mathrm{d}\mu < \infty$ for each n. ☐

Lemma 5.43 (Range of function vs. range of its normalized integral)**.** *Let $\mu : \mathfrak{M} \to [0, \infty)$ be a positive measure and $f \in L^1(X \to \mathbb{C}; \mu)$. Let $F \in \mathrm{Closed}(\mathbb{C})$ such that*

$$\frac{\int_E f \mathrm{d}\mu}{\mu(E)} \in F \qquad (E \in \mathfrak{M} : \mu(E) > 0) \,.$$

Then for μ-almost-all $x \in X$, $f(x) \in F$.

Proof. Assume $F \neq \mathbb{C}$ since otherwise we are finished. Since F^c is open, it contains some ball, say, $B_\varepsilon(z) \subseteq F^c$. It is well-known that F^c is the *countable* union of such balls, say $\{ B_{\varepsilon_n}(z_n) \}_{n \in \mathbb{N}}$:

$$F^c = \bigcup_{n \in \mathbb{N}} B_{\varepsilon_n}(z_n) \,.$$

Let us show that for each $n \in \mathbb{N}$,

$$\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right) = 0 \,.$$

Assume otherwise. Then

$$\frac{\int_{f^{-1}(B_{\varepsilon_n}(z_n))} f \mathrm{d}\mu}{\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right)} \in F \,.$$

And in particular, since $B_{\varepsilon_n}(z_n) \subseteq F^c$,

$$\left|\frac{\int_{f^{-1}(B_{\varepsilon_n}(z_n))} f \mathrm{d}\mu}{\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right)} - z_n\right| > \varepsilon_n\,.$$

But now,

$$\begin{aligned}\left|\frac{\int_{f^{-1}(B_{\varepsilon_n}(z_n))} f \mathrm{d}\mu}{\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right)} - z_n\right| &= \frac{1}{\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right)}\left|\int_{f^{-1}(B_{\varepsilon_n}(z_n))}(f-z_n)\,\mathrm{d}\mu\right| \\ &\leq \frac{1}{\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right)}\int_{f^{-1}(B_{\varepsilon_n}(z_n))}|f-z_n|\,\mathrm{d}\mu \\ &< \frac{1}{\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right)}\int_{f^{-1}(B_{\varepsilon_n}(z_n))}\varepsilon_n \mathrm{d}\mu \\ &= \varepsilon_n\end{aligned}$$

which leads to a contradiction (we have used the fact that on $f^{-1}(B_{\varepsilon_n}(z_n))$, $|f-z_n| < \varepsilon_n$). Hence we reach a contradiction, so it must be that

$$\mu\left(f^{-1}\left(B_{\varepsilon_n}(z_n)\right)\right) = 0\,.$$

But since this is true for *any* $n \in \mathbb{N}$, this is true for all of F^c: $\mu\left(f^{-1}(F^c)\right) = 0$ which is tantamount to saying that $f \in F$ μ-almost-everywhere. □

Lemma 5.44. *Let $f : X \to [0,\infty]$ be measurable and $E \in \mathfrak{M}$ be such that $\int_E f \mathrm{d}\mu = 0$. Then $f = 0$ μ-almost-everywhere on E. If $f \in L^1(X \to \mathbb{C};\mu)$ and $\int_E f = 0$ for all $E \in \mathfrak{M}$ then $f = 0$ μ-almost-everywhere.*

Proof. For the first statement, let $A_n := f^{-1}\left(\left(\frac{1}{n},\infty\right)\right) \cap E$. Then

$$\frac{1}{n}\mu(A_n) \leq \int_{A_n} f \mathrm{d}\mu \leq \int_E f \mathrm{d}\mu = 0$$

by hypothesis. Hence $\mu(A_n) = 0$ for any $n \in \mathbb{N}$. But

$$f^{-1}((0,\infty)) \quad = \quad \bigcup_{n\in\mathbb{N}} A_n\,,$$

so

$$\mu\left(f^{-1}((0,\infty))\right) \leq \sum_{n\in\mathbb{N}}\mu(A_n) = 0\,.$$

Hence $f = 0$ μ-almost-everywhere.

Now assume $f \in L^1(X \to \mathbb{C};\mu)$. Apply the first statement on the measurable set

$$\mathbb{Re}\{f\}^{-1}([0,\infty])$$

to obtain that

$$\int_{\mathbb{Re}\{f\}^{-1}([0,\infty])}\mathbb{Re}\{f\}^+ \mathrm{d}\mu = \mathbb{Re}\left\{\int_{\mathbb{Re}\{f\}^{-1}([0,\infty])} f \mathrm{d}\mu\right\} = \mathbb{Re}\{0\} = 0\,.$$

Hence by the first statement, $\mathbb{Re}\{f\}^+$ is zero μ-almost-everywhere on $\mathbb{Re}\{f\}^{-1}([0,\infty])$. Similarly we deal with the other parts of f to get the result. □

It is customary (at least in mathematical physics) to further decompose the singular part of a measure further into its atomic, *pure point* part and its *singular continuous* part. To that end, let us define, for the measure space $(X,\mathfrak{M},\mu)$ and the measure $\lambda : \mathfrak{M} \to [0,\infty)$ the set

$$X_{\mathrm{pp}} := \{\, x \in X \mid \lambda(\{\, x\,\}) > 0\,\}\,.$$

Note that if X is σ-finite then $|X_{\mathrm{pp}}| \leq \aleph_0$. Then the pure-point measure λ_{pp} is given by

$$\lambda_{\mathrm{pp}} := \lambda(X_{\mathrm{pp}} \cap \cdot)\,.$$

By definition, λ_{pp} is concentrated on X_{pp}. The singular continuous part is whatever remains of λ_{s} after removing λ_{pp}:

$$\lambda_{\text{sc}} := \lambda_{\text{s}} - \lambda_{\text{pp}}\,.$$

Hence by construction, $\lambda_{\text{sc}}(\{\, x \,\}) = 0$ for all $x \in X$.

Example 5.45. Take $f : \mathbb{R} \to [0, \infty]$ given by $x \mapsto \frac{1}{1+x^2}$. Then clearly $f \in L^1(\mathbb{R} \to \mathbb{C}; \lambda)$. Hence

$$\mathcal{B}(\mathbb{R}) \ni A \mapsto \int_A f \mathrm{d}\lambda =: \varphi_{\lambda,f}(A) \in [0, \infty)$$

defines a Borel measure on $\mathbb{R}$, *which is finite*. This measure is *absolutely continuous* w.r.t. λ, and

$$\frac{\mathrm{d}\varphi_{\lambda,f}}{\mathrm{d}\lambda} = f\,.$$

Example 5.46 (Atomic measures)**.** The Dirac delta measure

$$A \mapsto \chi_A(x_0) \equiv \delta_{x_0}(A)$$

is a point mass measure that is mutually singular w.r.t. λ. In fact we can have infinitely many masses, and still have a finite measure, via, e.g.

$$\sum_{n\in\mathbb{N}} 2^{-n}\delta_n =: \mu_{\text{pp}}\,.$$

Then μ_{pp} is mutually singular w.r.t. λ.

Example 5.47 (The Cantor measure)**.** We know that the middle-$\frac{1}{3}$ Cantor set is bijective with $2^{\mathbb{N}}$, because for any $x \in C$, we may represent x unique in ternary as

$$x = 0.a_1a_2a_3\cdots$$

so that $a : \mathbb{N} \to \{\, 0, 2 \,\}$ (this avoids possibly infinitely repeating 1s). Then we define the Cantor function $f_C : [0, 1] \to [0, 1]$ via

$$f_C(x) := \begin{cases} \sum_{n=1}^{\infty} \frac{\frac{1}{2}a_n}{2^n} & x \in C \\ \sup_{y\in C: y\leq x} f_C(y) & x \in [0, 1] \setminus C \end{cases}.$$

We now seek to define a measure μ_C on $\mathcal{B}([0, 1])$ such

$$\mu_C([0, x]) = f_C(x) \qquad (x \in [0, 1])\,.$$

To that end, recall from HW3Q5 the definition of the Lebesgue-Stieltjes measure associated with an increasing right-continuous function (which f_C *is*). Essentially we have

$$\mu_C([a, b]) \equiv f_C(b) - f_C(a)$$

and extended to a premeasure, outermeasure and then measure by the Caratheodory procedure Figure 2.

One can show (see the upcoming HW) that μ_C is concentrated on C, it is mutually singular w.r.t. λ, and that it has *no atoms,* i.e., it is *singular continuous.*

In principle any given measure on $\mathbb{R}$ is the sum of these three basic types:

$$\mu = \mu_{\text{ac}} + \mu_{\text{sc}} + \mu_{\text{pp}}$$

with

$$\mathrm{d}\mu_{\text{ac}} = f\mathrm{d}\lambda$$

for some $f \in L^1(\mathbb{R} \to \mathbb{C}, \lambda)$ and

$$\mu_{\text{pp}} = \sum_{n\in\mathbb{N}} \alpha_n\delta_{x_n}$$

for some $\{\, x_n \,\}_n \subseteq \mathbb{R}$ and $\{\, \alpha_n \,\}_n \subseteq \mathbb{C}$. μ_{sc} is characterized as "anything that remains".

Definition 5.48 (The Stieltjes transform)**.** Let $\mu : \mathscr{B}(\mathbb{R}) \to [0,\infty]$ be a Borel measure on $\mathbb{R}$ and denote by $\mathbb{C}_+$ the open upper half complex plane

$$\mathbb{C}_+ := \{ z \in \mathbb{C} \mid \mathbb{Im}\{z\} > 0 \} .$$

We define the *Stieltjes* transform of μ, $H_\mu : \mathbb{C}_+ \to \mathbb{C}_+$via

$$H_\mu(z) := \int_{x\in\mathbb{R}} \frac{1}{x-z} \mathrm{d}\mu(x) \qquad (z \in \mathbb{C}_+)$$

for all z for which the $x \mapsto \frac{1}{x-z} \in L^1(\mathbb{R}\to\mathbb{C};\mu)$.

Proposition 5.49. *Given a Borel measure $\mu : \mathscr{B}(\mathbb{R}) \to [0,\infty]$ such that $x \mapsto \frac{1}{x^2+1} \in L^1(\mathbb{R}\to\mathbb{C};\mu)$, its Stieltjes transform H_μ is a well-defined analytic function.*

Analytic functions $\mathbb{C}_+ \to \mathbb{C}_+$ are called *Herglotz-Pick-Nevanlinna* functions. There is a representation theorem for all such functions, see [Tes09].

Proof. We first verify that that H_μ is well-defined, i.e., that $H_\mu(z) \in \mathbb{C}_+$ for all $z \in \mathbb{C}_+$:

$$\begin{aligned}
\mathbb{Im}\{H_\mu(z)\} &\equiv \frac{1}{2\mathrm{i}}\left[H_\mu(z) - \overline{H_\mu(z)}\right] \\
&= \frac{1}{2\mathrm{i}}\left[\int_{x\in\mathbb{R}} \frac{1}{x-z}\mathrm{d}\mu(x) - \overline{\int_{x\in\mathbb{R}} \frac{1}{x-z}\mathrm{d}\mu(x)}\right] \\
&= \int_{x\in\mathbb{R}} \frac{1}{2\mathrm{i}}\left[\frac{1}{x-z} - \overline{\frac{1}{x-z}}\right]\mathrm{d}\mu(x) \\
&= \int_{x\in\mathbb{R}} \frac{\mathbb{Im}\{z\}}{(x-\mathbb{Re}\{z\})^2 + \mathbb{Im}\{z\}^2}\mathrm{d}\mu(x) \\
&= \mathbb{Im}\{z\}\int_{x\in\mathbb{R}} \frac{1}{(x-\mathbb{Re}\{z\})^2 + \mathbb{Im}\{z\}^2}\mathrm{d}\mu(x) .
\end{aligned}$$

So $\mathbb{Im}\{H_\mu(z)\} > 0$ indeed. Next, we study analyticity, which follows similarly:

$$\begin{aligned}
H'_\mu(z) &\overset{?}{=} \lim_{w\to z} \frac{H_\mu(w) - H_\mu(z)}{w-z} \\
&= \lim_{w\to z} \frac{\int_{x\in\mathbb{R}} \frac{1}{x-w}\mathrm{d}\mu(x) - \int_{x\in\mathbb{R}} \frac{1}{x-z}\mathrm{d}\mu(x)}{w-z} \\
&= \lim_{w\to z} \int_{x\in\mathbb{R}} \frac{\frac{1}{x-w} - \frac{1}{x-z}}{w-z}\mathrm{d}\mu(x) \\
&= \lim_{w\to z} \int_{x\in\mathbb{R}} \frac{\frac{1}{x-w}(w-z)\frac{1}{x-z}}{w-z}\mathrm{d}\mu(x) \\
&= \lim_{w\to z} \int_{x\in\mathbb{R}} \frac{1}{x-w}\frac{1}{x-z}\mathrm{d}\mu(x) .
\end{aligned}$$

Now, the dominated convergence (2.17) may be invoked. The sequence of measurable functions $\left\{ \frac{1}{x-w}\frac{1}{x-z} \right\}_w$ converges pointwise to $\frac{1}{(x-z)^2}$, and

$$\left|\frac{1}{x-w}\frac{1}{x-z}\right| = \frac{1}{\sqrt{(x-\mathbb{Re}\{z\})^2 + \mathbb{Im}\{z\}^2}\sqrt{(x-\mathbb{Re}\{w\})^2 + \mathbb{Im}\{w\}^2}} .$$

At $x \approx \mathbb{Re}\{z\}$ and $w \approx z$ this expression is bounded from above by

$$\frac{1}{|\mathbb{Im}\{z\}|}$$

whereas at $|x| \to \infty$ and $w \approx z$ it is bounded by $\frac{1}{x^2}$. So there should be some constant C so that for all x and all $|w-z|$ sufficiently small,

$$\left|\frac{1}{x-w}\frac{1}{x-z}\right| \leq C\frac{1}{\mathbb{Im}\left\{z\right\}}\frac{1}{x^2+1}\,.$$

In either case, by our assumption, this is integrable, so we may invoke the dominated convergence theorem and get that $H_\mu'(z)$ exists and equals

$$H_\mu'(z) = \int_{x\in\mathbb{R}} \frac{1}{(x-z)^2}\mathrm{d}\mu(x)\,.$$

□

Definition 5.50 (The support of a measure)**.** Let X be a topological space and $\mu : \mathcal{B}(X) \to [0,\infty]$ a Borel measure. The support of the measure, a subset of X, is defined via

$$\operatorname{supp}(\mu) := \left\{\, x \in X \mid \forall U \in \operatorname{Open}(X) : x \in U, \mu(U) > 0 \,\right\}.$$

Theorem 5.51. *Let a Borel measure $\mu : \mathcal{B}(\mathbb{R}) \to [0,\infty]$ be given such that $x \mapsto \frac{1}{x^2+1} \in L^1(\mathbb{R}\to\mathbb{C};\mu)$ and $H_\mu : \mathbb{C}_+ \to \mathbb{C}_+$ is Stieltjes transform. Let*

$$\mu = \mu_{ac} + \mu_{sc} + \mu_{pp}$$

be the Lebesgue decomposition of μ w.r.t. the Lebesgue measure $\lambda : \mathcal{B}(\mathbb{R}) \to [0,\infty]$. Then there exists some $M \in (0,\infty)$ such that

$$|H_\mu(x+\mathrm{i}\varepsilon)| \leq \frac{M}{\varepsilon}\,.$$

Moreover,

$$\operatorname{supp}(\mu_{ac}) = \left\{\, x \in \mathbb{R} \,\middle|\, \lim_{\varepsilon\to0^+} \mathbb{Im}\left\{H_\mu(x+\mathrm{i}\varepsilon)\right\} \in (0,\infty) \,\right\}$$

and

$$\frac{\mathrm{d}\mu_{ac}}{\mathrm{d}\lambda}(x) = \lim_{\varepsilon\to0^+}\frac{1}{\pi}\mathbb{Im}\left\{H_\mu(x+\mathrm{i}\varepsilon)\right\} \qquad (x\in\mathbb{R})\,.$$

Moreover,

$$\operatorname{supp}(\mu_s) = \left\{\, x \in \mathbb{R} \,\middle|\, \lim_{\varepsilon\to0^+} \mathbb{Im}\left\{H_\mu(x+\mathrm{i}\varepsilon)\right\} = \infty \,\right\}, \qquad \operatorname{supp}(\mu_{pp}) = \left\{\, x \in \mathbb{R} \,\middle|\, \lim_{\varepsilon\to0^+} \varepsilon\,\mathbb{Im}\left\{H_\mu(x+\mathrm{i}\varepsilon)\right\} \in (0,\infty) \,\right\}.$$

Proof. TODO □

5.6 Total variation and complex measures [Rudin]

We remind the reader that in our convention (following Rudin) complex measures $\mu : \mathfrak{M} \to \mathbb{C}$ are by definition *finite* measures. Since they have countable additivity (as they are measures), we must have

$$\mu\left(\sqcup_j A_j\right) = \sum_{j\in\mathbb{N}} \mu(A_j)$$

for any collection $\left\{\, A_j \,\right\}_j \subseteq \mathfrak{M}$ of pairwise disjoint measurable sets. Since μ is a finite measure this implies that $\sum_{j\in\mathbb{N}}\mu(A_j)$ converges to some (finite) complex number. However, the order of terms here was arbitrary, so in principle the series converges for *any rearrangement*.

Claim 5.52. If a series $\sum_j a_j$ of complex numbers $\left\{\, a_j \,\right\}_j \subseteq \mathbb{C}$ converges to the same value for *any rearrangement of its terms* then it actually converges absolutely.

Proof. Suppose for contradiction that $\sum_j |a_j| = \infty$. Then

$$\infty = \sum_j |a_j| \leq \sum_j |\mathbb{Re}\{a_j\}| + \sum_j |\mathbb{Im}\{a_j\}|$$

so that either $\{ \mathbb{Re}\{a_j\} \}_j$ or $\{ \mathbb{Im}\{a_j\} \}_j$ defines a conditionally convergent series of real numbers. But then the Riemann series theorem implies there is a rearrangement $\sigma : \mathbb{N} \to \mathbb{N}$such that, say, $\sum_j \mathbb{Re}\{a_{\sigma(j)}\}$ converges to something other than $\sum_j \mathbb{Re}\{a_j\}$. But that would contradict

$$\sum_j \mathbb{Re}\{a_{\sigma(j)}\} + \mathrm{i} \sum_j \mathbb{Im}\{a_j\} = \sum_j a_j \, .$$

□

As a result, we get the important fact about complex measures which is that the series associated to their countable additivity converges *absolutely*, i.e., for any any $\{ A_j \}_j \subseteq \mathfrak{M}$ pairwise disjoint,

$$\sum_j |\mu(A_j)|$$

converges. This suggests that there might be a *positive measure* $|\mu|$ associated to μ which dominated μ in the sense that

$$|\mu(A)| \leq |\mu|(A) \qquad (A \in \mathfrak{M}) \, .$$

It turns out that defining

$$|\mu|(A) := |\mu(A)|$$

will *not* yield a measure (clearly, it will violate additivity). Instead, what works is:

Definition 5.53 (Total variation measure)**.** Let $\mu : \mathfrak{M} \to \mathbb{C}$ be a measure. Let us define a new "measure" (putatively) $|\mu| : \mathfrak{M} \to [0, \infty)$ called the total variation measure via

$$|\mu|(A) := \sup_{\{ A_j \}_{j \in \mathbb{N}}} \sum_{j \in \mathbb{N}} |\mu(A_j)| \qquad (A \in \mathfrak{M})$$

where the supremum is taken over all collections $\{ A_j \}_{j \in \mathbb{N}} \subseteq \mathfrak{M}$ such that $A_i \cap A_j = \varnothing$ for all $i \neq j$ and such that $\bigcup_j A_j = A$. Such collections are called *partitions of* A.

Note that even though μ only takes on finite values (as it is a complex measure), it is not a-priori clear that $|\mu|$, even if it *is* a measure, is a finite measure (but it will turn out that it is indeed). This is because one could perhaps construct a sequence of partitions of X, where, even though each sum of each partition is finite, increases to infinity. We will exclude that possibility.

Theorem 5.54. *Let $\mu : \mathfrak{M} \to \mathbb{C}$ be a measure. Then its total variation measure $|\mu| : \mathfrak{M} \to [0, \infty)$ is indeed a measure. It satisfies:*

1. *$|\mu|(A) \geq |\mu(A)|$ for any $A \in \mathfrak{M}$.*
2. *$|\mu|$ is minimal in the sense that for any other positive measure $\nu : \mathfrak{M} \to [0, \infty)$, if $\nu(A) \geq |\mu(A)|$ for all $A \in \mathfrak{M}$ then $\nu \geq |\mu|$.*
3. *If μ is a positive measure then $|\mu| = \mu$.*
4. *$|\mu|(X) < \infty$ indeed.*

Proof. TODO □

Definition 5.55 (The $\mathbb{C}$-vector space of complex measures)**.** If $\mu, \nu : \mathfrak{M} \to \mathbb{C}$ are two complex measures, we define $\mu + \nu$ pointwise:

$$(\mu + \nu)(A) := \mu(A) + \nu(A) \qquad (A \in \mathfrak{M}) .$$

This is indeed a complex measure. Moreover, for any $\alpha \in \mathbb{C}$, we define $\alpha\mu$ also pointwise as

$$(\alpha\mu)(A) := \alpha\mu(A) \qquad (A \in \mathfrak{M}) .$$

This furnishes the set of all complex measures on $\mathfrak{M}$ as a $\mathbb{C}$-vector-space. It is in fact normed because

$$\|\mu\| := |\mu|(X)$$

is indeed a norm.

Theorem 5.56. *In fact $\mu \mapsto |\mu|(X)$ is* complete *so the $\mathbb{C}$-vector space of complex measures on X is a Banach space.*

Proof. TODO □

Lemma 5.57. *Given any complex measure $\mu : \mathfrak{M} \to \mathbb{C}$, its real and imaginary parts $\mathbb{Re}\{\mu\}, \mathbb{Im}\{\mu\} : \mathfrak{M} \to \mathbb{R}$ are also (complex) measures.*

Proof. Since μ is a measure, we have

$$\mathbb{Re}\{\mu\}(\varnothing) \equiv \mathbb{Re}\{\mu(\varnothing)\} = \mathbb{Re}\{0\} = 0 ;$$

similarly for the imaginary part. Next, to show countable additivity, let $\{ A_i \}_i \subseteq \mathfrak{M}$ be a sequence of pairwise disjoint sets. Then

$$\mathbb{Re}\{\mu\}(\sqcup_j A_j) \equiv \mathbb{Re}\{\mu(\sqcup_j A_j)\} = \mathbb{Re}\left\{\sum_j \mu(A_j)\right\} = \sum_j \mathbb{Re}\{\mu(A_j)\} \equiv \sum_j \mathbb{Re}\{\mu\}(A_j) ;$$

similarly for the imaginary part. □

Definition 5.58 (Jordan decomposition)**.** For any complex measure, using the above lemma, we may decompose it as

$$\mu \quad = \quad \mathbb{Re}\{\mu\} + \mathrm{i}\,\mathbb{Im}\{\mu\} .$$

Moreover, for any *real* finite measure $\nu : \mathfrak{M} \to \mathbb{R}$, we may decompose it as

$$\nu \quad = \quad \nu^+ - \nu^- \tag{5.10}$$

where

$$\nu^\pm := \frac{1}{2}(|\nu| \pm \nu) . \tag{5.11}$$

The decomposition Definition 5.58 with the choice (5.11) is known as the *Jordan decomposition* of a real measure. Clearly $\nu^\pm$ are also finite measures.

Lemma 5.59. *If $\nu : \mathfrak{M} \to \mathbb{R}$ is a measure then $\nu^\pm$ from the Jordan decomposition are both finite positive measures.*

Proof. The fact $\nu^{\pm}$ are finite measures is clear since we have proven in Theorem 5.54 that $|\mu|$ is a finite positive measure. We are left to verify that $\nu^{\pm}$ are *positive* measures. To that end, let $A \in \mathfrak{M}$. We want to show that

$$|\nu|(A) \geq \nu(A) .$$

To that end, we use the definition

$$|\nu|(A) \equiv \sup_{\{A_j\}_{j\in\mathbb{N}}} \sum_{j\in\mathbb{N}} |\nu(A_j)| .$$

Let $\{A_j\}_j$ be *any* partition of A. Then

$$\nu(A) \leq |\nu(A)| = \left|\sum_j \nu(A_j)\right| \leq \sum_j |\nu(A_j)| .$$

Taking now supremum of the above inequality over all partitions we obtain the desired result. ☐

Proposition 5.60. *Let $\mu, \nu : \mathfrak{M} \to \mathbb{C}$ be a measure.*

1. *If μ is concentrated on $A \in \mathfrak{M}$, so is $|\mu|$.*
2. *If $\mu \perp \nu$ then $|\mu| \perp |\nu|$.*
3. *If $\nu \blacktriangleleft \mu$ and μ is positive then $|\nu| \blacktriangleleft \mu$.*

Proof. TODO ☐

Theorem 5.61 (Yet another characterization of absolute continuity)**.** *If $\mu : \mathfrak{M} \to [0, \infty]$ and $\nu : \mathfrak{M} \to \mathbb{C}$ are two measures then $\nu \blacktriangleleft \mu$ iff*

$$\forall \varepsilon > 0 \exists \delta > 0 : \text{ If } A \in \mathfrak{M} : \mu(A) < \delta \text{ then } |\nu(A)| < \varepsilon . \tag{5.12}$$

Proof. Assume (5.12) holds. Assume that $A \in \mathfrak{M}$ is such that $\mu(A) = 0$. We want to show that $\nu(A) = 0$. But (5.12) implies that $|\nu(A)| < \varepsilon$ for any $\varepsilon > 0$, and hence it is zero as needed.

Conversely, assume $\nu \blacktriangleleft \mu$ but somehow (5.12) were false. Then $\exists \varepsilon > 0$ and some sequence $\{A_n\}_n \subseteq \mathfrak{M}$ with $\mu(A_n) < 2^{-n}$ and yet

$$|\nu(A_n)| \geq \varepsilon .$$

This implies that $|\nu|(A_n) \geq \varepsilon$ since the total variation measure dominates $|\nu(A_n)|$. Now

$$\mu\left(\bigcup_{i=n}^{\infty} A_i\right) \leq \sum_{i=n}^{\infty} \mu(A_i) \leq \sum_{i=n}^{\infty} 2^{-i} = 2^{-n+1}$$

so $\{\bigcup_{i=n}^{\infty} A_i\}_n$ is a decreasing sequence where at least one element is finite. Hence by (2.5) we have

$$\mu\left(\bigcap_{n\in\mathbb{N}} \bigcup_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty} \mu\left(\bigcup_{i=n}^{\infty} A_i\right) \leq 0$$

and moreover

$$|\nu|\left(\bigcap_{n\in\mathbb{N}} \bigcup_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty} |\nu|\left(\bigcup_{i=n}^{\infty} A_i\right) \geq \varepsilon > 0 .$$

So $\nu \blacktriangleleft \mu$ is violated. ☐

Theorem 5.62. *Let $\mu : \mathfrak{M} \to \mathbb{C}$ be a complex measure. Then $\mu \blacktriangleleft |\mu|$ and*

$$\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|} \in L^1(X \to \mathbb{C}, |\mu|)$$

actually takes values within $\mathbb{S}^1 \equiv \{ z \in \mathbb{C} \mid |z| = 1 \}$.

Proof. TODO □

Theorem 5.63. *Let $\mu : \mathfrak{M} \to [0, \infty]$ be a σ-finite positive measure and $\nu : \mathfrak{M} \to \mathbb{C}$ another measure such that $\nu \blacktriangleleft \mu$. Then $|\nu| \blacktriangleleft \mu$ too and*

$$\frac{\mathrm{d}|\nu|}{\mathrm{d}\mu} = \left|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right| .$$

Proof. TODO □

Theorem 5.64 (Hahn decomposition theorem)**.** *Let $\mu : \mathfrak{M} \to \mathbb{R}$ be a measure. Then there exist sets $A^\pm \in \mathfrak{M}$ such that*

$$A^+ \sqcup A^- = X ,$$

$A^+ \cap A^- = \varnothing$ and such that

$$\mu^+(E) = \mu\left(A^+ \cap E\right) , \qquad \mu^-(E) = -\mu\left(A^- \cap E\right) \qquad (E \in \mathfrak{M}) .$$

Proof. By Theorem 5.62, we have $\mu \blacktriangleleft |\mu|$, $\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|} \in L^1(|\mu|)$ and

$$\left|\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}(x)\right| = 1 \qquad (x \in X) .$$

Since μ is real, $\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}$ may be chosen to take real values (by Lemma 5.43) (first $|\mu|$-a.e., then redefine it to actually be so on a set of measure zero). Thus,

$$\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}(x) = \pm 1 \qquad (x \in X) .$$

Set

$$A^\pm := \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}^{-1}(\{\pm 1\}) .$$

Then,

$$\begin{aligned}
\mu^+(E) &= \frac{1}{2}(|\mu|(E) + \mu(E)) \\
&= \frac{1}{2}\left(\int_E \mathrm{d}|\mu| + \int_E \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\mathrm{d}|\mu|\right) \\
&= \int_E \frac{1}{2}\left(1 + \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\right)\mathrm{d}|\mu| \\
&= \int_{E\cap A^+} \underbrace{\frac{1}{2}\left(1 + \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\right)}_{=\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}}\mathrm{d}|\mu| + \underbrace{\int_{E\cap A^-} \frac{1}{2}\left(1 + \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\right)\mathrm{d}|\mu|}_{=0} \\
&= \int_{E\cap A^+} \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\mathrm{d}|\mu| \\
&= \mu\left(E \cap A^+\right) .
\end{aligned}$$

The other part follows similarly. □

Corollary 5.65. *If $\mu : \mathfrak{M} \to \mathbb{R}$ is a measure and $\mu = \lambda_1 - \lambda_2$ for two non-negative measures $\lambda_1, \lambda_2 : \mathfrak{M} \to [0,\infty)$ then necessarily*

$$\lambda_1 \geq \mu^+ , \qquad \lambda_2 \geq \mu^- .$$

Proof. We have $\mu \leq \lambda_1$ so

$$\mu^+ (E) \equiv \mu \left(E \cap A^+ \right) \leq \lambda_1 \left(E \cap A^+ \right) \leq \lambda_1 (E) \qquad (E \in \mathfrak{M}) .$$

The other equation follows similarly, from the fact that $-\mu \leq \lambda_2$. □

5.6.1 Integration with respect to complex measures

Let $\mu : \mathfrak{M} \to \mathbb{C}$ be a complex measure. Then we know now we may write

$$\mu = \mathbb{Re}\{\mu\} + \mathrm{i}\,\mathbb{Im}\{\mu\}$$

and each of which we can further decompose into positive and negative parts. Then we *define,* for any $f : X \to \mathbb{C}$ measurable,

$$\int_X f \mathrm{d}\mu \equiv \int_X f \mathrm{d}\mathbb{Re}\{\mu\}^+ - \int_X f \mathrm{d}\mathbb{Re}\{\mu\}^- + \mathrm{i} \int_X f \mathrm{d}\mathbb{Im}\{\mu\}^+ - \mathrm{i} \int_X f \mathrm{d}\mathbb{Im}\{\mu\}^- .$$

Each of these integrals is in turn further decomposed since f itself is written as the sum of four positive functions.

6 Differentiation of measures on $\mathbb{R}^n$ [Rudin]

In this chapter we are back to the special case of the Lebesgue measure $\lambda : \mathcal{B}(\mathbb{R}^n) \to [0,\infty]$ (although some of the definitions and theorems below may be generalized to any reference positive Borel measure on $\mathbb{R}^n$, see e.g. [Jak06]).

6.1 The Lebesgue differentiation theorem

While we are motivated on our goal to prove the change of variables formula in $\mathbb{R}^n$, another motivation to study the Lebesgue differentiation theorem is to establish the analog of the fundamental theorem of calculus in the context of the Lebesgue integral. The question is, if $f : \mathbb{R} \to \mathbb{C}$ is in L^1, then

$$\partial \int_a^{\cdot} f \mathrm{d}\lambda \overset{?}{=} f .$$

To answer this question we examine the definition of the derivative:

$$\begin{aligned} \left(\partial \int_a^{\cdot} f \mathrm{d}\lambda \right) (x) &\equiv \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \left[\int_a^{x+\varepsilon} f \mathrm{d}\lambda - \int_a^x f \mathrm{d}\lambda \right] \\ &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \int_x^{x+\varepsilon} f \mathrm{d}\lambda . \end{aligned}$$

Clearly if f is continuous at x then we get the result we are looking for. Indeed,

$$\frac{1}{\varepsilon} \int_x^{x+\varepsilon} f \mathrm{d}\lambda = \frac{1}{\varepsilon} \int_x^{x+\varepsilon} [f(y) - f(x)] \mathrm{d}\lambda(y) + f(x)$$

and

$$\left| \frac{1}{\varepsilon} \int_x^{x+\varepsilon} [f(y) - f(x)] \mathrm{d}\lambda(y) \right| \leq \frac{1}{\varepsilon} \int_x^{x+\varepsilon} |f(y) - f(x)| \mathrm{d}\lambda(y) .$$

Now the continuity of f at x is tantamount to the fact that for any $\tilde{\varepsilon} > 0$ there exists some $\tilde{\delta}(\tilde{\varepsilon}) > 0$ such that if $y \in B_{\tilde{\delta}(\tilde{\varepsilon})}(x)$ then $f(y) \in B_{\tilde{\varepsilon}}(f(x))$. Since $\varepsilon \to 0^+$, for fixed $\tilde{\varepsilon} > 0$, eventually we will have $\varepsilon < \tilde{\delta}(\tilde{\varepsilon})$. Thus, for all $\varepsilon \in \left(0, \tilde{\delta}(\tilde{\varepsilon})\right)$ we get

$$\left| \frac{1}{\varepsilon} \int_x^{x+\varepsilon} [f(y) - f(x)] \, \mathrm{d}\lambda(y) \right| \leq \tilde{\varepsilon} \,.$$

But since $\tilde{\varepsilon} > 0$ was arbitrary we get the result.

More generally, there is a vast gap between continuous functions and L^1 functions, and it is not clear how to proceed for L^1 functions. Motivated by Theorem 2.54, $f \mathrm{d}\lambda$ defines a new measure $\varphi_{\lambda,f}$ so we are essentially asking

$$\lim_{\varepsilon \to 0^+} \frac{\varphi_{\lambda,f}([x, x+\varepsilon])}{\lambda([x, x+\varepsilon])} \stackrel{?}{=} f(x) \,.$$

Motivated by this we make the

Definition 6.1 (The symmetric derivative)**.** Let $\mu : \mathcal{B}(\mathbb{R}^n) \to \mathbb{C}$ be a complex measure. Then *the symmetric derivative of μ at $x \in \mathbb{R}^n$ w.r.t.* λ, is given by

$$(\mathcal{D}_\lambda \mu)(x) := \lim_{\varepsilon \to 0^+} \frac{\mu(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))}$$

for all points $x \in \mathbb{R}^n$ at which this limit exists. Here

$$B_\varepsilon(x) \equiv \{ y \in \mathbb{R}^n \mid \|x - y\| < \varepsilon \} \,.$$

We also define the *Hardy-Littlewood maximal function* $\mathcal{M}_\lambda \mu : \mathbb{R}^n \to [0, \infty]$

$$(\mathcal{M}_\lambda \mu)(x) := \sup_{\varepsilon > 0} \frac{|\mu|(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} \qquad (x \in \mathbb{R}^n) \,.$$

In the special case of measures $\varphi_{\lambda,f}$ which are derived from functions f, we use the shorthand notation

$$\mathcal{M}_\lambda f \equiv \mathcal{M}_\lambda \varphi_{\lambda,f} \,.$$

Note that in $\mathbb{R}^n$,

$$\lambda(B_\varepsilon(x)) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} \varepsilon^n$$

where Γ is the Euler gamma function. Since the RHS is rather complicated and not more transparent than the LHS (except that it gives the explicit scaling as $\varepsilon \to 0$) we will usually continue to use the LHS.

Our main motivation in studying the symmetric derivative is that if $\mu \blacktriangleleft \lambda$ then we can calculate its Radon-Nikodym derivative via the symmetric derivative. Indeed, we will see that in this case,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = \mathcal{D}\mu \,.$$

To get to that statement we build some machinery.

Claim 6.2. If $\mu : \mathcal{B}(\mathbb{R}^n) \to \mathbb{C}$ is a complex measure then $\mathcal{M}_\lambda \mu$ is lower semicontinuous.

Proof. Let $a \in \mathbb{R}$. Lower semicontinuity would be implied if we show that the following set is open:

$$E := \{ x \in \mathbb{R}^n \mid (\mathcal{M}_\lambda \mu)(x) > a \} \in \mathrm{Open}(\mathbb{R}^n) \,.$$

Let $x \in E$. Then

$$\sup_{\varepsilon > 0} \frac{|\mu|(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} \quad > \quad a \,.$$

So there must exist some $\varepsilon > 0$ such that there exists some $b > a$ with which

$$\frac{|\mu|(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} = b \,.$$

Since $\frac{b}{a}>1$, pick some $\delta\in\left(0,\varepsilon\left(\left(\frac{b}{a}\right)^{\frac{1}{n}}-1\right)\right)$, so that $(\varepsilon+\delta)^n<\frac{b}{a}\varepsilon^n$. Now, if $y\in B_\delta(x)$, $B_{\varepsilon+\delta}(y)\supseteq B_\varepsilon(x)$ so that

$$|\mu|\left(B_{\varepsilon+\delta}(y)\right)\geq|\mu|\left(B_\varepsilon(x)\right)=b\lambda\left(B_\varepsilon(x)\right)=b\left(\frac{\varepsilon}{\varepsilon+\delta}\right)^n\lambda\left(B_{\varepsilon+\delta}(y)\right)>a\lambda\left(B_{\varepsilon+\delta}(y)\right)$$

which implies

$$\frac{|\mu|\left(B_{\varepsilon+\delta}(y)\right)}{\lambda\left(B_{\varepsilon+\delta}(y)\right)}>a\Longrightarrow\left(\mathcal{M}_\lambda\mu\right)(y)>a\Longrightarrow y\in E\,.$$

Hence we have established that $B_\delta(x)\subseteq E$ and hence $E\in\operatorname{Open}(\mathbb{R}^n)$ as needed. ☐

Corollary 6.3. *The Hardy-Littlewood maximal function $\mathcal{M}_\lambda\mu$ is measurable, since every lower semicontinuous function is.*

Example 6.4. We list a few examples of the Hardy-Littlewood maximal function for various measures:

- Clearly $\mathcal{M}_\lambda\lambda$ equals the constant function 1.
- Consider the Dirac delta measure δ_{x_0} for some $x_0\in\mathbb{R}^n$. Then $|\delta_{x_0}|=\delta_{x_0}=\chi_\cdot(x_0)$. As such,

$$\frac{|\mu|\left(B_\varepsilon(x)\right)}{\lambda\left(B_\varepsilon(x)\right)}=\frac{\chi_{B_\varepsilon(x)}(x_0)}{\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}\varepsilon^n}\,.$$

  We see that for any given $x\neq x_0$, if $\varepsilon<\|x-x_0\|$ then we get zero, and if $\varepsilon\geq\|x-x_0\|$ then the function starts decreasing from its maximal value of

$$\frac{1}{\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}\|x-x_0\|^n}$$

  to zero as $\varepsilon\to\infty$. Conversely, if $x=x_0$ then any ball contains x_0 and so we can shrink $\varepsilon\to0$ and get ∞. Hence

$$\left(\mathcal{M}_\lambda\delta_{x_0}\right)(x)=\begin{cases}\frac{1}{\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}\|x-x_0\|^n} & x\neq x_0\\ \infty & x=x_0\end{cases}\sim\text{decays like }\frac{1}{\|x-x_0\|^n}\text{ away from }x_0\,.$$

  Compare this with

$$\left(\mathcal{D}_\lambda\delta_{x_0}\right)(x)=\begin{cases}0 & x\neq x_0\\ \infty & x=x_0\end{cases}\,.$$

- Let $f:\mathbb{R}^n\to\mathbb{C}$ be continuous. Then as we saw,

$$\left(\mathcal{D}_\lambda\varphi_{\lambda,f}\right)(x)=f(x)\,.$$

  On the other hand, for the Hardy-Littlewood maximal function, we always have

$$\left(\mathcal{M}_\lambda f\right)(x)\geq|f(x)|\qquad(x\in\mathbb{R}^n)\,.$$

  However, $(\mathcal{M}_\lambda f)(x)$ may exceed $|f(x)|$ if somehow the average over a point exceeds the value at that point, for *some* ball. For example, fix some $\varepsilon>0$. Then define

$$f(x):=\begin{cases}\frac{\|x\|}{\varepsilon} & \|x\|\leq\varepsilon\\ 1 & \|x\|>\varepsilon\end{cases}\qquad(x\in\mathbb{R}^n)$$

  which is clearly continuous. Now $f(0)=0$ but if we average over balls of radius larger than ε, we get the value 1 for the function and the ε ball matters less and less, so that the supremum yields the value 1. Hence

$$\left(\mathcal{M}_\lambda f\right)(0)=1\neq0=|f(0)|\,.$$

- TODO: present example of measure μ where $\left(\mathcal{D}_\lambda\mu\right)(x)$ does not exist for some x.

Lemma 6.5 (Vitali's covering)**.** *Let* $\{ x_i \}_{i=1}^N \subseteq \mathbb{R}^n$ *and* $\{ r_i \}_{i=1}^N \subseteq (0,\infty)$*. We use the abbreviation* $B_i := B_{r_i}(x_i)$ *and* $3B_i := B_{3r_i}(x_i)$ *for* $i=1,\cdots,N$*. Then there exists some* $S \subseteq \{ 1,\cdots,N \}$ *such that:*

1. $B_i \cap B_j = \varnothing$ *if* $i,j\in S$ *and* $i\neq j$*.*
2. $\bigcup_{i=1}^N B_i \subseteq \bigcup_{i\in S} 3B_i$*.*
3. $\lambda\left(\bigcup_{i=1}^N B_i\right) \leq 3^n \sum_{i\in S}\lambda(B_i)$*.*

Proof. Without loss of generality assume $r_i \geq r_j$ for all $i<j\in\{ 1,\cdots,N \}$. We define S by including it in members of $\{ 1,\cdots,N \}$ in descending order according to the following rule: $1\in S$. The next element in S should be the smallest index j after 1 so that $B_j\cap B_1=\varnothing$. Continue in this fashion until there are no more indices left. The resulting collection S is clearly pairwise disjoint. Moreover, $\bigcup_{i=1}^N B_i\subseteq\bigcup_{i\in S}3B_i$ necessarily holds. Indeed, let $i\in\{ 1,\cdots,N \}\setminus S$. That means there exists some $j\in S$ with $j<i$ such that $B_i\cap B_j\neq\varnothing$. But since the radii are ordered, necessarily $r_i\leq r_j$, so in the worst case scenario, $B_{3r_j}(x_j)\supseteq B_{r_i}(x_i)$ which is what we needed. ☐

Theorem 6.6. *Let* $\mu:\mathcal{B}(\mathbb{R}^n)\to\mathbb{C}$ *be a measure and* $a>0$*. Then*

$$\lambda(\{ x\in\mathbb{R}^n \mid (\mathcal{M}_\lambda\mu)(x)>a \}) \leq 3^n\frac{\|\mu\|}{a}\,.$$

Proof. As we have seen in the proof of Claim 6.2, the set $E_a := \{ x\in\mathbb{R}^n \mid (\mathcal{M}_\lambda\mu)(x)>a \}$ is open. We know that the Lebesgue measure λ is regular (see Definition 3.1), so we have

$$\lambda(E_a) = \sup(\{ \lambda(K) : \text{Compact}(\mathbb{R}^n)\ni K\subseteq E_a \})\,.$$

Hence, let a compact $K\subseteq E_a$ be given. For any $x\in K$, by definition, since $K\subseteq E_a$, $(\mathcal{M}_\lambda\mu)(x)>a$, there exists some $\varepsilon_x>0$ such that

$$\frac{|\mu|(B_{\varepsilon_x}(x))}{\lambda(B_{\varepsilon_x}(x))}>a\,.$$

Since K is compact, for the open cover

$$\bigcup_{x\in K}B_{\varepsilon_x}(x)\supseteq K$$

there exists a finite sub-cover $\{ x_1,\cdots,x_N \}\subseteq K$:

$$\bigcup_{j=1}^N B_{\varepsilon_{x_j}}(x_j)\supseteq K\,.$$

Now using the Vitali covering Lemma 6.5, we get a subcollection $S\subseteq\{ 1,\cdots,N \}$ of pairwise disjoint balls whose three-fold inflation covers the original union. Thus,

$$\lambda(K)\leq\lambda\left(\bigcup_{j=1}^N B_{\varepsilon_{x_j}}(x_j)\right)\leq 3^n\sum_{j\in S}\lambda\left(B_{\varepsilon_{x_j}}(x_j)\right)\leq 3^n\sum_{j\in S}\frac{1}{a}|\mu|\left(B_{\varepsilon_{x_j}}(x_j)\right)=3^n\frac{1}{a}|\mu|\left(\sqcup_{j\in S}B_{\varepsilon_{x_j}}(x_j)\right)\leq 3^n\frac{1}{a}|\mu|(\mathbb{R}^n)\,.$$

We note that in the equality used here, we invoked the pairwise disjoint property of the collection S (so that we could invoke additivity rather than subadditivity in the opposite direction); taking now supremum over K we obtain the desired result. ☐

Corollary 6.7 (Hardy-Littlewood maximal inequality)**.** *Let* $f\in L^1(\mathbb{R}^n\to\mathbb{C},\lambda)$*. Then for any* $a>0$*,*

$$\lambda(\{ x\in\mathbb{R}^n \mid (\mathcal{M}_\lambda f)(x)>a \})\leq\frac{3^n}{a}\|f\|_{L^1}\,.$$

This corollary prompts us to define a new space of functions which generalizes the L^1 functions:

Definition 6.8 (Weak L^1)**.** We define a new space of functions,

$$L^1_{\text{weak}}\left(\mathbb{R}^n\to\mathbb{C},\lambda\right):=\left\{ f:\mathbb{R}^n\to\mathbb{C}\;\middle|\;\sup_{a>0}a\lambda\left(\left\{ x\in\mathbb{R}^n\mid\left|f\left(x\right)\right|>a\right\}\right)<\infty\right\} .$$

Claim 6.9. We have

$$L^1\left(\mathbb{R}^n\to\mathbb{C},\lambda\right)\subsetneq L^1_{\text{weak}}\left(\mathbb{R}^n\to\mathbb{C},\lambda\right).$$

Proof. Let $f\in L^1\left(\mathbb{R}^n\to\mathbb{C},\lambda\right)$. Then for any $a>0$,

$$\begin{aligned} a\lambda\left(\left\{ x\in\mathbb{R}^n\mid\left|f\left(x\right)\right|>a\right\}\right) & \leq \int_{\left\{ x\in\mathbb{R}^n\mid\left|f\left(x\right)\right|>a\right\}}\left|f\right|\mathrm{d}\lambda\\ & \leq \int_{\mathbb{R}^n}\left|f\right|\mathrm{d}\lambda\\ & \equiv \left\Vert f\right\Vert_{L^1}\\ & < \infty. \end{aligned}$$

To show that the inclusion is strict, consider the function

$$(0,1)\ni x\mapsto\frac{1}{x}.$$

That function is *not* $L^1\left(\left(0,1\right)\to\mathbb{C},\lambda\right)$ but it is in $L^1_{\text{weak}}\left(\left(0,1\right)\to\mathbb{C},\lambda\right)$. Indeed,

$$a\lambda\left(\left\{ x\in\left(0,1\right)\;\middle|\;\frac{1}{x}>a\right\}\right)=a\frac{1}{a}=1<\infty.$$

□

Remark 6.10. Putting everything together, we find that the Hardy-Littlewood maximal function can be interpreted as a map

$$\mathcal{M}_\lambda:L^1\left(\mathbb{R}^n\to\mathbb{C},\lambda\right)\to L^1_{\text{weak}}\left(\mathbb{R}^n\to\mathbb{C},\lambda\right).$$

Definition 6.11 (Lebesgue points)**.** Let $f\in L^1\left(\mathbb{R}^n\to\mathbb{C},\lambda\right)$. Then $x\in\mathbb{R}^n$ is called a Lebesgue point of f iff

$$\lim_{\varepsilon\to0^+}\frac{1}{\lambda\left(B_\varepsilon\left(0\right)\right)}\int_{B_\varepsilon\left(x\right)}\left|f\left(y\right)-f\left(x\right)\right|\mathrm{d}\lambda\left(y\right)=0.$$

In particular, at such points,

$$\lim_{\varepsilon\to0^+}\frac{1}{\lambda\left(B_\varepsilon\left(0\right)\right)}\int_{B_\varepsilon\left(x\right)}f\left(y\right)\mathrm{d}\lambda\left(y\right)=f\left(x\right).$$

Hence, going back to our motivating question from the beginning of this section, we give a special name to those points where the function equals its infinitesimal average. By that discussion, all points at which f is continuous are Lebesgue points.

Theorem 6.12 (Lebesgue differentiation theorem)**.** *If $f\in L^1\left(\mathbb{R}^n\to\mathbb{C},\lambda\right)$ then λ-almost-all points of $\mathbb{R}^n$ are Lebesgue points. I.e., for almost all points of $\mathbb{R}^n$ we have*

$$\lim_{\varepsilon\to0^+}\frac{1}{\lambda\left(B_\varepsilon\left(x\right)\right)}\int_{B_\varepsilon\left(x\right)}f\mathrm{d}\lambda=f\left(x\right).$$

Proof. Define for $r > 0$,

$$\left(T_r f\right)(x) := \frac{1}{\lambda\left(B_r(0)\right)} \int_{B_r(x)} \left|f(y) - f(x)\right| \mathrm{d}\lambda(y) \qquad \left(x \in \mathbb{R}^n\right)$$

and

$$\left(T f\right)(x) := \limsup_{r \to 0^+} \left(T_r f\right)(x) \qquad \left(x \in \mathbb{R}^n\right) .$$

Our goal is to prove that $Tf = 0$ λ-a.e.. Let $a > 0, m \in \mathbb{N}$.

We know that $C_c\left(\mathbb{R}^n\right)$ is dense in $L^1\left(\mathbb{R}^n\right)$ (this was proven in HW5Q15). Now if $f \in L^1$, then

$$\int_{\mathbb{R}^n} |f| \, \mathrm{d}\lambda < \infty ,$$

so in particular,

$$\lim_{R \to \infty} \int_{B_R(0)} |f| \, \mathrm{d}\lambda < \infty .$$

As such, for any $\varepsilon > 0$ there exists some $R_\varepsilon < \infty$ such that

$$\int_{B_{R_\varepsilon}(0)^c} |f| \, \mathrm{d}\lambda < \frac{\varepsilon}{2} .$$

Then approximate $f \chi_{B_{R_\varepsilon}(0)} \in L^1\left(\chi_{B_{R_\varepsilon}(0)}\right)$ by a continuous function g with compact support within $B_{R_\varepsilon}(0)$, up to precision $\frac{\varepsilon}{2}$. So we get

$$\begin{aligned}
\|f - g\|_{L^1(\mathbb{R}^n)} &\leq \left\|f - f\chi_{B_{R_\varepsilon}(0)}\right\|_{L^1(\mathbb{R}^n)} + \left\|f\chi_{B_{R_\varepsilon}(0)} - g\right\|_{L^1(\mathbb{R}^n)} \\
&= \left\|f\chi_{B_{R_\varepsilon}(0)^c}\right\|_{L^1(\mathbb{R}^n)} + \left\|f\chi_{B_{R_\varepsilon}(0)} - g\right\|_{L^1(B_{R_\varepsilon}(0))} \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon .
\end{aligned}$$

Pick $\varepsilon = \frac{1}{m}$ for some $m \in \mathbb{N}$ to get

$$\|f - g\|_1 < \frac{1}{m} .$$

Define $h := f - g$. Now, by continuity of g, $Tg = 0$ everywhere. Moreover,

$$\begin{aligned}
\left(T_r h\right)(x) &\equiv \frac{1}{\lambda\left(B_r(0)\right)} \int_{B_r(x)} \left|h(y) - h(x)\right| \mathrm{d}\lambda(y) \\
&\leq \frac{1}{\lambda\left(B_r(0)\right)} \int_{B_r(x)} \left|h(y)\right| \mathrm{d}\lambda(y) + |h(x)| \\
&= \frac{\varphi_{|h|,\lambda}\left(B_r(x)\right)}{\lambda\left(B_r(x)\right)} + |h(x)| \\
&= \frac{\left|\varphi_{h,\lambda}\right|\left(B_r(x)\right)}{\lambda\left(B_r(x)\right)} + |h(x)|
\end{aligned}$$

As a result, taking the lim sup $r \to 0$,

$$Th \leq \mathcal{M}\varphi_{h,\lambda} + h .$$

Hence

$$\left\{ x \in \mathbb{R}^n \mid (Th)(x) > 2a \right\} \subseteq \left\{ x \in \mathbb{R}^n \mid \left(\mathcal{M}\varphi_{h,\lambda}\right)(x) > a \right\} \cup \left\{ x \in \mathbb{R}^n \mid h(x) > a \right\} .$$

Moreover $\|h\|_1 < \frac{1}{m}$ and

$$\lambda\left(\left\{ x \in \mathbb{R}^n \mid |h(x)| > a \right\}\right) \leq \frac{1}{a}\|h\|_1$$

since

$$a\lambda\left(\left\{ x \in \mathbb{R}^n \mid |h(x)| > a \right\}\right) \leq \int_{\left\{ x \in \mathbb{R}^n \mid |h(x)| > a \right\}} |h| \, \mathrm{d}\lambda \leq \int_{\mathbb{R}^n} |h| \, \mathrm{d}\lambda = \|h\|_1 .$$

Moreover Theorem 6.6 now implies

$$\lambda\left(\left\{ x\in\mathbb{R}^n \mid \left(\mathcal{M}\varphi_{h,\lambda}\right)(x) > a \right\}\right) \leq 3^n \frac{\|\varphi_{h,\lambda}\|}{a} = 3^n \frac{|\varphi_{h,\lambda}|(X)}{a} = 3^n \frac{\|h\|_1}{a} \leq \frac{3^n}{am}\,.$$

We thus find

$$\lambda\left(\left\{ x\in\mathbb{R}^n \mid \left(\mathcal{M}\varphi_{h,\lambda}\right)(x) > a \right\} \cup \left\{ x\in\mathbb{R}^n \mid h(x) > a \right\}\right) \leq \frac{3^n+1}{am}\,.$$

Actually since this was true for arbitrary m, we find

$$\left\{ x\in\mathbb{R}^n \mid (Th)(x) > 2a \right\} \subseteq \bigcap_{m=1}^{\infty} \left\{ x\in\mathbb{R}^n \mid \left(\mathcal{M}\varphi_{h,\lambda}\right)(x) > a \right\} \cup \left\{ x\in\mathbb{R}^n \mid h(x) > a \right\}\,.$$

We find that $\left\{ x\in\mathbb{R}^n \mid (Th)(x) > 2a \right\}$ is a measurable subset of measure zero. Since this holds for any $a>0$, we find that $Th = 0$ λ-almost-everywhere, as desired. □

Theorem 6.13. *Let $\mu : \mathcal{B}(\mathbb{R}^n) \to \mathbb{C}$ be a measure and $\mu \blacktriangleleft \lambda$. Then*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = \mathcal{D}\mu$$

λ-almost-everywhere.

Proof. Let $x\in\mathbb{R}^n$ be a Lebesgue point of $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\in L^1$. Then, in particular,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x) = \lim_{\varepsilon\to 0^+} \frac{1}{\lambda(B_\varepsilon(0))}\int_{B_\varepsilon(x)} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\mathrm{d}\lambda \overset{\mu\blacktriangleleft\lambda}{=} \lim_{\varepsilon\to 0^+} \frac{\mu(B_\varepsilon(x))}{\lambda(B_\varepsilon(0))} = \lim_{\varepsilon\to 0^+} \frac{\mu(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} \equiv (\mathcal{D}\mu)(x)\,.$$

□

Theorem 6.14 (The fundamental theorem of calculus)**.** *Let $f\in L^1(\mathbb{R}\to\mathbb{C},\lambda)$. Then if x is a Lebesgue point of f,*

$$\left(\partial\int_{-\infty}^{\cdot} f\mathrm{d}\lambda\right)(x) = f(x)\,.$$

In particular the above equation holds λ-almost-everywhere.

Proof. Let $x\in\mathbb{R}$ be a Lebesgue point of f. Then we know by Theorem 6.12 that

$$\begin{aligned}\frac{1}{\lambda((x,x+\varepsilon))}\int_{(x,x+\varepsilon)} |f(y)-f(x)|\,\mathrm{d}\lambda(y) \quad &\leq \quad 2\frac{1}{\lambda(B_\varepsilon(x))}\int_{B_\varepsilon(x)} |f(y)-f(x)|\,\mathrm{d}\lambda(y) \\ &\overset{\varepsilon\to 0^+}{\to} \quad 0\,.\end{aligned}$$

A similar consideration leads also to

$$\lim_{\varepsilon\to 0^+}\frac{1}{\lambda((x-\varepsilon,x))}\int_{(x-\varepsilon,x)} |f(y)-f(x)|\,\mathrm{d}\lambda(y) = 0\,.$$

In particular we thus have

$$\lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\left[\int_{-\infty}^{x+\varepsilon} f\mathrm{d}\lambda - \int_{-\infty}^{x} f\mathrm{d}\lambda\right] \quad = \quad \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\int_{x}^{x+\varepsilon} f\mathrm{d}\lambda = f(x)$$

and similarly

$$\lim_{\varepsilon\to 0^+}\frac{1}{-\varepsilon}\left[\int_{-\infty}^{x-\varepsilon} f\mathrm{d}\lambda - \int_{-\infty}^{x} f\mathrm{d}\lambda\right] = \lim_{\varepsilon\to 0^+}\frac{1}{-\varepsilon}\left[\int_{x-\varepsilon}^{x} f\mathrm{d}\lambda\right] = f(x)\,.$$

□

We now only state, but do not prove, the other half of the fundamental theorem of calculus:

Definition 6.15 (Absolutely continuous functions)**.** Let $f : [a, b] \to \mathbb{C}$ is called *absolutely continuous* iff for any $\varepsilon > 0$ there exists some $\delta > 0$ such that

$$\sum_{j=1}^{n} |f(\beta_j) - f(\alpha_i)| < \varepsilon$$

for any $n \in \mathbb{N}$ and any disjoint collection of segments $(\alpha_1, \beta_1), \cdots, (\alpha_n, \beta_n) \subseteq [a, b]$ which satisfies

$$\sum_{j=1}^{n} \beta_j - \alpha_j < \delta\,.$$

Theorem 6.16. *Let $f : [a, b] \to \mathbb{R}$ be continuous and nondecreasing. Then the following are equivalent:*

1. *f is absolutely continuous.*
2. *f maps sets of measure zero to sets of measure zero.*
3. *f is differentiable λ-almost-everywhere on $[a, b]$, $f' \in L^1([a, b])$ and*

$$f(x) = f(a) + \int_a^x f' \mathrm{d}\lambda \qquad (x \in [a, b])\ .$$

Theorem 6.17 (2nd half of the fundamental theorem of calculus)**.** *Let $f : [a, b] \to \mathbb{C}$ be absolutely continuous. Then f is differentiable λ-almost-everywhere on $[a, b]$, $f' \in L^1([a, b] \to \mathbb{C}, \lambda)$ and*

$$f(x) = f(a) + \int_a^x f' \mathrm{d}\lambda \qquad (x \in [a, b])\ .$$

6.2 The change of variable formula revisited on $\mathbb{R}^n$

We now want to go back to the change of variables formula (5.2). So we assume that $(X, \mathcal{M}, \mu)$ is a measure space, $(Y, \mathcal{N})$ is a measurable space and $\varphi : X \to Y$ be measurable and *injective*. We know, by (5.3), that for any measurable $f : Y \to \mathbb{C}$,

$$\int_A f \circ \varphi \mathrm{d}\mu = \int_{\varphi(A)} f \mathrm{d}\mu_\varphi\,.$$

We may slightly rephrase this: assume φ is *bijective* so that $\varphi^{-1} : Y \to X$ is measurable. Then with $g := f \circ \varphi : X \to \mathbb{C}$, $f = g \circ \varphi^{-1}$ so that we get

$$\int_A g \mathrm{d}\mu = \int_{\varphi(A)} g \circ \varphi^{-1} \mathrm{d}\mu_\varphi \qquad (A \in \mathcal{M})\ .$$

Relabeling $\eta := \varphi^{-1}$ and $B := \eta^{-1}(A)$ we find

$$\int_{\eta(B)} g \mathrm{d}\mu = \int_B g \circ \eta \mathrm{d}\mu_{\eta^{-1}} \qquad (B \in \mathcal{N})\ .$$

Let us assume that there is yet another measure ν on Y and that additionally $\mu_{\eta^{-1}} \blacktriangleleft \nu$. Then by the Lebesgue decomposition theorem Theorem 5.42,

$$\int_{\eta(B)} g \mathrm{d}\mu = \int_B g \circ \eta \frac{\mathrm{d}\mu_{\eta^{-1}}}{\mathrm{d}\nu} \mathrm{d}\nu \qquad (B \in \mathcal{N})\ .$$

Our goal here is to explore the expression $\frac{\mathrm{d}\mu_{\eta^{-1}}}{\mathrm{d}\nu}$ when $\mu = \nu = \lambda$ is the Lebesgue measure on $Y := \mathbb{R}^n$. We will see that then

$$\frac{\mathrm{d}\lambda_{\eta^{-1}}}{\mathrm{d}\lambda} = \left|\det\left(\mathcal{D}\eta^{-1}\right)\right|\ .$$

where $\mathcal{D}\eta^{-1}$ is the Frechet derivative of η^{-1}.

Theorem 6.18. *Let $V \in \text{Open}(\mathbb{R}^n)$ and $\varphi : V \to \mathbb{R}^n$ be a continuously differentiable map whose inverse (defined on its image) is also continuously differentiable. Then the push-forward measure*

$$\lambda_{\varphi^{-1}} \equiv \lambda \circ \varphi$$

is absolutely-continuous w.r.t. the Lebesgue measure λ and its Radon-Nikodym derivative equals

$$\frac{\mathrm{d}\lambda_{\varphi^{-1}}}{\mathrm{d}\lambda} = |\det(\mathscr{D}\varphi)| \ . \tag{6.1}$$

As such, the change of variables formula becomes

$$\int_{\varphi(A)} f \mathrm{d}\lambda = \int_A f \circ \varphi \, |\det(\mathscr{D}\varphi)| \, \mathrm{d}\lambda \qquad (A \in \mathscr{B}(\mathbb{R}^n)) \ . \tag{6.2}$$

We begin with some preliminaries necessary for the proof of the theorem. The first one should be intuitive: it states that if we scale the Lebesgue measure by a matrix, then we get a factor of the determinant outside. Since the Lebesgue measure measures volume, and the determinant measures the volume of a parallelepiped, this makes sense.

Lemma 6.19. *Let $M : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map (an $n \times n$ matrix). Then*

$$\lambda(MA) = |\det(M)| \, \lambda(A) \qquad (A \in \mathscr{B}(\mathbb{R}^n)) \ .$$

Proof. If M does not have full rank, then $\det(M) = 0$. In which case, MA lies in a k-dimensional subspace of $\mathbb{R}^n$, the subspace $M\mathbb{R}^n$, whose measure is zero as it is lower dimensional. To see this, we may choose a coordinate base where $e_1, \cdots, e_k$ is an orthonormal basis of $M\mathbb{R}^n$ ($k < n$ by hypothesis) and $e_{k+1}, \cdots, e_n$ completes to a basis of $\mathbb{R}^n$. We then consider

$$\mathbb{R}^n \cong \mathbb{R}^k \times \mathbb{R}^{n-k}$$

where the first factor is $M\mathbb{R}^n$ and the other one is its orthogonal complement. Then for any $A \in \mathscr{B}(\mathbb{R}^n)$, MA is a rectangular set in this product structure, and so

$$\begin{aligned} \lambda_{\mathbb{R}^n}(MA) &= \lambda_{\mathbb{R}^k}(MA) \, \lambda_{\mathbb{R}^{n-k}}(\{0\}) \\ &= \lambda_{\mathbb{R}^k}(MA) \cdot 0 \\ &= 0 \, . \end{aligned}$$

Now if M does have full rank, then it is invertible and its inverse is continuous, so that $MA = \left(M^{-1}\right)^{-1}(A)$ is also a Borel set. Then translation invariance of λ and linearity of M implies

$$\lambda(M(A + x)) = \lambda(MA + Mx) = \lambda(MA) \ .$$

Hence the positive Borel measure $\lambda(M\cdot)$ is translation-invariant. We define the normalized measure

$$\mu := \frac{\lambda(M\cdot)}{\lambda(M[0,1]^n)} \, ,$$

which must, by the uniqueness Theorem 4.12, *equals* to the Lebesgue measure. We thus find

$$\lambda(MA) = \lambda(M[0,1]^n) \, \lambda(A) \qquad (A \in \mathscr{B}(\mathbb{R}^n)) \ .$$

We are thus left with proving, for all invertible M,

$$\lambda(M[0,1]^n) = |\det(M)| \ .$$

For diagonal matrices M this is true by the stated scaling property in the proof of Theorem 4.12:

$$M = \begin{bmatrix} m_1 & & & \\ & m_2 & & \\ & & \ddots & \\ & & & m_n \end{bmatrix}$$

and then

$$\lambda\left(M\left[0,1\right]^n\right) \quad = \quad \lambda\left(\left[0,\left|m_1\right|\right]\times\cdots\times\left[0,\left|m_n\right|\right]\right) = \left|m_1\cdots m_n\right| = \left|\det\left(M\right)\right| \,.$$

In the more general case, any matrix M may be factorized into $M = LU$ where $\left|\det\left(L\right)\right| = 1$ because it is a "shear"–it does not change volume–and U is upper triangular. U maps $\left[0,1\right]^n$ to a parallelepiped whose volume is $\left|\det\left(U\right)\right|$. Admittedly to follow through this proof completely one has to have a geometric interpretation of λ as measuring volume using the premeasure on $\mathbb{R}^n$ directly rather than with the product measure construction we have presented above. □

The above statement was for a constant matrix. If we have a general map, then this gets changed by the *differentiable* of a map. Recall that a map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is said to be differentiable at some $x_0 \in \mathbb{R}^n$ iff there exists a linear map $M : \mathbb{R}^n \to \mathbb{R}^n$ (dependent on φ and on x_0) such that the following limit exists and equals zero:

$$\lim_{y\to 0}\frac{\left\|\varphi\left(x_0+y\right)-\varphi\left(x_0\right)-My\right\|}{\left\|y\right\|} = 0\,.$$

When that happens, we say that M is the (total, or, Frechet) derivative of φ at x_0 and denote that linear map as

$$\left(\mathcal{D}\varphi\right)\left(x_0\right)\,.$$

In principle,

$$\mathbb{R}^n \ni x_0 \mapsto \left(\mathcal{D}\varphi\right)\left(x_0\right) \in \mathrm{Mat}_{n\times n}\left(\mathbb{R}\right)$$

defines a new map. The determinant of this map is called the Jacobian function associated to φ.

Theorem 6.20. *Let $\varphi : V \to \mathbb{R}^n$ be continuous where $V \in \mathrm{Open}\left(\mathbb{R}^n\right)$. Assume that φ is differentiable at $x \in V$. Then*

$$\lim_{r\to 0^+}\frac{\lambda\left(\varphi\left(B_r\left(x\right)\right)\right)}{\lambda\left(B_r\left(x\right)\right)} = \left|\det\left(\left(\mathcal{D}\varphi\right)\left(x\right)\right)\right|\,.$$

Proof. By possibly shifting the coordinate axes and shifting φ by a constant, assume that $x = 0$ and $\varphi\left(x\right) = 0$. Define $M := \left(\mathcal{D}\varphi\right)\left(0\right)$.

Case 1: M is invertible. Define $\Phi := M^{-1}\circ\varphi$. Then by the Leibniz rule,

$$\left(\mathcal{D}\Phi\right)\left(0\right) = M^{-1}\left(\mathcal{D}\varphi\right)\left(0\right) = \mathbb{1}_n\,.$$

We want to show that

$$\lim_{r\to 0^+}\frac{\lambda\left(\Phi\left(B_r\left(0\right)\right)\right)}{\lambda\left(B_r\left(0\right)\right)} = 1\,.$$

Let $\varepsilon > 0$. Then $\Phi\left(0\right) = 0$, and $\left(\mathcal{D}\Phi\right)\left(0\right) = \mathbb{1}_n$, so there is some $\delta > 0$ such that if $x \in B_\delta\left(0\right)\setminus\left\{\,0\,\right\}$ then

$$\Phi\left(x\right) \approx \underbrace{\Phi\left(0\right)}_{=0} + \underbrace{\left(\mathcal{D}\Phi\right)\left(0\right)}_{=\mathbb{1}_n} x + \cdots$$

or more preicsely,

$$\left\|\Phi\left(x\right)-x\right\| \leq \varepsilon\left\|x\right\|. \tag{6.3}$$

Now, if $r \in \left(0,\delta\right)$ then

$$B_{\left(1-\varepsilon\right)r}\left(0\right) \subseteq \Phi\left(B_r\left(0\right)\right)\,.$$

This will be proven right below in Lemma 6.21. Assuming this, we get

$$\Phi\left(B_r\left(0\right)\right) \subseteq B_{\left(1+\varepsilon\right)r}\left(0\right)$$

from (6.3). Thus

$$B_{(1-\varepsilon)r}(0) \subseteq \Phi(B_r(0)) \subseteq B_{(1+\varepsilon)r}(0)$$

which implies

$$(1-\varepsilon)^n \leq \frac{\lambda(\Phi(B_r(0)))}{\lambda(B_r(0))} \leq (1+\varepsilon)^n .$$

Taking $\varepsilon \to 0^+$ we obtain the claim. But now,

$$\lambda(\varphi(B)) = \lambda(M\Phi(B)) = |\det(M)|\,\lambda(\Phi(B))$$

for every ball B. As a result, we get the result.

Case 2: M is not invertible. Since φ is *continuous* differentiable, for any $\eta > 0$ there exists some $\delta > 0$ such that if $x \in B_\delta(0)$ then

$$\|\varphi(x) - Mx\| \leq \eta\|x\| .$$

Since M is not invertible, its image is a subspace of dimension $k < n$, and hence of measure zero. That means that for any $\varepsilon > 0$ there exists some $\eta > 0$ such that

$$\lambda(\{\, x \in \mathbb{R}^n \mid \operatorname{dist}(MB_1(0), x) < \eta \,\}) < \varepsilon .$$

Hence if $r < \delta$,

$$\varphi(B_r(0)) \subseteq \{\, x \in \mathbb{R}^n \mid \operatorname{dist}(MB_r(0), x) < r\eta \,\}$$

and so

$$\begin{aligned} \lambda(\varphi(B_r(0))) &\leq \lambda(\{\, x \in \mathbb{R}^n \mid \operatorname{dist}(MB_r(0), x) < r\eta \,\}) \\ &\leq \varepsilon r^n \end{aligned}$$

for any $r \in (0, \delta)$. Hence

$$\lim_{r\to 0^+} \frac{\lambda(\varphi(B_r(0)))}{\lambda(B_r(0))} = \lim_{r\to 0^+} \frac{\lambda(\varphi(B_r(0)))}{r^n\lambda(B_1(0))} \leq \lim_{r\to 0^+} \varepsilon$$

for any $\varepsilon > 0$ and thus the limit is zero. □

Lemma 6.21. *Let* $\mathbb{S}^{n-1} \subseteq \mathbb{R}^n$ *be given by*

$$\mathbb{S}^{n-1} \equiv \{\, x \in \mathbb{R}^n \mid \|x\| = 1 \,\} .$$

Note $\mathbb{S}^{n-1} = \partial B_1(0)$*. If*

$$f : \overline{B_1(0)} \to \mathbb{R}^n$$

is continuous and $\varepsilon \in (0, 1)$ *is such that*

$$\|f(x) - x\| < \varepsilon \qquad \left(x \in \mathbb{S}^{n-1}\right)$$

then

$$f(B_1(0)) \supseteq B_{1-\varepsilon}(0) .$$

Proof. Assume there exists some point

$$a \in B_{1-\varepsilon}(0) \setminus f(B_1(0)) .$$

Then, for any $x \in \mathbb{S}^{n-1}$,

$$\varepsilon > \|f(x) - x\| \geq \|x\| - \|f(x)\| = 1 - \|f(x)\|$$

i.e.,

$$\|f(x)\| > 1 - \varepsilon .$$

As a result, $a \notin f\left(\mathbb{S}^{n-1}\right)$. So $a \notin f\left(\overline{B_1(0)}\right)$. So define a continuous $G : \overline{B_1(0)} \to \overline{B_1(0)}$ via

$$x \mapsto \frac{a - f(x)}{\|a - f(x)\|} .$$

We now show that G fixes *no point of* $\overline{B_1(0)}$, in contradiction with Brouwer's fixed point theorem [Hat01] (Section 2.2). Indeed, if $x \in \mathbb{S}^{n-1}$, then

$$\langle x, a - f(x)\rangle = \langle x, a\rangle + \langle x, x - f(x)\rangle - \langle x, x\rangle = \langle x, a\rangle + \langle x, x - f(x)\rangle - 1 < \|x\|\|a\| + \|x\|\|x - f(x)\| - 1 < \|a\| + \varepsilon - 1 < 0 .$$

Hence

$$\langle x, G(x)\rangle < 0$$

so in particular $x \neq G(x)$ for all $x \in \mathbb{S}^{n-1}$. If $x \in B_1(0)$, then $x \notin \mathbb{S}^{n-1}$ but by definition $\operatorname{im}(G) \subseteq \mathbb{S}^{n-1}$, so $x \neq G(x)$. □

Lemma 6.22. *Let $E \subseteq \mathbb{R}^n$ be a null set: $\lambda(E) = 0$, $\varphi : E \to \mathbb{R}^n$ and assume that*

$$\limsup_{E \ni y \to x} \frac{\|\varphi(y) - \varphi(x)\|}{\|y - x\|} < \infty \qquad (x \in E) . \tag{6.4}$$

Then $\lambda(\varphi(E)) = 0$.

Proof. Let $m, p \in \mathbb{N}$ be given and define

$$F_{m,p} := \left\{ x \in E \;\middle|\; \|\varphi(y) - \varphi(x)\| \leq m\|y - x\| \forall y \in B_{\frac{1}{p}}(x) \cap E \right\} .$$

Let $\varepsilon > 0$. We have $\lambda(F_{m,p}) \leq \lambda(E) = 0$, so there are balls $B_{r_i}(x_i)$ such that $x_i \in F_{m,p}$, $r_i < \frac{1}{p}$ and

$$\bigcup_i B_{r_i}(x_i) \supseteq F_{m,p} \wedge \sum_i \lambda(B_{r_i}(x_i)) < \varepsilon .$$

Now if $x \in F_{m,p} \cap B_{r_i}(x_i)$, then $\|x - x_i\| < r_i < \frac{1}{p}$ and hence $x_i \in F_{m,p}$. Thus

$$\|\varphi(x_i) - \varphi(x)\| \leq m\|x_i - x\| < m r_i$$

and so

$$\varphi(F_{m,p} \cap B_{r_i}(x_i)) \subseteq B_{m r_i}(\varphi(x_i))$$

and thus

$$\varphi(F_{m,p}) \subseteq \bigcup_i B_{m r_i}(\varphi(x_i)) .$$

We estimate

$$\lambda\left(\bigcup_i B_{m r_i}(\varphi(x_i))\right) \leq \sum_i \lambda(B_{m r_i}(\varphi(x_i))) \leq m^n \sum_i B_{r_i}(x_i) < m^n \varepsilon .$$

But the Lebesgue measure is complete and ε was arbitrary, so $\varphi(F_{m,p})$ must be measurable and $\lambda(\varphi(F_{m,p})) = 0$. To complete the proof we note that

$$E = \bigcup_{m,p} F_{m,p} .$$

□

We are now ready for the

Proof of Theorem 6.18. The first item on the list is the proof that

$$\lambda_{\varphi} \quad \blacktriangleleft \quad \lambda\,.$$

This is a consequence of the fact that once φ is differentiable, it obeys (6.4), so in particular it maps sets of measure to sets of measure zero. In fact φ maps measurable sets to measurable sets.

Moreover, we have just seen above that

$$\frac{\mathrm{d}\lambda_{\varphi}}{\mathrm{d}\lambda}\left(x\right)=\lim_{r\to0^{+}}\frac{\lambda\left(\varphi\left(B_{r}\left(x\right)\right)\right)}{\lambda\left(B_{r}\left(x\right)\right)}=\left|\det\left(\left(\mathscr{D}\varphi\right)\left(x\right)\right)\right|\,.$$

□

7 Probability theory–measure theory with a soul [Folland]

[TODO: remove reference for "distribution" by itself, always specify "cumulative" or "density"]

In this chapter we use the tools we've developed to *introduce* the basics of probability theory.

A *probability space* is a measure space $(\Omega,\mathfrak{M},\mu)$ with Ω a non-empty set, $\mathfrak{M}$ some σ-algebra on it, and $\mu:\mathfrak{M}\to[0,\infty]$ a measure such that $\mu\left(\Omega\right)=1$[3]. Clearly, given any finite measure space $\mu\left(\Omega\right)<\infty$ we may re-define μ to normalize it so the main question is whether $\mu\left(\Omega\right)=\infty$ or not. Probability theory has a few different notations and terminology compared with (and sometimes in contrast to) the rest of measure theory:

- The normalized measure μ is usually denoted by $\mathbb{P}$:

  $$\mathbb{P}\left(\Omega\right)=1$$

  and we even go further and denote the measure of sets with square rather than round brackets for some reason:

  $$\mathbb{P}\left(A\right)\mapsto\mathbb{P}\left[A\right]\,.$$

  We will follow suit.

- The (Lebesgue) integral (2.15) with respect to the (fixed) probability measure $\mathbb{P}$ is denoted by $\mathbb{E}$ and a measurable function $f:\Omega\to\mathbb{C}$, usually denoted by $X:\Omega\to\mathbb{C}$ rather than f, is called a *random variable*. Then it is customary to use the notation

  $$\int_{\Omega}X\mathrm{d}\mathbb{P}\equiv\int_{\omega\in\Omega}X\left(\omega\right)\mathrm{d}\mathbb{P}\left(\omega\right)\mapsto\mathbb{E}\left[X\right]\,.$$

  So in probability we almost never write out the integration variable explicitly. This integral of X is referred to as the *expectation* of the random variable X.

- Rather than speak of the "bare" probability measures, we are often more interested in the probability distributions *induced* by *random variables*, as it were. That means, if $X:\Omega\to\mathbb{C}$ is a random variable with associated probability measure $\mathbb{P}:\mathrm{Msrbl}\left(\Omega\right)\to[0,1]$, then a probability measure is induced on $\mathscr{B}\left(\mathbb{C}\right)$ via the *push forward* construction:

  $$\mathbb{P}_{X}\left[A\right]:=\mathbb{P}\left[X^{-1}\left(A\right)\right]\qquad\left(A\in\mathscr{B}\left(\mathbb{C}\right)\right)\,.$$

  In more concrete terms this is the probability that X takes on values in A. It is called *the law of X*. Using (5.2), if $f:\mathbb{C}\to\mathbb{C}$ is measurable, then

  $$\mathbb{E}\left[f\left(X\right)\right]\equiv\int_{\omega\in\Omega}f\left(X\left(\omega\right)\right)\mathrm{d}\mathbb{P}\left(\omega\right)=\int_{x\in\mathbb{C}}f\left(x\right)\mathrm{d}\mathbb{P}_{X}\left(x\right)\equiv\mathbb{E}_{X}\left[f\right]\,.$$

- If $X:\Omega\to\mathbb{R}$ (i.e. it is real-valued) then the function

  $$\mathbb{R}\ni t\mapsto\mathbb{P}_{X}\left[(-\infty,t]\right]$$

  is called *the cumulative distribution function*. Formally its derivative is the distribution of the random variable X (although sometimes it is not differentiable, so the cumulative distribution function is somewhat more "robust" object to handle). When $\mathbb{P}_{X}\blacktriangleleft\lambda$ then we have a Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}_{X}}{\mathrm{d}\lambda}$ and

  $$\int_{x\in\mathbb{C}}f\left(x\right)\mathrm{d}\mathbb{P}_{X}\left(x\right)=\int_{x\in\mathbb{C}}f\left(x\right)\frac{\mathrm{d}\mathbb{P}_{X}}{\mathrm{d}\lambda}\left(x\right)\mathrm{d}\lambda\left(x\right).$$

[3] In this chapter we use a different convention where our ambient set is not X but Ω, since X will be used for measurable functions.

The function

$$\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\lambda}$$

is called *the distribution (density) function* of X. There are many *standard* distribution functions that we will encounter.

- Clearly we can create new functions out of old ones via algebraic manipulations. These then yield new random variables. We define the *variance* of a random variable $X:\Omega\to\mathbb{C}$ as

$$\mathbb{Var}[X] \quad := \quad \mathbb{E}\left[(X-\mathbb{E}[X])^2\right]=\mathbb{E}\left[X^2\right]-\mathbb{E}[X]^2\,.$$

  Usually this is only defined for real-valued random variables; the variance measures the typical degree of deviation of the function from its mean, average value, i.e., *the standard deviation*

$$\sqrt{\mathbb{Var}[X]}$$

  is the "typical" deviation of X away from its average $\mathbb{E}[X]$ as we shall see. Note the variance has the scaling

$$\mathbb{Var}[\alpha X]=\alpha^2\mathbb{Var}[X]\qquad(\alpha\in\mathbb{R})\,. \tag{7.1}$$

- The *characteristic function* $\varphi_X:\mathbb{C}\to\mathbb{C}\cup\{\,\infty\,\}$ of a random variable $X:\Omega\to\mathbb{C}$ is given by

$$\varphi_X(t):=\mathbb{E}[\exp(\mathrm{i}tX)]\qquad(t\in\mathbb{C})\,.$$

  The *moment generating function* $M_X:\mathbb{C}\to\mathbb{C}\cup\{\,\infty\,\}$ of a random variable $X:\Omega\to\mathbb{C}$ is given by

$$M_X(t):=\mathbb{E}[\exp(tX)]\qquad(t\in\mathbb{C})\,.$$

  Note that it's not a-priori clear that both these functions exist on the entire complex plane. Both of these functions yield *the moments of* X by taking derivatives w.r.t. t:

$$\varphi_X^{(n)}(0)=(\mathrm{i})^n\,\mathbb{E}[X^n]$$

  and

$$M_X^{(n)}(0)=\mathbb{E}[X^n]$$

  which is the reason for the name, at least of the latter.
  It should be noted that in the definition of both φ_X and M_X, usually one takes $t\in\mathbb{R}$. Our point of view of taking $t\in\mathbb{C}$ is not common and is just meant for later flexibility.

- We also have the *cumulant-generating function*

$$K_X(t):=\log(M_X(t))\qquad(t\in\mathbb{C})\,.$$

  Its derivatives at zero give linear combinations of moments called *cumulants:*

$$K_X^{(n)}(0)\quad=\quad\partial_t^n|_{t=0}\log(\mathbb{E}[\exp(tX)])\,.$$

  Clearly, the zeroth cumulant is zero, the first cumulant is the mean, the second is the variance, but the third and so on are already different. Two random variables have the same sequence of moments iff they have the same sequence of cumulants.

- In statistical mechanics, if we start with a finite measure $\mu(\Omega)<\infty$ which is not necessarily normalized, then $Z:=\mu(\Omega)$ is called *the partition function.*

- In statistical mechanics we are often times interested in random variables on *product spaces*, i.e., situations where $\Omega=\prod_{i\in\{\,1,\cdots,N\,\}}S$ for some space S, and then μ is given as a density w.r.t. the product measure. HW6Q5 becomes a useful resource then.

Example 7.1 (Standard normal random variable)**.** Choose as the probability space

$$(\Omega,\mathrm{Msrbl}(\Omega),\mathbb{P})=:(\mathbb{R},\mathcal{B}(\mathbb{R}),f\mathrm{d}\lambda)$$

with $f:\mathbb{R}\to\mathbb{R}$ given by the Gaussian

$$f(x):=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)\qquad(x\in\mathbb{R})\,.$$

Define the random variable $X : \mathbb{R} \to \mathbb{R}$ given by the identity map $x \mapsto x$. We say then that X is a *standard normal* random variable, denoted by

$$X \sim \mathcal{N}(0,1) \,,$$

One verifies that then, $\mathbb{E}[X] = 0$ and $\mathbb{E}\left[X^2\right] = 1$. In this case the function $X : \mathbb{R} \to \mathbb{R}$ is just the identity map, so the measure $\mathbb{P}$ carries all the probabilistic information already.

When we have only one random variable then it will be convenient to choose $X : \mathbb{R} \to \mathbb{R}$ as simply the identity map.

Example 7.2 (Cauchy random variable)**.** We say X is a *standard Cauchy* random variable, denoted by

$$X \sim \text{Cauchy}(0,1)$$

iff $X : \mathbb{R} \to \mathbb{R}$ $x \mapsto x$ is distributed according to

$$\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\lambda}(x) = \frac{1}{\pi}\frac{1}{x^2+1} \,.$$

We note the peculiar fact that $\mathbb{E}[X^n] = \infty$ for all $n \in \mathbb{N}$, i.e., *the Cauchy distribution has no moments.*

Example 7.3 (Uniform random variable)**.** Let $-\infty < a < b < \infty$ be given. We define a measure $\mathbb{P}$ on $\text{Msrbl}(\Omega) := \mathfrak{B}(\mathbb{R})$ via

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x) = \chi_{[a,b]}(x)\frac{1}{b-a} \,.$$

Then if $X : \Omega \to \mathbb{R}$ is again the identity map $x \mapsto x$, we get that the law of X is the *uniform distribution*

$$X \sim \text{Uniform}(a,b) \,.$$

We don't *always* have to pick the identity map. For instance, consider the random variable $Y : \Omega \to \mathbb{R}$ as $x \mapsto \frac{1}{2}x$ implies

$$\begin{aligned}\mathbb{P}_Y[A] &\equiv \int_{x\in\mathbb{R}} \chi_A(Y(x))\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x)\,\mathrm{d}\lambda(x) \\ &= \int_{x\in\mathbb{R}} \chi_A\left(\frac{1}{2}x\right)\chi_{[a,b]}(x)\frac{1}{b-a}\mathrm{d}\lambda(x)\end{aligned}$$

We now use the change of variable formula (6.2) (with $\varphi : x \mapsto \frac{1}{2}x$) to get

$$\begin{aligned}\mathbb{P}_Y[A] &= \int_{y\in\mathbb{R}} \chi_A(y)\chi_{[a,b]}(2y)\frac{1}{b-a}2\mathrm{d}\lambda(y) \\ &= \int_{y\in\mathbb{R}} \chi_A(y)\chi_{\left[\frac{a}{2},\frac{b}{2}\right]}(y)\frac{1}{\frac{b}{2}-\frac{a}{2}}\mathrm{d}\lambda(y)\end{aligned}$$

so that

$$Y \sim \text{Uniform}\left(\frac{a}{2},\frac{b}{2}\right) \,.$$

7.1 Multiple random variables

Often we are interested in situations where there are multiple (or even infinitely many) random variables. They all have their domain as the same *probability space*, however, sometimes that probability space takes itself the form of a product space. For instance, consider the case

$$\Omega = \Omega_1 \times \Omega_2$$

is the space of events, with $\mathfrak{M} := \mathfrak{M}_1 \otimes \mathfrak{M}_2$. Consider then that we are given the total probability measure $\mathbb{P} : \mathfrak{M} \to [0,1]$ (we do not know that it is a product measure). Then if

$$\pi_j : \Omega \to \Omega_j$$

is the projection onto the jth coordinate, it is natural to take $X_j := \pi_j$ as two random variables. Then the induced laws are

$$\mathbb{P}_{X_j}[A] \equiv \mathbb{P}\left[\pi_j^{-1}(A)\right] \,.$$

Example 7.4. Consider $\Omega := \mathbb{R}^2$ and what is usually called the *joint probability distribution function*

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\lambda}(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(x_1^2 + x_2^2\right)\right) \qquad (x_1, x_2 \in \mathbb{R}) \ .$$

Then with $X_j := \pi_j$ for $j = 1, 2$ indeed, we have

$$\begin{aligned} \mathbb{P}_{X_1}[A] &\equiv \mathbb{P}\left[\pi_1^{-1}(A)\right] \\ &= \int_{x \in \mathbb{R}^2} \chi_{\pi_1^{-1}(A)}(x_1, x_2) \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(x_1^2 + x_2^2\right)\right) \mathrm{d}\lambda(x_1, x_2) \\ &= \int_{x_1 \in \mathbb{R}} \chi_A(x_1) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) \mathrm{d}\lambda(x_1) \end{aligned}$$

so we recognize that

$$\frac{\mathrm{d}\mathbb{P}_{X_1}}{\mathrm{d}\lambda}(x_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) \qquad (x_1 \in \mathbb{R}) \ .$$

The law $\mathbb{P}_{X_1}$ is called *the marginal.* We may alternatively also consider *the conditional probability measure*

$$\mathbb{P}_X[X \in A | Y \in B] := \frac{\mathbb{P}[X \in A \wedge Y \in B]}{\mathbb{P}[Y \in B]} \qquad (A, B \text{ msrbl.}) \ . \tag{7.2}$$

Clearly

$$A \mapsto \mathbb{P}_X[X \in A | Y \in B] \in [0, 1]$$

is a probability distribution.

Definition 7.5. For any two real-valued random variables X, Y, we define their *covariance* as

$$\mathbb{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \ .$$

This may be considered a measure of their mutual dependence.

Example 7.6. Of course the most dependence we could have is for a random variable with itself, whence

$$\mathbb{Cov}[X, X] = \mathbb{Var}[X] \ .$$

As we shall see, if two random variables are *independent* then their covariance vanishes. The converse is false.

Definition 7.7 (Stochastic process)**.** Given a probability space $(\Omega, \mathrm{Msrbl}(\Omega), \mathbb{P})$, a *stochastic process* is a sequence of random variables $\{ X_n : \Omega \to \mathbb{C} \}_{n \in \mathbb{N}}$. In principle we could also consider the indexing set of the sequence a continuous variable.

Remark 7.8. Do not confuse marginal with martingale: the latter is a *type* of stochastic process, which we will cover later in TODO: CITE.

7.2 Independence

Definition 7.9 (Independent events)**.** Let A be some index set. The collection of events $\{ E_\alpha \}_{\alpha \in A} \subseteq \mathrm{Msrbl}(\Omega)$ is said to be independent iff for any $S \subseteq A$ such that $|S| < \infty$,

$$\mathbb{P}\left[\bigcap_{\alpha \in S} E_\alpha\right] = \prod_{\alpha \in S} \mathbb{P}[E_\alpha] \ .$$

Definition 7.10 (Independent random variables)**.** Let A be some set. The collection of random variables $\{ X_\alpha : \Omega \to \mathbb{C} \}_{\alpha \in A}$ are said to be independent iff $\left\{ X_\alpha^{-1}(B_\alpha) \right\}_{\alpha \in A} \subseteq \Omega$ is an independent collection of events, for *any* collection of Borel sets $\{ B_\alpha \}_\alpha \subseteq \mathfrak{B}(\mathbb{C})$.

Claim 7.11. The collection of random variables $\{ X_\alpha : \Omega \to \mathbb{C} \}_{\alpha \in A}$ are independent iff for any $S = \{ \alpha_1, \cdots, \alpha_n \} \subseteq A$,

$$\mathbb{P}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)} = \prod_{j=1}^{n} \mathbb{P}_{X_{\alpha_j}} \tag{7.3}$$

where by $\mathbb{P}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)}$ we mean the induced distribution of the variables $X_{\alpha_1}, \cdots, X_{\alpha_n}$ where all others are "integrated out":

$$\mathbb{P}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)}[A] \equiv \mathbb{P}\left[(X_{\alpha_1}, \cdots, X_{\alpha_n}) \in A\right] \equiv \mathbb{P}\left[(X_{\alpha_1}, \cdots, X_{\alpha_n})^{-1}(A)\right] \qquad (A \in \mathfrak{B}(\mathbb{C}^n)) .$$

On the RHS we mean the product measure construction Definition 5.6.

Proof. Let us assume (7.3). Let $\{ B_\alpha \}_\alpha \subseteq \mathfrak{B}(\mathbb{C})$ be given and pick any $S \subseteq A$ such that $|S| < \infty$. Then we want to show that

$$\mathbb{P}\left[\bigcap_{\alpha \in S} X_\alpha^{-1}(B_\alpha)\right] = \prod_{\alpha \in S} \mathbb{P}\left[X_\alpha^{-1}(B_\alpha)\right] .$$

We recognize the RHS as

$$\mathbb{P}\left[X_\alpha^{-1}(B_\alpha)\right] \equiv \mathbb{P}_{X_\alpha}[B_\alpha] .$$

On the LHS, we have

$$\mathbb{P}\left[\bigcap_{\alpha \in S} X_\alpha^{-1}(B_\alpha)\right] = \mathbb{P}\left[(X_{\alpha_1}, \cdots, X_{\alpha_n})^{-1}\left(\prod_{\alpha \in S} B_\alpha\right)\right] \equiv \mathbb{P}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)}\left[\prod_{\alpha \in S} B_\alpha\right] .$$

Since we are assuming (7.3), we get

$$\mathbb{P}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)}\left[\prod_{\alpha \in S} B_\alpha\right] = \prod_{\alpha \in S} \mathbb{P}_{X_\alpha}[B_\alpha]$$

so the two are indeed equal as needed. For the other direction, if two Borel measures agree on all rectangular Borel sets, then they agree by regularity. ☐

Corollary 7.12. *Let A be some indexing set and $\{ X_\alpha : \Omega \to \mathbb{C} \}_{\alpha \in A}$ be an independent sequence of random variables. Then for any $S = \{ \alpha_1, \cdots, \alpha_n \} \subseteq A$,*

$$\mathbb{E}\left[\prod_{\alpha \in S} X_\alpha\right] = \prod_{\alpha \in S} \mathbb{E}[X_\alpha] .$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[\prod_{\alpha \in S} X_\alpha\right] &= \mathbb{E}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)}\left[\prod_{\alpha \in S} X_\alpha\right] \\
&= \int_{(x_1,\cdots,x_n) \in \mathbb{C}^n} x_1 \cdots x_n \mathrm{d}\mathbb{P}_{\left(X_{\alpha_1},\cdots,X_{\alpha_n}\right)}(x_1, \cdots, x_n) \\
&= \int_{(x_1,\cdots,x_n) \in \mathbb{C}^n} x_1 \cdots x_n \mathrm{d}\left(\prod_{j=1}^{n} \mathbb{P}_{X_{\alpha_j}}\right)(x_1, \cdots, x_n) && \text{(independence)} \\
&= \prod_{j=1}^{n} \int_{x_j \in \mathbb{C}} x_j \mathrm{d}\mathbb{P}_{X_{\alpha_j}}(x_j) && \text{(Fubini)} \\
&= \prod_{j=1}^{n} \mathbb{E}\left[X_{\alpha_j}\right] .
\end{aligned}$$

$\square$

Corollary 7.13. *Let A be some indexing set and $\{ X_\alpha : \Omega \to \mathbb{C} \}_{\alpha \in A}$ be an independent sequence of random variables. Then for any $\alpha, \beta \in A$,*

$$\mathbb{Cov}[X_\alpha, X_\beta] = 0\,.$$

Proof. We have

$$\begin{aligned}\mathbb{Cov}[X_\alpha, X_\beta] &= \mathbb{E}[X_\alpha X_\beta] - \mathbb{E}[X_\alpha]\,\mathbb{E}[X_\beta] \\ &\overset{\text{indep.}}{=} \mathbb{E}[X_\alpha]\,\mathbb{E}[X_\beta] - \mathbb{E}[X_\alpha]\,\mathbb{E}[X_\beta] \\ &= 0\,.\end{aligned}$$

$\square$

Example 7.14. The converse is false: there are random variables which are not independent yet their covariance is zero.

Thanks to Akshat for this counter example. Take X uniform in $[-1,1]$ and $Y := X^2$. Clearly X and Y are *not* independent since Y is a function of X. More formally, consider that

$$\begin{aligned}\mathbb{P}_{(X,Y)}[A \times B] &\equiv \mathbb{P}[(X,Y) \in A \times B] \\ &= \mathbb{P}[X \in A \wedge Y \in B] \\ &= \mathbb{P}\left[X \in A \wedge X^2 \in B\right] \\ &= \frac{1}{2}\int_{x\in[-1,1]} \chi_A(x)\,\chi_B\left(x^2\right)\mathrm{d}\lambda(x) \\ &\neq \left(\frac{1}{2}\int_{x\in[-1,1]} \chi_A(x)\,\mathrm{d}\lambda(x)\right)\left(\frac{1}{2}\int_{x\in[-1,1]} \chi_B\left(x^2\right)\mathrm{d}\lambda(x)\right)\end{aligned}$$

for all $A, B \in \mathfrak{B}(\mathbb{R})$ (take e.g. $A = \left[-1, -\frac{1}{2}\right]$ and $B = \left[0, \frac{1}{4}\right]$), so that $\mathbb{P}_{(X,Y)}$ is *not* the product measure, so that X, Y are not independent. On the other hand,

$$\begin{aligned}\mathbb{Cov}[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\ &= \mathbb{E}\left[X^3\right] - \mathbb{E}[X]\,\mathbb{E}\left[X^2\right] \\ &= 0\end{aligned}$$

since X is an odd function. $\square$

Claim 7.15. Functions of independent random variables are themselves independent random variables.

Proof. For the sake of simplicity we only show this for the case of two random variables, to illustrate the principle. Let $f, g : \mathbb{C} \to \mathbb{C}$ be measurable and $X, Y : \Omega \to \mathbb{C}$ be two random variables. Then we want to show that $f(X), g(Y)$ are also independent. To that end, if $B_1, B_2 \in \mathfrak{B}(\mathbb{C})$, then we want to show that

$$\begin{aligned}\mathbb{P}_{(f(X),g(Y))}[B_1 \times B_2] &\equiv \mathbb{P}[(f(X), g(Y)) \in B_1 \times B_2] \\ &= \mathbb{P}\left[(X,Y) \in f^{-1}(B_1) \times g^{-1}(B_2)\right] \\ &= \mathbb{P}\left[X \in f^{-1}(B_1)\right]\mathbb{P}\left[Y \in g^{-1}(B_2)\right] \\ &= \mathbb{P}[f(X) \in B_1]\,\mathbb{P}[g(Y) \in B_2] \\ &\equiv \mathbb{P}_{f(X)}[B_1]\,\mathbb{P}_{g(Y)}[B_2]\,.\end{aligned}$$

$\square$

Remark 7.16. Since a probability space is a *finite* measure space, in this particular context, we always have

$$L^p(\Omega\to\mathbb{C},\mathbb{P}) \quad\subseteq\quad L^q(\Omega\to\mathbb{C},\mathbb{P}) \qquad (1\leq q\leq p<\infty)\ .$$

Proof. Let $X\in L^p(\Omega\to\mathbb{C},\mathbb{P})$. Let $q\in[1,p]$. Then

$$\int_\Omega |X|^q\,\mathrm{d}\mathbb{P} = \int_\Omega \left(|X|^p\right)^{\frac{q}{p}}\mathrm{d}\mathbb{P}\,.$$

Now, the map $\alpha\mapsto\alpha^{\frac{q}{p}}$ is *concave*. As such, Jensen's inequality (Theorem 5.25) implies that

$$\begin{aligned}\int_\Omega \left(|X|^p\right)^{\frac{q}{p}}\mathrm{d}\mathbb{P} &\leq \left(\int_\Omega |X|^p\,\mathrm{d}\mathbb{P}\right)^{\frac{q}{p}}\\ &< \infty\,.\end{aligned}$$

□

In particular, $L^{n+m}\subseteq L^n$, so having higher moments guarantees the existence of lower moments, but not vice versa. This is in contrast to the $\mu(X)=\infty$ scenario,

Proposition 7.17. *The product of L^1 independent random variables is itself an L^1 random variable, and the expectation of the product is the product of expectations.*

This is in stark contrast to the usual case where if $f,g\in L^1$ then it is certainly far from obvious whether $fg\in L^1$ (consider for instance $f=g=[0,1]\ni x\mapsto\frac{1}{\sqrt{x}}$ which are both L^1 but fg is not L^1; of course, f,g are *not* independent).

Proof. Let $X,Y:\Omega\to\mathbb{C}$ be two independent L^1 random variables. Then we want to show that $XY\in L^1$ too.

$$\begin{aligned}\mathbb{E}\left[|XY|\right] &= \mathbb{E}\left[|X|\,|Y|\right]\\ &= \mathbb{E}_{(X,Y)}\left[|X|\,|Y|\right]\\ &= \left(\mathbb{E}_X\times\mathbb{E}_Y\right)\left[|X|\,|Y|\right]\\ &= \mathbb{E}_X\left[|X|\right]\mathbb{E}_Y\left[|Y|\right]\end{aligned}$$

where in the last step we invoked Tonelli's Theorem 5.14. The step $\mathbb{E}_{(X,Y)}=\mathbb{E}_X\times\mathbb{E}_Y$ is where we invoked independence. Once this is known, the same maneuver without the absolute values shows that the expectation of the product is the product of the expectations. □

Proposition 7.18. *The sum of L^2 independent random variables is itself in L^2, and the variance of the sum is the sum of the variances.*

Note that the sum of L^2 functions is always L^2 by the triangle inequality (regardless of independence):

$$\|f+g\|_{L^2}\leq\|f\|_{L^2}+\|g\|_{L^2}\,.$$

Proof. Let $X_1,\cdots,X_n:\Omega\to\mathbb{C}$ be independent L^2 random variables. Let $S:=\sum_{j=1}^n X_j$. The following calculation shows both that $S \in L^2$ and our claim:

$$\begin{aligned}
\mathbb{Var}[S] &= \mathbb{E}\left[S^2\right] - \mathbb{E}[S]^2 \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{n} X_j\right)^2\right] - \left(\sum_{j=1}^{n}\mathbb{E}[X_j]\right)^2 \\
&= \mathbb{E}\left[\sum_{j,l=1}^{n} X_j X_l\right] - \sum_{j,l=1}^{n}\mathbb{E}[X_j]\,\mathbb{E}[X_l] \\
&= \sum_{j,l=1}^{n}\mathbb{E}[X_j X_l] - \mathbb{E}[X_j]\,\mathbb{E}[X_l] \\
&= \sum_{j=1}^{n}\underbrace{\mathbb{E}\left[X_j^2\right] - \mathbb{E}[X_j]^2}_{=\mathbb{Var}[X_j]} + \sum_{j,l=1, j\neq l}^{n}\underbrace{\mathbb{E}[X_j X_l] - \mathbb{E}[X_j]\,\mathbb{E}[X_l]}_{=\mathbb{Cov}[X_j, X_l]} \\
&= \sum_{j=1}^{n}\mathbb{Var}[X_j]\,. && \text{(independence)}
\end{aligned}$$

□

7.3 Important inequalities

Theorem 7.19 (Markov)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable and $\varphi : \mathbb{R} \to [0, \infty)$ a non-decreasing function so that $\varphi \circ X \in L^1$. Then*

$$\varphi(a)\,\mathbb{P}[X \geq a] \leq \mathbb{E}[\varphi(X)] \qquad (a \in \mathbb{R})\,. \tag{7.4}$$

Proof. We have

$$\begin{aligned}
\mathbb{E}[\varphi(X)] &= \int_{\omega\in\Omega}\varphi(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) \\
&= \int_{x\in\mathbb{R}}\varphi(x)\,\mathrm{d}\mathbb{P}_X(x) \\
&= \int_{x<a}\varphi(x)\,\mathrm{d}\mathbb{P}_X(x) + \int_{x\geq a}\varphi(x)\,\mathrm{d}\mathbb{P}_X(x) \\
&\geq \int_{x\geq a}\varphi(x)\,\mathrm{d}\mathbb{P}_X(x) && (\varphi \geq 0) \\
&\geq \int_{x\geq a}\varphi(a)\,\mathrm{d}\mathbb{P}_X(x) && (\varphi \text{ is nondecreasing}) \\
&= \varphi(a)\,\mathbb{P}[X \geq a]\,.
\end{aligned}$$

□

Theorem 7.20 (Chebyshev)**.** *Let $X : \Omega \to \mathbb{R}$ be an L^1 random variable. Then*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\mathbb{Var}[X]}{\varepsilon^2} \qquad (\varepsilon > 0)\,. \tag{7.5}$$

Proof. We apply Markov's inequality to the variable $Y := (X - \mathbb{E}[X])^2$ with the function $\varphi(a) := a$ and the value

$a := \varepsilon^2$ to get

$$\begin{array}{rcl} \mathbb{P}\left[Y \geq \varepsilon^2\right] & \leq & \frac{\mathbb{E}[Y]}{\varepsilon^2} \\ & \updownarrow & \\ \mathbb{P}\left[(X-\mathbb{E}[X])^2 \geq \varepsilon^2\right] & \leq & \frac{\mathbb{Var}[X]}{\varepsilon^2} \\ & \updownarrow & \\ \mathbb{P}\left[|X-\mathbb{E}[X]| \geq \varepsilon\right] & \leq & \frac{\mathbb{Var}[X]}{\varepsilon^2} . \end{array}$$

□

Lemma 7.21 (Borel-Cantelli)**.** *Let* $\{ E_n \}_{n\in\mathbb{N}} \subseteq \mathrm{Msrbl}(\Omega)$ *be such that*

$$\sum_{n\in\mathbb{N}} \mathbb{P}[E_n] < \infty .$$

Then the probability that infinitely many of the E_n *occur is zero, i.e.,*

$$\mathbb{P}\left[\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k\right] = 0 .$$

We note that $\omega \in \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k$ iff $\omega \in E_k$ for infinitely many k's:

$$\begin{array}{rcl} \omega & \in & \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k \\ & \updownarrow & \\ \omega \in \bigcup_{k\geq n} E_k & \forall & n \in \mathbb{N} \\ & \updownarrow & \\ \forall n \in \mathbb{N} \exists k \geq n : \omega \in E_k . & & \end{array}$$

That set is denoted

$$\limsup_n E_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k .$$

One may also prove that

$$\limsup_n E_n = \left\{ \omega \in \Omega \,\middle|\, \limsup_n \chi_{E_n}(\omega) = 1 \right\}$$

Proof. Consider the sequence $F_N := \bigcup_{n\geq N} E_n$. It is non-increasing:

$$F_N \supseteq F_{N+1}$$

and

$$F_N \supseteq \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k .$$

Hence

$$\mathbb{P}\left[\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k\right] = \lim_{N\to\infty} \mathbb{P}[F_N] \leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \mathbb{P}[E_n] = 0 .$$

□

Lemma 7.22 (Second Borel-Cantelli)**.** *Let* $\{ E_n \}_{n\in\mathbb{N}} \subseteq \operatorname{Msrbl}(\Omega)$ *be* independent *such that*

$$\sum_{n\in\mathbb{N}} \mathbb{P}[E_n] = \infty\,.$$

Then

$$\mathbb{P}\left[\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} E_k\right] = 1\,.$$

Proof. We have

$$\begin{aligned} 1-\mathbb{P}\left[\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} E_k\right] &= \mathbb{P}\left[\left(\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} E_k\right)^c\right] \\ &= \mathbb{P}\left[\bigcup_{n\in\mathbb{N}}\bigcap_{k\geq n} E_k^c\right] \\ &= \lim_{n\to\infty}\mathbb{P}\left[\bigcap_{k\geq n} E_k^c\right] \qquad \text{(Approximation property)} \end{aligned}$$

Now, $\left\{ \bigcap_{k=n}^N E_k^c \right\}_N$ is a decreasing sequence of events whose limit is $\bigcap_{k=n}^\infty E_k^c$. Hence by the approximation property of measures (2.5) in reverse we find

$$\begin{aligned} 1-\mathbb{P}\left[\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} E_k\right] &= \lim_{n\to\infty}\mathbb{P}\left[\bigcap_{k=n}^{\infty} E_k^c\right] \\ &= \lim_{n\to\infty}\lim_{N\to\infty}\mathbb{P}\left[\bigcap_{k=n}^{N} E_k^c\right] \\ &= \lim_{n\to\infty}\lim_{N\to\infty}\prod_{k=n}^{N}\mathbb{P}[E_k^c] \qquad \text{(Independence)} \\ &= \lim_{n\to\infty}\lim_{N\to\infty}\prod_{k=n}^{N}(1-\mathbb{P}[E_k])\,. \end{aligned}$$

But since we assume $\sum_{n\geq 1}\mathbb{P}[E_n]=\infty$, this implies $\lim_{n\to\infty}\prod_{k\geq n}(1-\mathbb{P}[E_k])=0$ and we are done. Indeed,

$$\exp\left(\log\left(\prod_{k\geq n}(1-\mathbb{P}[E_k])\right)\right) \;=\; \exp\left(\sum_{k\geq n}\log\left((1-\mathbb{P}[E_k])\right)\right)\leq \exp\left(-\sum_{k\geq n}\mathbb{P}[E_k]\right)=0\,.$$

□

Lemma 7.23 (Kolmogorov's inequality)**.** *On a probability space* $(\Omega, \operatorname{Msrbl}(\Omega), \mathbb{P})$ *let* $\{ X_n : \Omega\to\mathbb{C} \}_{n\in\mathbb{N}}$ *be a sequence of independent random variables such that* $\mathbb{E}[X_n]=0$ *for all* $n\in\mathbb{N}$*. Set* $S_k := X_1+\dots+X_k$*. Then*

$$\mathbb{P}\left[\max_{k\in\{1,\cdots,n\}}|S_k|\geq\varepsilon\right]\leq\frac{1}{\varepsilon^2}\mathbb{Var}[S_n] \qquad (\varepsilon>0)\,.$$

Proof. We note

$$\left\{\max_{k\in\{1,\cdots,n\}}|S_k|\geq\varepsilon\right\} = \sqcup_{k=1}^n A_k$$

with $A_k := \{ |S_k| \geq \varepsilon \wedge |S_j| < \varepsilon \forall j < k \}$. Hence

$$\mathbb{P}\left[\max_{k\in\{1,\cdots,n\}} |S_k| \geq \varepsilon\right] = \sum_{k=1}^{n} \mathbb{P}[A_k] \leq \varepsilon^{-2} \sum_{k=1}^{n} \mathbb{E}\left[\chi_{A_k} S_k^2\right]$$

because on A_k, $S_k^2 \geq \varepsilon^2$. Now,

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] &\geq \sum_{k=1}^{n} \mathbb{E}\left[\chi_{A_k} S_n^2\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\chi_{A_k} \left(S_k + (S_n - S_k)\right)^2\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[\chi_{A_k} \left(S_k^2 + 2S_k (S_n - S_k) + (S_n - S_k)^2\right)\right] \\
&\geq \sum_{k=1}^{n} \mathbb{E}\left[\chi_{A_k} S_k^2\right] + 2\sum_{k=1}^{n} \mathbb{E}\left[\chi_{A_k} S_k (S_n - S_k)\right] .
\end{aligned}$$

Actually, $\mathbb{E}[\chi_{A_k} S_k (S_n - S_k)] = 0$. Indeed, $(S_n - S_k)$ is independent from $\chi_{A_k} S_k$. The first expression depends on $X_{k+1}, \cdots, X_n$ whereas the latter on $X_1, \cdots, X_k$. But $\mathbb{E}[S_n - S_k] = 0$. Hence

$$\mathbb{P}\left[\max_{k\in\{1,\cdots,n\}} |S_k| \geq \varepsilon\right] \leq \varepsilon^{-2} \mathbb{E}\left[S_n^2\right] = \varepsilon^{-2} \mathbb{V}\mathrm{ar}[S_n] .$$

□

7.4 Convergence of sums of random variables: LLN, CLT, LIL and all of that

On a probability space $(\Omega, \mathrm{Msrbl}(\Omega), \mathbb{P})$ we are given a sequence of random variables $\{ X_n : \Omega \to \mathbb{C} \}_{n\in\mathbb{N}}$ and we assume they are independent, as in (7.3)[4]. This situation could experimentally arise when we sample many measurements of some experiment, each time restarting the apparatus from scratch so as to obtain independence. We are then interested in the distribution of the *average* of the first N measurements:

$$A_N := \frac{1}{N} \sum_{n=1}^{N} X_n \qquad (N \in \mathbb{N}) .$$

Calculating $\mathbb{P}_{A_N}$ in and of itself is not that interesting (and also difficult). We are more interested in the behavior of $\mathbb{P}_{A_N}$ as $N \to \infty$. It turns out that as $N \to \infty$, A_N behaves quite deterministically. Just by linearity of the integral we have

$$\mathbb{E}[A_N] = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[X_n]$$

and by (7.1) and independence (Proposition 7.18),

$$\mathbb{V}\mathrm{ar}[A_N] = \frac{1}{N^2} \sum_{n=1}^{N} \mathbb{V}\mathrm{ar}[X_n] .$$

A natural thing is that all our random variables would be identically distributed, or at least, have the same variance σ^2. In this case,

$$\mathbb{V}\mathrm{ar}[A_N] = \frac{1}{N^2} \sum_{n=1}^{N} \sigma^2 = \frac{1}{N} \sigma^2 \to 0 .$$

If the variance of A_N becomes arbitrarily small, it is to be expected that it becomes arbitrarily close to a *constant* (i.e., deterministic) random variable:

$$A_N \overset{N\to\infty}{\approx} \mathbb{E}[A_N] = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[X_n] .$$

[4]For some of our statements independence may actually be dropped

For example, if all the expectations values are the same, $\mathbb{E}[X_n] =: \mu$, we get

$$A_N \approx \mu\,.$$

In this situation we expect

$$A_N \approx \mu + \frac{\sigma}{\sqrt{N}} \times (\text{random fluctuations of order 1 in } N) + \cdots ?\,.$$

The above rationale prompts us to define the following random variable

$$Z_N := \frac{A_N - \mu}{\sigma/\sqrt{N}}$$

which captures, presumably, scaled (order 1) fluctuations of A_N about its mean. The amazing thing is that the distribution of Z_N as $N \to \infty$ is *universal* and completely independent of $\mathbb{P}_{X_j}$. In this sense the standard distribution is "universal": regardless of the distribution of the sequence $\{ X_n \}_n$ we choose, just the structure of independence and some mild assumptions, we will establish:

$$Z_N \overset{N\to\infty}{\sim} \mathcal{N}(0,1)\,.$$

As we shall see, in principle one may continue the asymptotic expansion of A_N to further powers of $\frac{1}{N}$:

$$A_N =: \mu + \frac{\sigma}{\sqrt{N}} Z_N + R_N\,.$$

The terms in R_N however *do* depend on the distribution of $\mathbb{P}_{X_j}$ and are no longer universal, as we shall see (TODO CITE).

7.4.1 The law of large numbers

The above intuitive discussion was rather vague about *how* various random variables converge. We now want to make this precise. The first form of convergence we want to discuss is

Definition 7.24 (Convergence in probability)**.** Let $\{ Y_n \}_{n\in\mathbb{N}}$ be a sequence of random variables and let Y be some other random variable. We say that $Y_n \to Y$ in probability,

$$Y_n \overset{P}{\to} Y$$

iff

$$\mathbb{P}[|Y_n - Y| \geq \varepsilon] \overset{n\to\infty}{\to} 0 \qquad (\varepsilon > 0)\,.$$

Remark 7.25. Note that this notion is *identical* to the notion of *convergence in measure* which appeared in the homework, with the specification that the measure is a probability measure.

With this at hand, we want to make precise the statement $A_N \to \mu$:

Theorem 7.26 (Khinchin's theorem, the weak law of large numbers (LLN))**.** *On a probability space* $(\Omega, \mathrm{Msrbl}(\Omega), \mathbb{P})$ *let* $\{ X_n : \Omega \to \mathbb{C} \}_{n\in\mathbb{N}}$ *be a sequence of* L^2 *independent random variables such that*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{n=1}^{N} \mathbb{Var}[X_n] = 0\,.$$

In particular we are not *assuming* $\{ X_n \}_n$ *are identically distributed. Then*

$$A_N \overset{P}{\to} \mathbb{E}[A_N]\,.$$

in the sense that

$$\lim_{N\to\infty} \mathbb{P}[|A_N - \mathbb{E}[A_N]| < \varepsilon] = 1 \qquad (\varepsilon > 0)\,.$$

Proof. We invoke the Chebyshev inequality Theorem 7.20 on the random variable $B_N := A_N - \mathbb{E}[A_N]$ which has, by construction $\mathbb{E}[B_N] = 0$ and $\mathbb{Var}[B_N] = \frac{1}{N^2}\sum_{n=1}^{N}\mathbb{Var}[X_n]$. We obtain

$$\mathbb{P}[|B_N| \geq \varepsilon] \leq \frac{1}{\varepsilon^2 N^2}\sum_{n=1}^{N}\mathbb{Var}[X_n] .$$

□

This theorem may be strengthened by dropping the L^2 assumption and truncating random variables. This allows one to handle, e.g., Cauchy distributions. One possible phrasing of a generalization is as:

Theorem 7.27 (Better version of WLLN)**.** *On a probability space* $(\Omega, \mathrm{Msrbl}(\Omega), \mathbb{P})$ *let* $\{ X_n : \Omega \to \mathbb{C} \}_{n\in\mathbb{N}}$ *be a sequence of independent identically distributed random variables such that*

$$\lim_{x\to\infty} x\mathbb{P}[|X_n| > x] = 0 .$$

Then

$$A_N - \mathbb{E}\left[X_1 \chi_{[-N,N]}(X_1)\right] \overset{P}{\to} 0 \qquad (N\to\infty) .$$

The proof may be found in standard probability texts, e.g., [Dur19] (Theorem 2.2.7 in the Edition 4.1, April 21, 2013; Theorem 2.2.12 in Edition 5 online).

Next, we want to strengthen the mode of convergence of the LLN above. We shall build towards

Definition 7.28 (Convergence almost-surely)**.** Let $\{ Y_n \}_{n\in\mathbb{N}}$ be a sequence of random variables and let Y be some other random variable. We say that $Y_n \to Y$ almost-surely,

$$Y_n \overset{\text{a.s.}}{\to} Y$$

iff

$$\mathbb{P}\left[\lim_{n\to\infty} Y_n = Y\right] = 1 .$$

Claim 7.29. If $\{ Y_n \}_{n\in\mathbb{N}}$ converges almost-surely to Y then it converges in probability to Y.

Proof. Let $Y_n \to Y$ almost-surely. We want to show that $Y_n \to Y$ in probability, i.e., we want to show that

$$\lim_{n\to\infty} \mathbb{P}[|Y_n - Y| < \varepsilon] = 1 \qquad (\varepsilon > 0) .$$

To re-iterate, we are assuming that

$$\mathbb{P}\left[\lim_{n\to\infty}|Y_n - Y| = 0\right] = 1 .$$

Let $\varepsilon > 0$. Then $A_n := \bigcup_{m\geq n}\{ |Y_m - Y| \geq \varepsilon \}$ defines a decreasing sequence of sets towards $\bigcap_{n\in\mathbb{N}} A_n$. Hence

$$\lim_n \mathbb{P}[A_n] = \mathbb{P}\left[\bigcap_{n\in\mathbb{N}} A_n\right] .$$

But, $\mathbb{P}[A_n] = \mathbb{P}\left[\bigcup_{m\geq n}\{ |Y_m - Y| \geq \varepsilon \}\right] \geq \mathbb{P}[|Y_n - Y| \geq \varepsilon]$ by monotonicity. We find

$$\lim_{n\to\infty} \mathbb{P}[|Y_n - Y| \geq \varepsilon] \leq \mathbb{P}\left[\bigcap_{n\in\mathbb{N}} A_n\right] .$$

But $\mathbb{P}\left[\bigcap_{n\in\mathbb{N}} A_n\right] = 0$. Indeed, this follows from

$$\bigcap_{n\in\mathbb{N}} A_n \subseteq \left\{ \lim_{n\to\infty}|Y_n - Y| = 0 \right\}^c .$$

To see this, let $\omega \in \{ \lim_{n\to\infty} |Y_n - Y| = 0 \}$. Then for that ω, $|Y_n - Y| < \varepsilon$ for all $n \geq N_\varepsilon(\omega)$. For such n, $\omega \notin A_n$. But since $\mathbb{P}[\lim_{n\to\infty} |Y_n - Y| = 0] = 1$, the probability of the complement is zero, so we get our result. ☐

Theorem 7.30 (Kolmogorov's strong LLN)**.** *On a probability space* $(\Omega, \mathrm{Msrbl}(\Omega), \mathbb{P})$ *let* $\{ X_n : \Omega \to \mathbb{C} \}_{n\in\mathbb{N}}$ *be a sequence of* L^2 *independent random variables such that*

$$\lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{n^2} \mathbb{Var}[X_n] < \infty .$$

Then $\mathbb{P}$*-almost-surely,*

$$\lim_{N\to\infty} \left(A_N - \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[X_n] \right) = 0 ,$$

i.e., $\lim_{N\to\infty} \left(A_N - \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[X_n] \right) = 0$ *almost-surely, that is,*

$$\mathbb{P}\left[\lim_{N\to\infty} \left(A_N - \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[X_n] \right) = 0 \right] = 1 .$$

Proof. Define $B_N := A_N - \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[X_n]$. Then $\mathbb{E}[B_N] = 0$. Let $\varepsilon > 0$. For fixed $k \in \mathbb{N}$, $|B_n| \geq \varepsilon \exists n \in [2^{k-1}, 2^k]$ implies $\max_{n=1,\cdots,2^k} n|B_n| \geq \varepsilon 2^{k-1}$, so

$$\begin{aligned} \mathbb{P}\left[|B_n| \geq \varepsilon \exists n \in [2^{k-1}, 2^k] \right] &\leq \mathbb{P}\left[\max_{n=1,\cdots,2^k} n|B_n| \geq \varepsilon 2^{k-1} \right] && \text{(Monotonicity)} \\ &\leq \frac{1}{(\varepsilon 2^{k-1})^2} \sum_{n=1}^{2^k} \mathbb{Var}[X_n] && \text{(Kolomogrov)} \end{aligned}$$

Summing this inequality from $k = 1$ to ∞ we find

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}\left[|B_n| \geq \varepsilon \exists n \in [2^{k-1}, 2^k] \right] &\leq \sum_{k=1}^{\infty} \frac{1}{(\varepsilon 2^{k-1})^2} \sum_{n=1}^{2^k} \mathbb{Var}[X_n] \\ &= \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sum_{n=1}^{2^k} 2^{-2k} \mathbb{Var}[X_n] \\ &= \frac{4}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{k=(\log_2(n))}^{\infty} 2^{-2k} \mathbb{Var}[X_n] \\ &\leq \frac{8}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{Var}[X_n] < \infty . \end{aligned}$$

Hence

$$\mathbb{P}\left[\limsup_k \{ |B_n| \geq \varepsilon \exists n \in [2^{k-1}, 2^k] \} \right] = 0$$

by the Borel-Cantelli Lemma 7.21. But

$$\limsup_k \{ |B_n| \geq \varepsilon \exists n \in [2^{k-1}, 2^k] \} \equiv \{ |B_N| \geq \varepsilon \text{ for infinitely many } N \} .$$

Hence

$$\mathbb{P}\left[\liminf_N \{ |B_N| < \varepsilon \} \right] = 1 .$$

If we now take a (countable) limit of $\varepsilon \to 0$, we conclude that $\lim_N |B_N| = 0$ almost surely. ☐

7.4.2 The central limit theorem

We now return to the asymptotic expansion

$$A_N \approx \mu + \frac{\sigma}{\sqrt{N}} \times (\text{random fluctuations of order 1 in } N) + \cdots ?\,.$$

To study it we have defined

$$Z_N := \frac{A_N - \mu}{\sigma/\sqrt{N}}\,.$$

We shall see that

$$Z_N \overset{N\to\infty}{\sim} \mathcal{N}(0,1)\,.$$

The mode of convergence of $Z_N \to \mathcal{N}(0,1)$ will be *in distribution*.

Definition 7.31 (Convergence in distribution)**.** Let $\{ Y_n \}_{n\in\mathbb{N}}$ be a sequence of random variables and let Y be some other random variable. We say that $Y_n \to Y$ in distribution,

$$Y_n \overset{\mathrm{d}}{\to} Y$$

iff for all bounded continuous functions $f : \mathbb{C} \to \mathbb{C}$,

$$\lim_{n\to\infty} \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)]\,.$$

Claim 7.32. $\{ Y_n \}_n \to Y$ in distribution iff

$$\lim_{n\to\infty} \mathbb{P}[Y_n \geq t] = \mathbb{P}[Y \geq t]$$

pointwise in t, for all points t at which $t \mapsto \mathbb{P}[Y \geq t]$ is continuous.

Proof. See HW9Q3. ☐

Theorem 7.33 (Lévy's continuity theorem)**.** *Let $\{ Y_n \}_n$ be a sequence of real-valued random variables and Y be a real-valued random variable. $\{ Y_n \}_n \to Y$ in distribution iff $\mathbb{E}[\exp(\mathrm{i}tY_n)] \to \mathbb{E}[\exp(\mathrm{i}tY)]$ pointwise in t as $t \in \mathbb{C}$ and $t \mapsto \mathbb{E}[\exp(\mathrm{i}tY)]$ is continuous at $t = 0$.*

Proof. Let $f : \mathbb{C} \to \mathbb{C}$ be the image of the Fourier transform of some L^1 function $g : \mathbb{C} \to \mathbb{C}$. Then

$$f(y) = \int_{\psi\in\mathbb{R}} \exp(\mathrm{i}y\psi)\, g(\psi)\, \mathrm{d}\lambda(\psi)\,.$$

Then

$$\begin{aligned}\mathbb{E}[f(Y_n)] &= \mathbb{E}\left[\int_{\psi\in\mathbb{R}} \exp(\mathrm{i}Y_n\psi)\, g(\psi)\, \mathrm{d}\lambda(\psi)\right] \\ &= \int_{\psi\in\mathbb{R}} \mathbb{E}[\exp(\mathrm{i}Y_n\psi)]\, g(\psi)\, \mathrm{d}\lambda(\psi)\,.\end{aligned}$$

Taking now the limit $n \to \infty$ on both sides and using the dominated convergence theorem we find

$$\lim_{n\to\infty} \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)]\,.$$

We now approximate every continuous bounded f by such functions as above. For the details, see HW9Q4. ☐

Claim 7.34. Convergence in probability implies convergence in distribution.

Proof. TODO ☐

Hence, we have the following hierarchy of convergence modes:

Almost sure $\Longrightarrow L^p \Longrightarrow$ In probability $\Longrightarrow$ In distribution.

[5]

We shall prove

Theorem 7.35 (CLT)**.** *On a probability space* $(\Omega, \mathrm{Msrbl}(\Omega), \mathbb{P})$ *let* $\{ X_n : \Omega \to \mathbb{C} \}_{n \in \mathbb{N}}$ *be a sequence of independent* L^2 *random variables such that there exists some* $\delta > 0$ *with which*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^N \mathbb{E}\left[|X_n - \mathbb{E}[X_n]|^{2+\delta}\right]}{\left(\sqrt{\sum_{n=1}^N \mathbb{Var}[X_n]}\right)^{2+\delta}} = 0\,.$$

Then $Z_N \to Z$ *(where* $Z \sim \mathcal{N}(0,1)$*)* in distribution.

Proof. We shall calculate the characteristic function of Z_N. For convenience we denote $\overline{X_n} := X_n - \mathbb{E}[X_n]$ and $\sigma_n^2 := \mathbb{Var}[X_n]$,

$$\sigma := \sqrt{\frac{1}{N}\sum_{n=1}^N \mathbb{Var}[X_n]}$$

and finally, $Y_n := \frac{X_n - \mathbb{E}[X_n]}{\sigma}$. Note that $\{ Y_n \}_n$ is an independent sequence with $\mathbb{E}[Y_n] = 0$ and $\mathbb{E}\left[Y_n^2\right] = 1$. Then

$$\begin{aligned}
\varphi_{Z_N}(t) &\equiv \mathbb{E}[\exp(\mathrm{i}tZ_N)] \\
&= \mathbb{E}\left[\exp\left(\mathrm{i}t\sum_{n=1}^N \frac{Y_n}{\sqrt{N}}\right)\right] \\
&= \mathbb{E}\left[\prod_{n=1}^N \exp\left(\mathrm{i}t\frac{Y_n}{\sqrt{N}}\right)\right] \\
&= \prod_{n=1}^N \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{Y_n}{\sqrt{N}}\right)\right] && \text{(independence)} \\
&= \exp\left(\sum_{n=1}^N \log\left(\mathbb{E}\left[\sum_{\ell=0}^\infty \frac{1}{\ell!}(\mathrm{i}t)^\ell Y_n^\ell\right]\right)\right) \\
&= \exp\left(\sum_{n=1}^N \log\left(1 - \frac{t^2}{2N} + \frac{\mathrm{i}}{6N^{\frac{3}{2}}}t^3\mathbb{E}\left[Y_n^3\right] + \mathcal{O}\left(N^{-2}\right)\right)\right) \\
&= \exp\left(\sum_{n=1}^N \left(-\frac{t^2}{2N} + \frac{\mathrm{i}}{6N^{\frac{3}{2}}}t^3\mathbb{E}\left[Y_n^3\right] + \mathcal{O}\left(N^{-2}\right)\right)\right) \\
&= \exp\left(-\frac{t^2}{2}\right)\exp\left(\left(\frac{\mathrm{i}}{6N^{\frac{1}{2}}}t^3\left(\frac{1}{N}\sum_{n=1}^N \mathbb{E}\left[Y_n^3\right]\right) + \mathcal{O}\left(N^{-1}\right)\right)\right) \\
&= \exp\left(-\frac{t^2}{2}\right)\left(1 + \sum_{j=1}^\infty \frac{P_j(\mathrm{i}t)}{N^{\frac{j}{2}}}\right)
\end{aligned}$$

where P_j is some polynomial of degree $3j$ whose coefficients depend on the moments of the Y_n's. We find this

[5] In principle there is also the total variation convergence: $\|\mu - \nu\|_{\mathrm{TV}} := \sup_A |\mu(A) - \nu(A)|$ and we then ask that $\|\mathbb{P}_{X_n} - \mathbb{P}_X\|_{\mathrm{TV}} \to 0$. This mode of convergence implies convergence in distribution but no other implication involves total variation distance. We do not need $\|\cdot\|_{\mathrm{TV}}$ yet.

converges pointwise in t to

$$\varphi_{Z_N}(t) \to \exp\left(-\frac{1}{2}t^2\right).$$

The convergence of the characteristic function implies convergence in distribution by Lévy's continuity Theorem 7.33. TODO: make this compatible with the assumption that variables are not identically distributed. □

Corollary 7.36 ("Small" deviations tail bound from CLT)**.** *Let $\{ X_n : \Omega \to \mathbb{R} \}_{n\in\mathbb{N}}$ be a sequence of IID random variables with mean μ and variance σ^2. Let $A_N := \frac{1}{N}\sum_{n=1}^{N} X_n$ as before. Then, for every $t > 0$ there exists some $N_0(t)$ such that if $N \geq N_0(t)$*

$$\mathbb{P}\left[\left|\frac{A_N-\mu}{\sigma}\right| > \frac{t}{\sqrt{N}}\right] \leq \frac{1}{t}\exp\left(-\frac{1}{2}t^2\right). \tag{7.6}$$

Proof. We have

$$\frac{A_N-\mu}{\sigma} = \frac{1}{\sqrt{N}} Z_N$$

and Z_N converges in distribution to a standard normal RV. Hence,

$$\begin{aligned}\mathbb{P}\left[\left|\frac{A_N-\mu}{\sigma}\right| > \frac{t}{\sqrt{N}}\right] &= \mathbb{P}\left[\frac{1}{\sqrt{N}}|Z_N| > \frac{t}{\sqrt{N}}\right] \\ &= \mathbb{E}\left[\chi_{[-t,t]^c}(Z_N)\right].\end{aligned}$$

The CLT now implies that

$$\lim_{N\to\infty}\mathbb{E}\left[\chi_{[-t,t]^c}(Z_N)\right] = \mathbb{E}\left[\chi_{[-t,t]^c}(Z)\right]$$

where $Z \sim \mathcal{N}(0,1)$. For Z, we have

$$\begin{aligned}\mathbb{E}\left[\chi_{[-t,t]^c}(Z)\right] &= \frac{1}{\sqrt{2\pi}}\int_{z\in[-t,t]^c}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}\lambda(z) \\ &= \frac{2}{\sqrt{2\pi}}\int_{z=t}^{\infty}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}\lambda(z) \\ &\leq \frac{2}{\sqrt{2\pi}}\int_{z=t}^{\infty}\frac{z}{t}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}\lambda(z) && (1\leq \tfrac{z}{t}) \\ &= \frac{2}{t\sqrt{2\pi}}\int_{z=t}^{\infty}-\left[\partial_z\exp\left(-\frac{1}{2}z^2\right)\right]\mathrm{d}\lambda(z) \\ &= \frac{2}{t\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right). && \text{(FTC)}\end{aligned}$$

Hence,

$$\lim_{N\to\infty}\mathbb{P}\left[\left|\frac{A_N-\mu}{\sigma}\right| > \frac{t}{\sqrt{N}}\right] \leq \frac{2}{t\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right) \qquad (t>0).$$

□

Compare this with Hoeffding's inequality:

Theorem 7.37 (Heoffding)**.** *Let $\{ X_n : \Omega \to [a,b] \}_n$ be IID random variables and $A_N := \frac{1}{N}\sum_{n=1}^{N} X_n$. Then*

$$\mathbb{P}\left[\left|\frac{A_N-\mu}{b-a}\right| \geq t\right] \leq 2\exp\left(-2Nt^2\right) \qquad (t>0).$$

7.4.3 Higher order terms in the asymptotic expansion: an Edgeworth expansion

As we see in the above proof of the CLT, we could in principle continue the expansion of $\varphi_{Z_N}(t)$ (pointwise in t) to obtain any order in $N^{-\frac{1}{2}}$. Such an expansion is called an Edgeworth expansion. It expands the characteristic function φ_{Z_N} in terms of the characteristic function of the standard normal $t \mapsto \exp\left(-\frac{1}{2}t^2\right)$.

Once we have the characteristic function, we can invert it back to get an expansion for the distribution of Z_N:

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}_{Z_N}}{\mathrm{d}\lambda}(z) &= \frac{1}{2\pi}\int_{t\in\mathbb{R}}\exp(-\mathrm{i}tz)\,\varphi_{Z_N}(t)\,\mathrm{d}\lambda(t)\\ &= \frac{1}{2\pi}\int_{t\in\mathbb{R}}\exp(-\mathrm{i}tz)\exp\left(-\frac{1}{2}t^2\right)\left[1-\mathrm{i}\frac{1}{6}\frac{t^3}{N^{-\frac{1}{2}}}\mathbb{E}\left[Y_n^3\right]\right]\mathrm{d}\lambda(t)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\left[1-\frac{1}{6}\frac{1}{N^{-\frac{1}{2}}}\mathbb{E}\left[Y_n^3\right]z\left(z^2-3\right)+\mathcal{O}\left(N^{-1}\right)\right].\end{aligned}$$

Hence the distribution of Z_N *does* depend on $\mathbb{P}_X$, just, asymptotically this dependence converges to zero and we just get a standard normal variable for the zero order term in the asymptotic $N^{-\frac{1}{2}}$ expansion.

7.4.4 Law of iterated logarithm (LIL) [extra]

The tail bound derived from the CLT says that given $t>0$, there exists some $N_0(t)\in\mathbb{N}$ such that if $N\geq N_0(t)$, then

$$\mathbb{P}\left[|Z_N|>t\right]\leq\mathbb{P}\left[|Z|>t\right].$$

This situation does not preclude that almost-surely, as $N\to\infty$, Z_N becomes in fact unbounded.

Claim 7.38. Let $\{X_n:\Omega\to\mathbb{R}\}_n$ be a sequence of IID random variables whose mean is μ and variance is σ^2. Then, with the convention

$$A_N := \frac{1}{N}\sum_{n=1}^N X_n$$

and

$$A_N =: \mu+\frac{\sigma}{\sqrt{N}}Z_N$$

we have

$$\mathbb{P}\left[\limsup_{N\to\infty} Z_N=\infty\right]=1.$$

Proof. Let $Y_n := \frac{X_n-\mu}{\sigma}$. $S_N := Y_1+\dots+Y_N$. Let $M>0$. Let $m_1 := 2$ and $m_{k+1} := m_k^3$ for all $k\in\mathbb{N}$. Then $\frac{m_k}{\sqrt{m_{k+1}}}=m_k^{-\frac{1}{2}}\to 0$. Then

$$S_{m_{k+1}} = S_{m_k}+\underbrace{Y_{m_k+1}+\dots+Y_{m_{k+1}}}_{=:B_k}.$$

By construction, $\mathbb{E}[B_k]=0$ and $\mathbb{Var}[B_k]=m_{k+1}-m_k$ and $\{B_k\}_k$ are independent, so we may apply the CLT on it to get

$$\mathbb{P}\left[B_k>2M\sqrt{m_{k+1}}\right]\approx\mathbb{P}\left[Z>2M\right].$$

In writing this, we use $\frac{\sqrt{m_{k+1}}}{\sqrt{\mathbb{Var}[B_k]}}\to 1$. Hence, $\sum_k\mathbb{P}\left[B_k>2M\sqrt{m_{k+1}}\right]=\infty$. Thus, by Lemma 7.22,

$$\mathbb{P}\left[B_k>2M\sqrt{m_{k+1}}\text{ for infinitely many }k\text{'s}\right] = 1.$$

Now, by the strong LLN,

$$\frac{S_{m_k}}{\sqrt{m_{k+1}}}=\frac{S_{m_k}}{m_k}\frac{m_k}{\sqrt{m_{k+1}}}\to 0$$

almost-surely. So almost-surely, for sufficiently large k, $\left|\frac{S_{m_k}}{\sqrt{m_{k+1}}}\right|<M$.

Whenever both $B_k>2M\sqrt{m_{k+1}}$ and $\left|\frac{S_{m_k}}{\sqrt{m_{k+1}}}\right|<M$, we get

$$S_{m_{k+1}}=S_{m_k}+B_k\geq B_k-|S_{m_k}|\geq 2M\sqrt{m_{k+1}}-M\sqrt{m_{k+1}}=M\sqrt{m_{k+1}}.$$

Hence,

$$Z_{m_{k+1}} \equiv \frac{S_{m_{k+1}}}{\sqrt{m_{k+1}}} > M\,.$$

This event can be obtained as the countable intersection of the two almost-sure events, so we are done. □

So if Z_N almost-surely grows to $\pm\infty$, can we characterize how quickly?

Theorem 7.39. *Let $\{ X_n : \Omega \to \mathbb{R} \}_n$ be a sequence of IID random variables whose mean is μ and variance is σ^2. Then, with the convention*

$$A_N := \frac{1}{N} \sum_{n=1}^{N} X_n$$

and

$$A_N =: \mu + \frac{\sigma}{\sqrt{N}} Z_N$$

we have

$$\mathbb{P}\left[\limsup_{N\to\infty} \frac{Z_N}{\sqrt{2\log\left(\log\left(N\right)\right)}} = 1\right] = 1 \qquad \wedge \qquad \mathbb{P}\left[\liminf_{N\to\infty} \frac{Z_N}{\sqrt{2\log\left(\log\left(N\right)\right)}} = -1\right] = 1\,.$$

I.e., almost-surely, the maximum of $|Z_N|$ grows like $\sqrt{2\log\left(\log\left(N\right)\right)}$ as $N \to \infty$.

Note the contrast between the CLT and the LIL. CLT says that for a *single arbitrarily large* N, $\mathbb{P}_{Z_N} \approx \mathbb{P}_Z$ with $Z \sim \mathcal{N}(0,1)$. On the other hand, the LIL says that when taken as a whole, the sequence $\{ |Z_N| \}_N$ grows, almost-surely, like $\sqrt{2\log\left(\log\left(N\right)\right)}$, which is *very* slow.

Proof of Theorem 7.39. TODO □

7.5 Large deviations [Varadhan]

In this section we follow Varadhan's lecture notes [Var84]; see also [DZ09].

To motivate the "large deviations" question, recall the approximation theorem due to Laplace:

Theorem 7.40 (Laplace asymptotics)**.** *Fix some $n \in \mathbb{N}$. Let $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{C}$ be given. Assume that f has continuous Hessian*

$$\mathbb{H}f : \mathbb{R}^n \to \mathrm{Mat}_{n\times n}\left(\mathbb{R}\right)$$

at some $x_0 \in \mathbb{R}^n$ and g is continuous and non-vanishing at x_0. Assume further that

$$\left(\nabla f\right)\left(x_0\right) = 0 \wedge \left(\mathbb{H}f\right)\left(x_0\right) > 0\,.$$

Finally, assume that there exists some $\eta_\star > 0$ such that

$$\int_{x\in\mathbb{R}^n} \mathrm{e}^{-\eta_\star f(x)} g\left(x\right) \mathrm{d}\lambda\left(x\right) < \infty\,.$$

Then

$$\lim_{\eta\to\infty} \frac{\int_{x\in\mathbb{R}^n} \mathrm{e}^{-\eta f(x)} g\left(x\right) \mathrm{d}\lambda\left(x\right)}{\eta^{-\frac{n}{2}} \mathrm{e}^{-\eta f(x_0)}} = \frac{g\left(x_0\right)}{\sqrt{\det\left(\frac{1}{2\pi}\left(\mathbb{H}f\right)\left(x_0\right)\right)}}\,.$$

In particular,

$$\lim_{\eta\to\infty} -\frac{1}{\eta}\log\left(\int_{x\in\mathbb{R}^n} \mathrm{e}^{-\eta f(x)} g\left(x\right) \mathrm{d}\lambda\left(x\right)\right) = f\left(x_0\right)\,.$$

The proof of this classical result may be found, e.g., in [Sha23a].

We want to ask the following basic question: if we have a sequence of probability measures $\mathbb{P}_\eta$ on a fixed measurable space $\left(\Omega, \mathrm{Msrbl}\left(\Omega\right)\right)$ and X is some fixed random variable $X : \Omega \to \mathbb{C}$, in analogy with can we compute

$$\lim_{\eta\to\infty} -\frac{1}{\eta}\log\left(\mathbb{E}_\eta\left[g\left(X\right)\right]\right) = ?$$

What about

$$\lim_{\eta\to\infty} -\frac{1}{\eta}\log\left(\mathbb{E}_\eta\left[g_\eta\left(X\right)\right]\right) = ?$$

To see why this might make sense, let us consider the case of IID random variables from the previous section, $\{ X_n \}_{n\in\mathbb{N}}$ whose mean is μ and variance σ^2; with

$$A_N \equiv \frac{1}{N}\sum_{n=1}^{N} X_n$$

we found that

$$Z_N := \frac{A_N-\mu}{\sigma/\sqrt{N}}$$

asymptotically behaves like a standard normal. This means that the tail bound behaves like

$$\begin{aligned}
\mathbb{P}\left[\left|\frac{A_N-\mu}{\sigma}\right|\geq t\right] &= \mathbb{P}\left[\left|\frac{1}{\sqrt{N}}Z_N\right|\geq t\right] \\
&= \mathbb{P}\left[|Z_N|\geq\sqrt{N}t\right] \\
&\overset{\star}{=} \int_{|z|\geq\sqrt{N}t}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z \\
&= 2\int_{z\in\left[\sqrt{N}t,\infty\right)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z \\
&= 2\int_{z\in[0,\infty)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(z+\sqrt{N}t\right)^2\right)\mathrm{d}z \\
&= 2\int_{z\in[0,\infty)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}N\left(\frac{1}{\sqrt{N}}z+t\right)^2\right)\mathrm{d}z \\
&= 2\frac{\sqrt{N}}{\sqrt{2\pi}}\int_{z\in[0,\infty)}\exp\left(-\frac{1}{2}N\left(z+t\right)^2\right)\mathrm{d}z\,.
\end{aligned}$$

Remark 7.41. Note that the step $\star$ above is *not* justified but only heuristic because we may only apply the CLT with *fixed* t, which does not depend on N.

Employing the Laplace asymptotics cited above yields

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\log\left(\mathbb{P}\left[\left|\frac{A_N-\mu}{\sigma}\right|\geq t\right]\right) &= -\frac{1}{2}t^2 \\
&\downarrow \\
\mathbb{P}\left[\left|\frac{A_N-\mu}{\sigma}\right|\geq t\right] &\approx \mathrm{e}^{-\frac{1}{2}Nt^2}\,.
\end{aligned} \tag{7.7}$$

We thus expect that the sequence of probability measures $\left\{ \mathbb{P}_{\frac{1}{\sqrt{N}}Z_N} \right\}_{N\in\mathbb{N}}$ will behave, asymptotically, in an *exponentially decaying way.*

After these intuitive motivation, let us now make this more precise. In what follows $(\Omega, \mathrm{Msrbl}(\Omega))$ is a fixed measurable space. For simplicity we assume that the ambient space Ω is a complete separable metric space and $\mathrm{Msrbl}(\Omega) = \mathfrak{B}(\Omega)$. We switch from the asymptotic parameter $\eta\to\infty$ to $\varepsilon\to 0^+$, with the rule $\varepsilon := \frac{1}{\eta}$. We thus consider sequences of probability measures $\mathbb{P}_\varepsilon : \mathfrak{B}(\Omega)\to[0,1]$ parametrized by a continuous parameter $\varepsilon>0$. We want to formalize the relation

$$\mathbb{P}_\varepsilon[A] \sim \sup_{a\in A}\mathrm{e}^{-\frac{1}{\varepsilon}I(a)} \qquad \left(\varepsilon\to 0^+\right)$$

for some function $I:\Omega\to[0,\infty]$. To that end, let us make the

Definition 7.42 (Rate function)**.** A rate function $I:\Omega\to[0,\infty]$ is some lower semicontinuous map $\Omega\to[0,\infty]$ (i.e., for any $t\in(0,\infty)$), $I^{-1}((t,\infty])\in\mathrm{Open}(\Omega)$ or equivalently, $I^{-1}([0,t])\in\mathrm{Closed}(\Omega)$ for any $t\in(0,\infty)$) such that for any $\ell\in(0,\infty)$,

$$I^{-1}([0,\ell])\in\mathrm{Compact}(\Omega)\,.$$

Remark 7.43. Varadhan uses this notion whereas other authors only require a rate function to be lower semicontinuous whereas the additional requirement on compactness is called a "good rate function" (see e.g. Dembo's textbook [DZ09]). The difference between the two offers some technical advantages down the road. For now we stick with Varadhan's simpler phrasing.

Definition 7.44 (Large deviations principle)**.** Let $\{ \mathbb{P}_\varepsilon \}_{\varepsilon>0}$ be a sequence of probability measures. We say that $\{ \mathbb{P}_\varepsilon \}_{\varepsilon>0}$ *obeys a large deviations principle with rate function* $I : \Omega \to [0,\infty]$ *iff*

1. For any $F \in \text{Closed}(\Omega)$,
$$\limsup_{\varepsilon\to 0^+} \varepsilon \log (\mathbb{P}_\varepsilon [F]) \leq - \inf_{\omega\in F} I(\omega) \ .$$
2. For any $U \in \text{Open}(\Omega)$,
$$\liminf_{\varepsilon\to 0^+} \varepsilon \log (\mathbb{P}_\varepsilon [U]) \geq - \inf_{\omega\in U} I(\omega) \ .$$

Claim 7.45. Let $\{ \mathbb{P}_\varepsilon \}_{\varepsilon>0}$ be a sequence of probability measures that obeys a large deviations principle with rate function I as above. If $A \in \mathfrak{B}(\Omega)$ is such that

$$\inf_{\omega\in A^\circ} I(\omega) = \inf_{\omega\in A} I(\omega) = \inf_{\omega\in \bar{A}} I(\omega) \tag{7.8}$$

then

$$\lim_{\varepsilon\to 0^+} \varepsilon \log (\mathbb{P}_\varepsilon [A]) = - \inf_{\omega\in A} I(\omega) \ .$$

Proof. By definition we have

$$\limsup_{\varepsilon\to 0^+} \varepsilon \log \left(\mathbb{P}_\varepsilon \left[\overline{A}\right]\right) \leq - \inf_{\omega\in \overline{A}} I(\omega) = - \inf_{\omega\in A^\circ} I(\omega) \leq \liminf_{\varepsilon\to 0^+} \varepsilon \log (\mathbb{P}_\varepsilon [A^\circ]) \ .$$

Moreover, since $A^\circ \subseteq A \subseteq \overline{A}$, we always have $\mathbb{P}_\varepsilon [A^\circ] \leq \mathbb{P}_\varepsilon [A] \leq \mathbb{P}_\varepsilon \left[\overline{A}\right]$. Hence

$$\limsup_{\varepsilon\to 0^+} \varepsilon \log (\mathbb{P}_\varepsilon [A]) \leq \liminf_{\varepsilon\to 0^+} \varepsilon \log (\mathbb{P}_\varepsilon [A])$$

so the limit exists and equals $-\inf_{\omega\in A} I(\omega)$ as desired. □

In particular, if $\Omega = \mathbb{R}^n$ and $I : \Omega \to [0,\infty]$ is monotone increasing away from the origin, we have

$$\mathbb{P}_\varepsilon [B_t(0)^c] \sim \exp\left(-\frac{1}{\varepsilon} \inf_{\omega\in B_t(0)^c} I(\omega)\right) = \exp\left(-\frac{1}{\varepsilon} I(t)\right) \ .$$

Example 7.46. Let Z be a standard normal RV, i.e., $Z \sim \mathcal{N}(0,1)$. Then, we claim, the sequence of measures $\left\{ \mathbb{P}_{\sqrt{\varepsilon}Z} \right\}_{\varepsilon>0}$ obeys an LDP with the rate function $I(z) = \frac{1}{2}z^2$. To see this, first we note that clearly $I : \mathbb{R} \to [0,\infty]$ is continuous so it is lower semicontinuous, and moreover,

$$I^{-1}([0,t]) = \left[-\sqrt{2t}, \sqrt{2t}\right] \in \text{Compact}(\mathbb{R}) \ .$$

So I is indeed a rate function. Moreover, let $C\in\text{Closed}(\mathbb{R})$. Then

$$\begin{aligned}\mathbb{P}_{\sqrt{\varepsilon}Z}[C] &\equiv \mathbb{P}\left[\sqrt{\varepsilon}Z\in C\right]\\ &= \mathbb{P}\left[Z\in\frac{1}{\sqrt{\varepsilon}}C\right]\\ &= \int_{z\in\mathbb{R}}\chi_{\frac{1}{\sqrt{\varepsilon}}C}(z)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}\lambda(z)\\ &= \frac{1}{\sqrt{2\pi\varepsilon}}\int_{z\in\mathbb{R}}\chi_C(z)\exp\left(-\frac{1}{2\varepsilon}z^2\right)\mathrm{d}\lambda(z)\\ &\leq \frac{1}{\sqrt{2\pi\varepsilon}}\int_{z\in\mathbb{R}}\chi_C(z)\exp\left(-\frac{1}{2\varepsilon}\inf_{z\in C}z^2\right)\mathrm{d}\lambda(z)\\ &= \frac{\exp\left(-\frac{1}{2\varepsilon}\inf_{z\in C}z^2\right)}{\sqrt{2\pi\varepsilon}}\lambda(C)\,.\end{aligned}$$

Hence

$$\limsup_{\varepsilon\to0^+}\varepsilon\log(\mathbb{P}_\varepsilon[C])\leq-\frac{1}{2}\inf_{z\in C}z^2=-\inf_{z\in C}I(z)\,.$$

Moreover, if $U\in\text{Open}(\mathbb{R})$ and $\delta>0$, let $z_\delta\in U$ be such that

$$I(z_\delta)-\delta<\inf_{z\in U}I(z)\,.$$

We may choose some $r>0$ such that $B_r(z_\delta)\subseteq U$ and $I(B_r(z_\delta))\subseteq B_{2\delta}(\inf_{z\in U}I(z))$. Then,

$$\begin{aligned}\mathbb{P}_{\sqrt{\varepsilon}Z}[U] &= \frac{1}{\sqrt{2\pi\varepsilon}}\int_{z\in U}\exp\left(-\frac{1}{2\varepsilon}z^2\right)\mathrm{d}\lambda(z)\\ &\geq \frac{1}{\sqrt{2\pi\varepsilon}}\int_{z\in B_r(z_\delta)}\exp\left(-\frac{1}{\varepsilon}\inf_{z\in U}I(z)-\frac{1}{\varepsilon}2\delta\right)\mathrm{d}\lambda(z)\\ &= \frac{2r}{\sqrt{2\pi\varepsilon}}\exp\left(-\frac{1}{\varepsilon}\inf_{z\in U}I(z)-\frac{1}{\varepsilon}2\delta\right)\,.\end{aligned}$$

Hence

$$\liminf_{\varepsilon\to0^+}\varepsilon\log(\mathbb{P}_\varepsilon[U])=-\inf_{z\in U}I(z)-2\delta\,.$$

Since $\delta>0$ was arbitrary we get the result.

Lemma 7.47 (Varadhan's lemma)**.** *Let $\mathbb{P}_\varepsilon$ satisfy the LDP with a rate function I. Then for any bounded continuous random variable $X:\Omega\to\mathbb{R}$*

$$\lim_{\varepsilon\to0^+}\varepsilon\log\left(\mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\right]\right)=\sup_{\omega\in\Omega}(X(\omega)-I(\omega))\,.$$

Proof. Let $M>0$. Note that $I:\Omega\to[0,\infty]$ is itself a random variable. Hence

$$\mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\right] = \mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\chi_{I\leq M}\right]+\mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\chi_{I>M}\right]\,.$$

Since X is bounded, we get

$$\mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\chi_{I>M}\right] \leq \exp\left(\frac{1}{\varepsilon}\|X\|_\infty\right)\mathbb{P}_\varepsilon[I>M]\,.$$

By the fact that $\mathbb{P}_\varepsilon$ obeys an LDP, we have

$$\mathbb{P}_\varepsilon[I>M]\lesssim\mathrm{e}^{-\frac{1}{\varepsilon}M}\,.$$

Hence we get

$$\mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\chi_{I>M}\right]\lesssim\exp\left(-\frac{1}{\varepsilon}\left(M-\|X\|_\infty\right)\right).$$

Eventually we shall take $M\to\infty$ but for now all we need is that

$$M\gg\|X\|_\infty+\sup_{\omega\in\Omega}\left(X\left(\omega\right)-I\left(\omega\right)\right)$$

so that

$$\lim_{\varepsilon\to0^+}\frac{\exp\left(-\frac{1}{\varepsilon}\left(M-\|X\|_\infty\right)\right)}{\exp\left(-\frac{1}{\varepsilon}\left(\sup_{\omega\in\Omega}\left(X\left(\omega\right)-I\left(\omega\right)\right)\right)\right)}\to0\,. \tag{7.9}$$

Let us now study the other term. We prove first an upper bound. Let $\delta>0$. By assumption, $K_M:=\{\,I\leq M\,\}$ is compact, so X is uniformly continuous on it. As such, there exists some $\eta>0$ such that if $\omega_1,\omega_2\in K_M$ are such that

$$|\omega_1-\omega_2|<\eta$$

then

$$|X\left(\omega_1\right)-X\left(\omega_2\right)|<\delta\,.$$

Since K_M is compact, the open cover $\left\{\,B_{\frac{1}{2}\eta}\left(\omega\right)\,\right\}_{\omega\in K_M}$ admits a finite sub-cover $\left\{\,B_{\frac{1}{2}\eta}\left(\omega_j\right)\,\right\}_{j=1,\cdots,n}$. The collection

$$\left\{\,C_j:=\overline{B_{\frac{1}{2}\eta}\left(\omega_j\right)}\,\right\}_{j=1,\cdots,n}$$

still covers K_M (it is just bigger), each C_j is closed, and has diameter less than η, so obeys uniform continuity of X. Then

$$\begin{aligned}\int_{\omega\in K_M}\exp\left(\frac{1}{\varepsilon}X\left(\omega\right)\right)\mathrm{d}\mathbb{P}_\varepsilon\left(\omega\right)&\leq\sum_{j=1}^n\int_{\omega\in C_j}\exp\left(\frac{1}{\varepsilon}X\left(\omega\right)\right)\mathrm{d}\mathbb{P}_\varepsilon\left(\omega\right)\\&\leq\sum_{j=1}^n\int_{\omega\in C_j}\exp\left(\frac{1}{\varepsilon}\left(X_j+\delta\right)\right)\mathrm{d}\mathbb{P}_\varepsilon\left(\omega\right)\end{aligned}$$

where $X_j:=\inf_{\omega\in C_j}X\left(\omega\right)$. Hence

$$\begin{aligned}\limsup_{\varepsilon\to0^+}\varepsilon\log\left(\int_{\omega\in\Omega}\exp\left(\frac{1}{\varepsilon}X\left(\omega\right)\right)\mathrm{d}\mathbb{P}_\varepsilon\left(\omega\right)\right)&=\limsup_{\varepsilon\to0^+}\varepsilon\log\left(\int_{\omega\in K_M}\exp\left(\frac{1}{\varepsilon}X\left(\omega\right)\right)\mathrm{d}\mathbb{P}_\varepsilon\left(\omega\right)\right)\qquad(\text{Using (7.9)})\\&\leq\sup_{j\in\{\,1,\cdots,n\,\}}\left(X_j+\delta-\inf_{\omega\in C_j}I\left(\omega\right)\right)\\&\leq\sup_{j\in\{\,1,\cdots,n\,\}}\sup_{\omega\in C_j}\left(X\left(\omega\right)-I\left(\omega\right)\right)+\delta\\&=\sup_{\omega\in K_M}\left(X\left(\omega\right)-I\left(\omega\right)\right)+\delta\,.\\&\leq\sup_{\omega\in\Omega}\left(X\left(\omega\right)-I\left(\omega\right)\right)+\delta\,.\end{aligned}$$

Since $\delta>0$ was arbitrary, we are done.

For the lower bound, given some $\delta>0$, by the approximation property of the supremum, there is a point $\tilde{\omega}\in K_M$ such that

$$X\left(\tilde{\omega}\right)-I\left(\tilde{\omega}\right)\geq\sup_{\omega\in K_M}\left(X\left(\omega\right)-I\left(\omega\right)\right)-\frac{\delta}{2}\,.$$

By continuity of X, we may find a neighborhood $U\subseteq K_M$ of $\tilde{\omega}$ such that $X(\omega)\geq X(\tilde{\omega})-\frac{1}{2}\delta$ for all $\omega\in U$. Then

$$\begin{aligned}\liminf_{\varepsilon\to0^+}\varepsilon\log\left(\int_{\omega\in\Omega}\exp\left(\frac{1}{\varepsilon}X(\omega)\right)\mathrm{d}\mathbb{P}_\varepsilon(\omega)\right) &\geq \liminf_{\varepsilon\to0^+}\varepsilon\log\left(\int_{\omega\in K_M}\exp\left(\frac{1}{\varepsilon}X(\omega)\right)\mathrm{d}\mathbb{P}_\varepsilon(\omega)\right)\\ &\geq \liminf_{\varepsilon\to0^+}\varepsilon\log\left(\int_{\omega\in U}\exp\left(\frac{1}{\varepsilon}X(\omega)\right)\mathrm{d}\mathbb{P}_\varepsilon(\omega)\right)\\ &\geq X(\tilde{\omega})-\frac{\delta}{2}-\inf_{\omega\in U}I(\omega)\\ &\geq X(\tilde{\omega})-I(\tilde{\omega})-\frac{\delta}{2}\\ &\geq \sup_{\omega\in K_M}(X(\omega)-I(\omega))-\delta\,.\end{aligned}$$

Again, since $\delta>0$ and M are arbitrary, we are done. □

We illustrate the utility of these notions by applying them to the sum of IID random variables which we saw before in the CLT section, Corollary 7.36.

Example 7.48 (CLT Improvement via large deviations, Cramer's theorem)**.** Let $\{X_n\}_n$ be IID random variables with mean μ and standard deviation $\sigma>0$. Define

$$A_N:=\frac{1}{N}\sum_{n=1}^{N}X_n=:\mu+\frac{\sigma}{\sqrt{N}}Z_N\,.$$

We want to revisit Section 7.5.

We thus want to establish that the sequence of probability measures $\mathbb{P}_{A_N}$ obeys an LDP with rate function (*Cramer's function*)

$$I(x):=\sup_{\theta\in\mathbb{R}}\left(\theta x-\log\left(\mathbb{E}\left[\mathrm{e}^{\theta X_1}\right]\right)\right)\qquad(x\in\mathbb{R})\;;$$

(recall that $\theta\mapsto\log\left(\mathbb{E}\left[\mathrm{e}^{\theta X_1}\right]\right)=:\kappa(\theta)$ is the *cumulant generating function*). The mapping

$$\{\theta\mapsto\kappa(\theta)\}\mapsto\left\{x\mapsto\sup_{\theta\in\mathbb{R}}(\theta x-\kappa(\theta))\right\}$$

is called a *Legendre transform*. The Legendre transform is always lower semicontinuous (prove this). For instance, for standard normal,

$$\kappa(\theta)=\log\left(\mathrm{e}^{\frac{1}{2}\theta^2}\right)=\frac{1}{2}\theta^2$$

and then

$$I(x)=\sup_{\theta\in\mathbb{R}}\left(\theta x-\frac{1}{2}\theta^2\right)=\frac{1}{2}x^2\,.$$

Moreover, we have

$$\begin{aligned}I(x) &\geq \sup_{|\theta|\leq a}(\theta x-\kappa(\theta))\\ &\geq a|x|-\sup_{|\theta|\leq a}\kappa(\theta)\,.\end{aligned}$$

Presumably $\mathbb{P}_{X_1}$ is sufficiently nice so that $M_a:=\sup_{|\theta|\leq a}\kappa(\theta)$ is finite for some $a>0$. If that is the case, then

$$I(x)\geq t$$

for all $|x|\geq\frac{t+M_a}{a}$ so that $\{I\leq t\}$ is necessarily bounded. Since it is a closet subset of $\mathbb{R}$ it is compact, and that makes I a rate function.

We now prove that $\mathbb{P}_{A_N}$ obeys an LDP with Cramer's rate function. For convenience, let us instead work with the normalized

$$B_N:=\frac{A_N-\mu}{\sigma}=\frac{1}{\sqrt{N}}Z_N\,.$$

Let now $F \in \text{Closed}(\mathbb{R})$ and $\theta \in \mathbb{R}$. Then

$$\begin{aligned}\mathbb{P}[B_N \in F] &= \mathbb{P}[B_N = c \quad \exists c \in F] \\ &\leq \mathbb{P}[B_N \geq c \quad \exists c \in F] \\ &= \mathbb{P}\left[\mathrm{e}^{N\theta B_N} \geq \mathrm{e}^{N\theta c} \text{ for some } c \in F\right] \\ &\leq \mathrm{e}^{-N\theta c}\mathbb{E}\left[\mathrm{e}^{N\theta B_N}\right] \qquad \text{(Markov)} \\ &= \mathrm{e}^{-N\theta c}\mathrm{e}^{N\kappa(\theta)} = \mathrm{e}^{-N(\theta c - \kappa(\theta))}\,.\end{aligned}$$

since this holds for *every* θ and every $c \in F$, we get

$$\frac{1}{N}\log(\mathbb{P}[B_N \in F]) \leq -\sup_\theta \inf_{c\in F}(\theta c - \kappa(\theta)) = -\inf_{c\in F}\sup_\theta(\theta c - \kappa(\theta)) = -\inf_{c\in F} I(c)\,.$$

TODO: complete the lower bound on open sets.

Now that we have the LDP for $\{ A_N \}$, we get an improvement of (7.6): Let us assume that for any t, $[\mu - t\sigma, \mu + t\sigma]^c$ is such that (7.8) is obeyed. Then

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\log\left(\mathbb{P}\left[\left|\frac{A_N - \mu}{\sigma}\right| > t\right]\right) &= \lim_{N\to\infty}\frac{1}{N}\log\left(\mathbb{P}\left[A_N \in [\mu - t\sigma, \mu + t\sigma]^c\right]\right) \\ &= -\inf_{\theta\in[\mu - t\sigma, \mu + t\sigma]^c} I(\theta)\,.\end{aligned}$$

We note in passing that the LDP has also been established for Brownian motion, see [Sch66].

7.6 The Kolmogorov extension theorem [Biskup]

This is also known as the Kolmogorov existence theorem, the Kolmogorov consistency theorem or the Daniell-Kolmogorov theorem. We follow mostly Folland and [Bis].

Let $(\Omega, \text{Msrbl}(\Omega), \mathbb{P})$ be some probability space and A be *any* non-empty index set. Usually we suppose that for any $\alpha \in A$, we are given a random variable $X_\alpha : \Omega \to \mathbb{R}$. Now we are interested in the following "inverse" problem: given all laws of all random variables, can we build a mutual probability space $(\Omega, \text{Msrbl}(\Omega), \mathbb{P})$ where all X_α are random variables on this mutual "sample space"?

Let us make this more precise: given the distributions $\mathbb{P}_{X_\alpha}$ for any $\alpha \in A$ (these are probability measures on $\mathbb{R}$), can we reconstruct $(\Omega, \text{Msrbl}(\Omega), \mathbb{P})$, whose marginals are $\{ \mathbb{P}_{X_\alpha} \}_{\alpha\in A}$? This question is still too simplistic because sometimes we may want to encode the dependence structure between the various random variables. So we should rather ask, the following: For any $n < \infty$ and any *injective* map

$$\alpha : \{ 1, \cdots, n \} \to A$$

we are given $\mathbb{P}_{(X_{\alpha_1}, \cdots, X_{\alpha_n})}$ as a probability measure on $\mathbb{R}^n$. Can we then reconstruct $(\Omega, \text{Msrbl}(\Omega), \mathbb{P})$ so that for any such α, $\mathbb{P}_{(X_{\alpha_1}, \cdots, X_{\alpha_n})}$ is indeed the marginal of $\mathbb{P}$?

There is an obvious solution to this problem in the following special case: say that all variables are independent and $|A| < \infty$. Then we can take

$$\Omega := \mathbb{R}^{|A|}, \qquad \text{Msrbl}(\Omega) := \mathfrak{B}\left(\mathbb{R}^{|A|}\right), \qquad \mathbb{P} := \prod_{\alpha\in A}\mathbb{P}_{X_\alpha}\,.$$

In principle this product structure is how we want to think of the underlying probability space for *any* sequence of random variables, i.e., the random variables are the coordinate projections of Ω into each individual component:

$$\Omega = \prod_{\alpha\in A}\mathbb{R}\,, \qquad X_\alpha := \pi_\alpha\,.$$

However, when A is infinite we must be careful with this construction since in principle we do not know how to make sense of $\prod_{\alpha\in A}\mathbb{P}_{X_\alpha}$ for infinite A.

Moreover, we want to emphasize that any reasonable probability space should obey the following so called Kolmogorov "consistency" conditions for its marginals:

$$\mathbb{P}_{(X_1,X_2)}[B_1 \times B_2] = \mathbb{P}_{(X_2,X_1)}[B_2 \times B_1] \qquad (B_1, B_2 \in \mathfrak{B}(\mathbb{R}))$$

and in fact this should hold not just for two random variables but for any finite subcollection and any permutation within that finite subcollection.

Moreover, if $k<n$,

$$\mathbb{P}_{(X_1,\cdots,X_n)}\left[B_1\times\cdots\times B_k\times\mathbb{R}^{n-k}\right]=\mathbb{P}_{(X_1,\cdots,X_k)}\left[B_1\times\cdots B_k\right]\qquad(B_1,\cdots,B_k\in\mathfrak{B}(\mathbb{R}))\ .$$

This leads us to

Theorem 7.49 (Kolmogorov's extension)**.** *Let A be an arbitrary nonempty set, and define*

$$I_n(A):=\{\,\alpha:\{\,1,\cdots,n\,\}\to A\mid\alpha\text{ is injective}\,\}\ ,\qquad I(A):=\bigsqcup_{n=1}^{\infty}I_n(A)$$

and

$$\mathcal{M}_n:=\{\,\mu:\mathfrak{B}(\mathbb{R}^n)\to[0,1]\mid\mu\text{ is a probability measure}\,\}\ ,\qquad\mathcal{M}:=\bigsqcup_{n=1}^{\infty}\mathcal{M}_n\,.$$

Assume that we are given a map

$$m:I(A)\to\mathcal{M}$$

such that for any $\alpha\in I_n(A)$,

1. *$m(\alpha)\in\mathcal{M}_n$, i.e., $m(\alpha)$ is a probability measure on $\mathbb{R}^n$.*
2. *For any $\pi\in S_n$ the group of permutations,*

$$m(\alpha)(B_1\times\cdots\times B_n)=m(\alpha\circ\pi)\left(B_{\pi(1)}\times\cdots\times B_{\pi(n)}\right)\qquad(B_1,\cdots,B_n\in\mathfrak{B}(\mathbb{R}))\ .$$

3. *For any $k<n$, let $\alpha|_{\{1,\cdots,k\}}\in I_k(A)$ be the restriction to the first k indices. Then*

$$m(\alpha)\left(B_1\times\cdots\times B_k\times\mathbb{R}^{n-k}\right)=m\left(\alpha|_{\{1,\cdots,k\}}\right)(B_1\times\cdots\times B_k)\qquad(B_1,\cdots,B_k\in\mathfrak{B}(\mathbb{R}))\ .$$

Then there exists a unique *probability space $(\Omega,\operatorname{Msrbl}(\Omega),\mathbb{P})$ which has a product structure $\Omega:=(\mathbb{R}\cup\{\,\infty\,\})^A$ so that if $\pi_j:\Omega\to\mathbb{R}$ is the projection to the jth coordinate, then for all $\alpha\in I_n(A)$ and $\{\,\alpha_1,\cdots,\alpha_n\,\}=\operatorname{im}(\alpha)$, then $m(\alpha)=\mathbb{P}_{\left(\pi_{\alpha_1},\cdots,\pi_{\alpha_n}\right)}$. Moreover, the measure $\mathbb{P}$ is Radon in the sense of Definition 2.79.*

Proof. The theorem may be proven either using the Kakutani-Markov-Riesz representation Theorem 2.84 or Caratheodory's extension Theorem 2.76. We shall use the latter approach as it is more elementary (see Folland for the former). Let us define

$$\Omega:=\mathbb{R}^A\equiv\{\,f:A\to\mathbb{R}\,\}\ .$$

This set *is* represented with a product structure so it is furnished with a natural σ-algebra, the product σ-algebra Definition 5.1:

$$\begin{aligned}\otimes_{\alpha\in A}\mathfrak{B}(\mathbb{R})&\equiv\sigma\left(\left\{\,\pi_\alpha^{-1}(E_\alpha)\;\middle|\;E_\alpha\in\mathfrak{B}(\mathbb{R})\wedge\alpha\in A\,\right\}\right)\\&=\sigma\left(\left\{\prod_{\alpha\in A}E_\alpha\;\middle|\;E_\alpha\in\mathfrak{B}(\mathbb{R})\,\forall\alpha\in A\wedge E_\alpha\neq\mathbb{R}\text{ for at most finitely-many }\alpha\text{'s}\right\}\right)\ .\end{aligned}$$

It will also be useful to set, for $S\subseteq A$ finite,

$$\mathcal{F}_S:=\sigma\left(\left\{\prod_{\alpha\in A}E_\alpha\;\middle|\;E_\alpha\in\mathfrak{B}(\mathbb{R})\,\forall\alpha\in A\wedge E_\alpha\neq\mathbb{R}\text{ onlyfor }\alpha\in S\right\}\right)\ .$$

Let

$$\mathcal{A}:=\bigcup_{S\subseteq A\text{ finite}}\mathcal{F}_S\,.$$

We claim that $\mathscr{A}$ is an algebra on Ω. Moreover, let us define a map

$$\rho : \mathscr{A} \to [0,1] \ .$$

To do so, for $S \subseteq A$ finite, let $\pi_S : \Omega \to \mathbb{R}^S$ by the projection onto the S coordinates. given $E \in \mathscr{A}$, there is some $S \subseteq A$ finite with $n := |S|$, so that $E \in \mathscr{F}_S$. Since $E \in \mathscr{F}_S$, really E only has n non-trivial factors and the rest are factors of $\mathbb{R}$, so it is "encoded" by $\pi_S(E)$ just as well. Let $\alpha \in I_n(A)$ be so that $\operatorname{im}(\alpha) = S$. Then we set

$$\rho(E) := m(\alpha)(\pi_S(E)) \ .$$

This map is well-defined (independent of the choices of $S \subseteq A$ and $\alpha \in I_n(A)$) by the consistency conditions we stipulate on m. We claim that ρ is a premeasure on $\mathscr{A}$. Then by Theorem 2.76, there is a unique measure, extending ρ on $\mathscr{A}$ onto $\mathbb{P} := \mu_{\varphi_\rho}$ on $\sigma(\mathscr{A}) = \otimes_{\alpha \in A} \mathfrak{B}(\mathbb{R})$. The marginals agree *by construction.*

To obtain regularity we need to associated the measure $\mathbb{P}$ with a measure on an extension onto $(\mathbb{R} \cup \{\infty\})^A$ where $\mathbb{R} \cup \{\infty\}$ is the *one-point compactification*, so that by Tychonoff, $(\mathbb{R} \cup \{\infty\})^A$ is a compact Hausdorff space. The difference between the original $\mathbb{P}$ and the extension to $(\mathbb{R} \cup \{\infty\})^A$ is not meaningful is we set $(\mathbb{R} \cup \{\infty\})^n \setminus \mathbb{R}^n$ to have measure zero for all n. Then the regularity theorems we have imply that the extension is a Radon measure. ☐

Corollary 7.50 (Simple random walk stochastic process). *We now know that the simple random walk exist. Let μ_0 be any a-priori measure on $\mathfrak{B}(\mathbb{R})$. Then we construct a measure $\mathbb{P}$ on $\Omega := \mathbb{R}^{\mathbb{N}}$ as the joint probability distribution of the sequence of projection maps $X_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ which are independent and identically distributed according to μ_0, i.e.,*

$$\mathbb{P}_{(X_{j_1}, \cdots, X_{j_n})} = \prod_{k=1}^{n} \mu_0 \qquad (j_1, \cdots, j_n \in \mathbb{N}) \ .$$

Then the simple random walk is $S_N := \sum_{n=1}^{N} X_n$. The simplest model is $\mu_0 = \frac{1}{2}(\delta_{-1} + \delta_1)$ (i.e. symmetric Bernoulli RVs). One may think of N as the time variable of a particle proceeding according to a (discrete) diffusion equation. One may verify the consistency conditions of this definition holds.

Corollary 7.51 (White noise stochastic process). *Let $\mathscr{D} := C_c^\infty(\mathbb{R})$ be the set of compactly supported smooth "test functions" $\mathbb{R} \to \mathbb{R}$. We shall take our index set $A = \mathscr{D}$. Then we take $\Omega := \mathbb{R}^{\mathscr{D}}$ and for any finite sub-collection $\varphi_1, \cdots, \varphi_n \in \mathscr{D}$, define the marginal $\mathbb{P}_{(W(\varphi_1), \cdots, W(\varphi_n))}$ to be given by the density*

$$\frac{\mathrm{d}\mathbb{P}_{(W(\varphi_1), \cdots, W(\varphi_n))}}{\mathrm{d}\lambda}(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(K)}} \exp\left(-\frac{1}{2}\left\langle x, K^{-1}x \right\rangle_{\mathbb{R}^n}\right) \qquad (x \in \mathbb{R}^n) ; \qquad K_{ij} := \langle \varphi_i, \varphi_j \rangle_{L^2(\mathbb{R}^n)} \ .$$

One may verify the consistency conditions for this definition hold. Then "white noise" is the resulting stochastic process $\{ W(\varphi) : \Omega \to \mathbb{R} \}_{\varphi \in \mathscr{D}}$ which is the coordinate projections (here our coordinates are $\varphi \in \mathscr{D}$), i.e.,

$$(W(\varphi))(\omega) \equiv \omega(\varphi) \qquad (\omega : \mathscr{D} \to \mathbb{R} \in \Omega) \ .$$

Even though we would like to really take $\mathscr{D}$ as the set of delta functions, i.e., $\{ W(t) \}_{t \in \mathbb{R}}$ are independent Gaussians with

$$\frac{\mathrm{d}\mathbb{P}_{(W(t_1), \cdots, W(t_n))}}{\mathrm{d}\lambda}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\langle x, \mathbb{1}_n x \rangle_{\mathbb{R}^n}\right) \qquad (x \in \mathbb{R}^n)$$

the resulting process $\mathbb{R} \ni t \mapsto W(t) \in \mathbb{R}$ (a random function) is not "regular" in any kind of sense. In fact it turns out we cannot even prove it is measurable, let alone continuous anywhere. For this reason we work with the dual object $\mathscr{D} \ni \varphi \mapsto W(\varphi) \in \mathbb{R}$.

Corollary 7.52 (Brownian motion)**.** *Let $A := [0,\infty)$ be the indexing set, $\Omega := \mathbb{R}^A$, and define a stochastic process $\{ B_t : \Omega \to \mathbb{R} \}_{t\in A}$ via the finite marginals, for $0 \le t_1 < \cdots < t_n$, $\mathbb{P}_{(B_{t_1},\cdots,B_{t_n})}$ given by the density*

$$\frac{\mathrm{d}\mathbb{P}_{(B_{t_1},\cdots,B_{t_n})}}{\mathrm{d}\lambda}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(K)}}\exp\left(-\frac{1}{2}\left\langle x, K^{-1}x\right\rangle_{\mathbb{R}^n}\right) \qquad (x\in\mathbb{R}^n)\,; \qquad K_{ij} := \min(\{ t_i, t_j \})\ .$$

Note in particular this implies

$$\mathbb{P}_{X_0} = \delta_0\,,$$

i.e., $X_0 = 0$ almost-surely. One may verify the consistency conditions for this definition hold. In fact, one may prove (eventually) that almost-surely, the (random) map $t \mapsto X_t$ is continuous (Kolmogorov–Chentsov continuity theorem), but nowhere differentiable. Another way to characterize white noise from above is

$$W_t \approx \partial_t X_t$$

which explains why W_t is "not a function".

Corollary 7.53 (Brownian bridge)**.** *Let $T > 0$ and $A := [0,T]$. We want to build Brownian motion which is conditioned such that $B_T = 0$ also. One way to achieve that*

$$B_t^{bridge} := B_t - \frac{t}{T}B_T$$

where $\{ B_t \}_{t\in[0,T]}$ is standard Brownian motion (as above) which is merely conditioned to have $B_0 = 0$. Another way to achieve this is via the joint density

$$\frac{\mathrm{d}\mathbb{P}_{(B_{t_1}^{bridge},\cdots,B_{t_n}^{bridge})}}{\mathrm{d}\lambda}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(K)}}\exp\left(-\frac{1}{2}\left\langle x, K^{-1}x\right\rangle_{\mathbb{R}^n}\right) \qquad (x\in\mathbb{R}^n)\,; \qquad K_{ij} := \min(\{ t_i, t_j \}) - \frac{t_i t_j}{T}\ .$$

Corollary 7.54 (Pinned Brownian motion)**.** *What is we wanted a skewed Brownian bridge which starts at some x and ends at some y? Then we could take*

$$B_t^{pinned} := x + (y-x)\frac{t}{T} + B_t - \frac{t}{T}B_T$$

where $\{ B_t \}_{t\in[0,T]}$ is the usual Brownian motion which only has $B_0 = 0$ almost-surely. Alternatively, we could specify the joint density

$$\frac{\mathrm{d}\mathbb{P}_{(B_{t_1}^{pinned},\cdots,B_{t_n}^{pinned})}}{\mathrm{d}\lambda}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(K)}}\exp\left(-\frac{1}{2}\left\langle (x-\mu), K^{-1}(x-\mu)\right\rangle_{\mathbb{R}^n}\right) \qquad (x\in\mathbb{R}^n)$$

where $K_{ij} := \min(\{ t_i, t_j \}) - \frac{t_i t_j}{T}$ and $\mu_i := x + \frac{t_i}{T}(y-x)$.

7.7 The Wiener measure [Simon functional integration] [extra]

Here we follow, for the most part, [Sim04].

We have already seen the "existence" of the Wiener measure as the measure in Corollary 7.52. We would like to establish some properties of it.

7.7.1 Scaling law

Theorem 7.55. *Let $\{ B_t \}_{t\geq 0}$ be the usual Brownian motion (which only has the $B_0 = 0$ conditioning). Then for any $c > 0$,*

$$B_t \overset{d}{=} \sqrt{c} B_{\frac{t}{c}} \, .$$

The equivalence in distribution is meant in the sense that for any $0 \leq t_1 < \cdots < t_n$,

$$\mathbb{P}_{\left(B_{t_1},\cdots,B_{t_n}\right)} = \mathbb{P}_{\left(\sqrt{c}B_{\frac{t_1}{c}},\cdots,\sqrt{c}B_{\frac{t_n}{c}}\right)} \, .$$

Proof. Calculate the density function for any finite sample vector of Brownian vector and conclude by uniqueness of Kolmogorov extension. Let $S \in \mathfrak{B}\left(\mathbb{R}^n\right)$. Then

$$\begin{aligned}
\mathbb{P}_{\left(\sqrt{c}B_{\frac{t_1}{c}},\cdots,\sqrt{c}B_{\frac{t_n}{c}}\right)}[S] &\equiv \mathbb{P}\left[\left(\sqrt{c}B_{\frac{t_1}{c}},\cdots,\sqrt{c}B_{\frac{t_n}{c}}\right) \in S\right] \\
&= \mathbb{P}\left[\left(B_{\frac{t_1}{c}},\cdots,B_{\frac{t_n}{c}}\right) \in \frac{1}{\sqrt{c}} S\right] \\
&= \mathbb{P}_{\left(B_{\frac{t_1}{c}},\cdots,B_{\frac{t_n}{c}}\right)}\left[\frac{1}{\sqrt{c}} S\right] .
\end{aligned}$$

Now,

$$\frac{\mathrm{d}\mathbb{P}_{\left(B_{\frac{t_1}{c}},\cdots,B_{\frac{t_n}{c}}\right)}}{\mathrm{d}\lambda}(x) \equiv \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\left(\frac{1}{c}K\right)}} \exp\left(-\frac{1}{2}\left\langle x, cK^{-1}x\right\rangle_{\mathbb{R}^n}\right) \qquad (x \in \mathbb{R}^n) \, ; \qquad K_{ij} := \min\left(\{ t_i, t_j \}\right) .$$

Hence

$$\begin{aligned}
\mathbb{P}_{\left(B_{\frac{t_1}{c}},\cdots,B_{\frac{t_n}{c}}\right)}\left[\frac{1}{\sqrt{c}} S\right] &= \int_{x\in\frac{1}{\sqrt{c}}S} \frac{\mathrm{d}\mathbb{P}_{\left(B_{\frac{t_1}{c}},\cdots,B_{\frac{t_n}{c}}\right)}}{\mathrm{d}\lambda}(x)\,\mathrm{d}\lambda(x) \\
&= \int_{x\in\frac{1}{\sqrt{c}}S} \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\left(\frac{1}{c}K\right)}} \exp\left(-\frac{1}{2}\left\langle x, cK^{-1}x\right\rangle_{\mathbb{R}^n}\right)\mathrm{d}\lambda(x) \\
&\overset{y:=\sqrt{c}x}{=} \int_{y\in S} \frac{1}{(2\pi)^{\frac{n}{2}} c^{-\frac{n}{2}}\sqrt{\det(K)}} \exp\left(-\frac{1}{2}\left\langle y, K^{-1}y\right\rangle_{\mathbb{R}^n}\right) c^{-\frac{n}{2}}\mathrm{d}\lambda(y) \\
&= \mathbb{P}_{\left(B_{t_1},\cdots,B_{t_n}\right)}[S] \, .
\end{aligned}$$

□

7.7.2 The Markov property

Definition 7.56 (Filtration)**.** Let $(\Omega, \mathfrak{A}, \mathbb{P})$ be a probability space. For every $t \geq 0$, let $\mathfrak{F}_t$ be a sub-σ-algebra of $\mathfrak{A}$. Then $(\mathfrak{F}_t)_{t\geq 0}$ is a filtration, iff

$$\mathfrak{F}_s \subseteq \mathfrak{F}_t \qquad (s \leq t) \, .$$

Then $(\Omega, \mathfrak{A}, \mathbb{P}, \mathfrak{F})$ is called a filtered probability space.

Definition 7.57 (Markov property)**.** Let $(\Omega, \mathfrak{A}, \mathbb{P}, \mathfrak{F})$ be a filtered probability space and $(X_t)_{t\geq 0}$ be a stochastic process. Then $(X_t)_t$ is said to have the *Markov property* iff

$$\mathbb{E}\left[f\left(X_t\right)|\mathfrak{F}_s\right] = \mathbb{E}\left[f\left(X_t\right)|\sigma\left(X_s\right)\right] \qquad (s < t, \; f : \Omega \to \mathbb{R} \text{ bounded and msrbl}) \, .$$

I.e., if the conditional expectation w.r.t. the entire past is the same as the conditional expectation w.r.t. the last point in the past.

Definition 7.58 (Stopping time)**.** Let $(\Omega, \mathfrak{A}, \mathbb{P}, \mathfrak{F})$ be a filtered probability space and $\tau : \Omega \to \mathbb{R}$ be a random variable. Then τ is called a *stopping time w.r.t.* $\mathfrak{F}$ *iff*

$$\{\, \tau \leq t \,\} \in \mathfrak{F}_t\,.$$

In words, that means that the set $\{\, \tau \leq t \,\}$ *is "determined" only by everything that happened until time* t*.*

Definition 7.59 (Strong Markov property)**.** Let $(\Omega, \mathfrak{A}, \mathbb{P}, \mathfrak{F})$ be a filtered probability space and $(X_t)_{t\geq 0}$ be a stochastic process. Then $(X_t)_t$ is said to have the *strong Markov property* iff for any stopping time τ, conditioned on $\{\, \tau < \infty \,\}$, $X_{\tau+t}$ is independent of $\mathcal{F}_\tau$ given X_τ.

Remark 7.60. Let $\Omega = \mathbb{R}^{[0,\infty)}$ be the sample space for Brownian motion taken with the product σ-algebra from the Borel σ-algebra on each copy of $\mathbb{R}$. For Brownian motion $(B_t)_{t\geq 0}$, a natural filtration is given by

$$\mathfrak{F}_t := \sigma\left(\{\, B_s \mid s \in [0,t] \,\}\right)\,.$$

Claim 7.61. Brownian motion $(B_t)_{t\geq 0}$ obeys the Markov property.

Proof. By definition, we have that $B_t - B_s$ is independent of $\mathfrak{F}_s$ and $B_t - B_s \sim \mathcal{N}(0, t-s)$. For $t > s$, write

$$B_t := B_s + X$$

with $X := B_t - B_s$. We note X is independent of $\mathfrak{F}_s$. Then

$$\begin{aligned} \mathbb{E}\left[f(B_t)\,|\mathfrak{F}_s\right] &= \mathbb{E}\left[f(B_s + X)\,|\mathfrak{F}_s\right] \\ &= \mathbb{E}\left[f(B_s + X)\,|\sigma(B_s)\right] \\ &= \int_{y\in\mathbb{R}} f(B_s + y)\,\frac{\mathrm{e}^{-\frac{y^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}\mathrm{d}y \\ &=: \varphi(B_s)\,. \end{aligned}$$

Hence $\mathbb{E}\left[f(B_t)\,|\mathfrak{F}_s\right]$ is $\sigma(B_s)$-measurable so that

$$\mathbb{E}\left[f(B_t)\,|\mathfrak{F}_s\right] = \mathbb{E}\left[f(B_t)\,|\sigma(B_s)\right]\,.$$

□

Claim 7.62. Brownian motion $(B_t)_{t\geq 0}$ obeys the strong Markov property.

Proof. Let τ be a stopping time. Our goal is to show that within the event $\{\, \tau < \infty \,\}$, $(B_{\tau+t} - B_\tau)_{t\geq 0}$ is standard Brownian motion which is independent of $\mathfrak{F}_\tau$. To that end, let

$$\tau_n := 2^{-n}\left\lceil 2^n \tau \right\rceil \geq \tau \qquad (n \in \mathbb{N})\,.$$

Then $\tau_n \to \tau$ from above, and each τ_n is itself a stopping time. TODO: complete this. □

7.7.3 Donsker's theorem

Another way to think about Brownian motion is as follows. Let $\{\, X_n \,\}_{n\in\mathbb{N}}$ be an IID sequence of Bernoulli ± 1 random variables, each with Bernoulli parameter $\frac{1}{2}$. Then $\sum_{n=1}^{N} X_n$ is a random walk up to time $N \in \mathbb{N}$. As we have seen, using the central limit Theorem 7.35,

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N} X_n \to Z$$

in distribution, where $Z \sim \mathcal{N}(0,1)$. What about the process $\{ B_t \}_{t\in[0,1]}$ defined via

$$B_t := \sqrt{t} \lim_{N\to\infty} \frac{1}{\sqrt{tN}} \sum_{n=1}^{\lfloor tN \rfloor} X_n$$

By the central limit Theorem 7.35, we only have $B_1 \to \mathcal{N}(0,1)$ in distribution. But what about other $t \in [0,1]$? It turns out B_t converges, in distribution, to the same Brownian motion we have already seen (we shall not prove this fact here but this statement is known as Donsker's theorem).

7.7.4 Continuity of Brownian motion

As constructed so far, Brownian motion B_t is an arbitrary function, it need not even be measurable. Here we want to establish that it is almost-surely continuous. By definition, given $0 \leq t < s$, the joint distribution of B_t and B_s is ac and is given by the density

$$\frac{\mathrm{d}\mathbb{P}_{(B_t,B_s)}}{\mathrm{d}\lambda}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(K)}} \exp\left(-\frac{1}{2}\left\langle x, K^{-1}x\right\rangle_{\mathbb{R}^2}\right)$$

where $K := \begin{bmatrix} t & t \\ t & s \end{bmatrix}$ so $K^{-1} = \frac{1}{s-t}\begin{bmatrix} \frac{s}{t} & -1 \\ -1 & 1 \end{bmatrix}$. We can thus calculate

$$\begin{aligned}
\mathbb{E}\left[(B_t - B_s)^{2m}\right] &= \int_{x\in\mathbb{R}^2} (x_1-x_2)^{2m} \frac{1}{2\pi}\frac{1}{\sqrt{t(s-t)}} \exp\left(-\frac{1}{2}\left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, K^{-1}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right\rangle_{\mathbb{R}^2}\right) \mathrm{d}\lambda(x) \\
&= \int_{x\in\mathbb{R}^2} (x_1-x_2)^{2m} \frac{1}{2\pi}\frac{1}{\sqrt{t(s-t)}} \exp\left(-\frac{1}{2}\left(\frac{sx_1^2 - 2tx_1x_2 + tx_2^2}{t(s-t)}\right)\right) \mathrm{d}\lambda(x) \\
&= \int_{x\in\mathbb{R}^2} (x_1-x_2)^{2m} \frac{1}{2\pi}\frac{1}{\sqrt{t(s-t)}} \exp\left(-\frac{1}{2}\left(\frac{(s-t)x_1^2 + t(x_1-x_2)^2}{t(s-t)}\right)\right) \mathrm{d}\lambda(x) \\
&= \int_{x\in\mathbb{R}^2} (x_1-x_2)^{2m} \frac{1}{2\pi}\frac{1}{\sqrt{t(s-t)}} \exp\left(-\frac{1}{2}\left(\frac{x_1^2}{t} + \frac{(x_1-x_2)^2}{s-t}\right)\right) \mathrm{d}\lambda(x) \\
&= \frac{1}{2\pi}\frac{1}{\sqrt{t(s-t)}} \int_{y\in\mathbb{R}^2} y_2^{2m} \exp\left(-\frac{1}{2}\left(\frac{y_1^2}{t} + \frac{y_2^2}{s-t}\right)\right) \mathrm{d}\lambda(y) \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{s-t}} \int_{y_2\in\mathbb{R}} y_2^{2m} \exp\left(-\frac{1}{2}\frac{y_2^2}{s-t}\right) \mathrm{d}\lambda(y) \\
&= 2^m\left(1 + (-1)^{2m}\right)(s-t)^m \Gamma\left(\frac{1}{2}+m\right).
\end{aligned}$$

In particular, for $m = 2$ we get

$$\mathbb{E}\left[(B_t - B_s)^4\right] = \frac{1}{2}(s-t)^2 \qquad (0 \leq t < s).$$

Theorem 7.63 (Kolmogorov continuity)**.** *Let (S,d) be a complete separable metric space and $X : [0,\infty) \times \Omega \to S$ be a stochastic process. Assume that for all $T > 0$, there are some α, β, K positive constants such that*

$$\mathbb{E}\left[d(X_t, X_s)^\alpha\right] \leq K|t-s|^{\beta+1} \qquad (0 \leq s, t \leq T).$$

Then there exists a modification $\tilde{X}$ of X which is a continuous process, i..e, $\tilde{X} : [0,\infty) \times \Omega \to S$ such that:

1. *For any $t \geq 0$, $\mathbb{P}\left[X_t = \tilde{X}_t\right] = 1$.*
2. *$\mathbb{P}\left[t \mapsto \tilde{X}_t \text{ is continuous}\right] = 1$.*
3. *In fact, $\mathbb{P}\left[t \mapsto \tilde{X}_t \text{ is locally } \gamma\text{-Hölder continuous for every } 0 < \gamma < \frac{\beta}{\alpha}\right] = 1$.*

Proof. Pick some $\gamma \in \left(0, \frac{\beta}{\alpha}\right)$. Then $1+\beta-\alpha\gamma =: \delta > 1$. For any $m \in \mathbb{N}$, partition $[0,T]$ into 2^m equal subintervals of length $\Delta_m := T2^{-m}$. The grid points are $t_{m,k} := k\Delta_m$ with $k = 0, \cdots, 2^m$. A "bad" event is when we violate the Hoelder continuity we are seeking, i.e.,

$$A_m := \left\{ \exists k = 1, \cdots, 2^m : \left|X_{t_{m,k}} - X_{t_{m,k-1}}\right| > 2^{-m\gamma} \right\} .$$

We bound the probability of A_m via Markov's inequality:

$$\begin{aligned} \mathbb{P}\left[A_m\right] &\leq \sum_{k=1}^{2^m} \mathbb{P}\left[\left|X_{t_{m,k}} - X_{t_{m,k-1}}\right| > 2^{-m\gamma}\right] \\ &\leq \sum_{k=1}^{2^m} \frac{\mathbb{E}\left[\left|X_{t_m,k} - X_{t_m,k-1}\right|^\alpha\right]}{2^{-m\alpha\gamma}} \\ &\leq 2^m \frac{K\Delta_m^{1+\beta}}{2^{-m\alpha\gamma}} = KT^{1+\beta}2^{-m\delta} . \end{aligned}$$

Since $\delta > 1$, this is finite. Applying now Lemma 7.22, we get that almost-surely, only finitely many of the A_m's may occur. Hence, there is a (random) $M(\omega)$ such that for all $m \geq M(\omega)$, and all k,

$$|X_{t_m,k} - X_{t_m,k-1}| \leq 2^{-m\gamma} .$$

Now, if $s < t$ on the m-grid, i.e., $t - s = J\Delta_m$ for some $J \in \mathbb{N}$, then

$$|X_t - X_s| \leq \sum_{i=1}^{J} |X_{t_m,i} - X_{t_m,i-1}| \leq J2^{-m\gamma} = \frac{t-s}{\Delta_m} 2^{-m\gamma} = T^{-\gamma}(t-s)^\gamma .$$

But the dyadic grid for all m is dense in $[0,T]$, and the grid-paths are uniformly Hölder continuous for all $m \geq M(\omega)$, $X_t(\omega)$ admits a unique continuous extension to all of t:

$$\tilde{X}_t := \lim_{m\to\infty} X_{t_m}$$

where t_m is a sequence which converges to t. □

7.7.5 The Feynman-Kac formula

TODO: complete this.

Let $V : \mathbb{R} \to \mathbb{R}$ be some sufficiently nice function. We want to prove the identity

$$\left(\exp\left(-t\left(-\Delta + V(X)\right)\right)\psi_0\right)(x) = \mathbb{E}\left[\psi_0(x+B_t)\exp\left(-\int_0^t V(x+B_s)\,\mathrm{d}s\right)\right] \qquad \left(\psi_0 \in L^2(\mathbb{R}), t > 0\right) .$$

To do so, we can use the Trotter product formula to build the path integral out of small increments. Another possibility is to use the fat that the heat kernel is the *unique* solution to the heat equation, so if

$$u_{\psi_0}(t,x) := \mathbb{E}\left[\psi_0(x+B_t)\exp\left(-\int_0^t V(x+B_s)\,\mathrm{d}s\right)\right] \qquad (t \geq 0, x \in \mathbb{R}) .$$

then we should show

$$\partial_t u_{\psi_0} = -Hu_{\psi_0}$$

for $H = -\Delta + V(X)$. Clearly u_{ψ_0} obeys the initial condition, since

$$u_{\psi_0}(0,\cdot) = \mathbb{E}\left[\psi_0(\cdot + B_0)\right] = \psi_0 .$$

Hence let us calculate $\partial_t u_{\psi_0}$. We note that because Brownian motion has independent increments and is conditioned to start at the origin at time zero, we actually have

$$B_{t+\varepsilon} \overset{d}{=} B_t + \tilde{B}_\varepsilon \qquad (t \geq 0, \varepsilon > 0)$$

where $\left(\tilde{B}_t\right)_t$ is another independent copy of Brownian motion. Moreover,

$$\begin{aligned}\int_0^{t+\varepsilon} V(x+B_s)\,\mathrm{d}s &= \int_0^t V(x+B_s)\,\mathrm{d}s + \int_t^{t+\varepsilon} V(x+B_s)\,\mathrm{d}s \\ &= \int_0^t V(x+B_s)\,\mathrm{d}s + \int_0^{\varepsilon} V(x+B_{t+s})\,\mathrm{d}s \\ &= \int_0^t V(x+B_s)\,\mathrm{d}s + \int_0^{\varepsilon} V\left(x+B_t+\tilde{B}_s\right)\mathrm{d}s\end{aligned}$$

As a result, we may separate the expectation to expectation w.r.t. B and w.r.t. $\tilde{B}$. Using Fubini we then have

$$\begin{aligned}\mathbb{E}_B\left[\psi_0(x+B_{t+\varepsilon})\exp\left(-\int_0^{t+\varepsilon} V(x+B_s)\,\mathrm{d}s\right)\right] &= \mathbb{E}_{\tilde{B}}\left[\mathbb{E}_B\left[\psi_0\left(x+B_\varepsilon+\tilde{B}_t\right)\exp\left(-\int_0^{\varepsilon} V(x+B_s)\,\mathrm{d}s + \int_0^t V\left(x+B_\varepsilon+\tilde{B}_s\right)\right.\right.\right. \\ &= \mathbb{E}_B\left[u_{\psi_0}(\varepsilon, x+B_t)\exp\left(-\int_0^{\varepsilon} V(x+B_s)\,\mathrm{d}s\right)\right].\end{aligned}$$

Now, for infinitesimal times,

$$u_{\psi_0}(\varepsilon,x) \equiv \mathbb{E}\left[\psi_0(x+B_\varepsilon)\exp\left(-\int_0^{\varepsilon} V(x+B_s)\,\mathrm{d}s\right)\right]$$

and

$$\exp\left(-\int_0^{\varepsilon} V(x+B_s)\,\mathrm{d}s\right) = 1-\int_0^{\varepsilon} V(x+B_s)\,\mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right).$$

Moreover,

$$\mathbb{E}\left[\psi_0(x+B_\varepsilon)\right] = \mathbb{E}\left[\psi_0(x+B_\varepsilon)\right]$$

7.7.6 The Karhunen–Loève expansion

Derive a spectral expansion of the random stochastic process of the form

$$X_t = \sum_{k=1}^{\infty} Z_k \varphi_k(t)$$

where $\{\, Z_k \,\}_k$ are pairwise uncorrelated random variables and φ_k are continuous real-valued functions on $[a,b]$ that are pairwise orthogonal in $L^2([a,b])$ and are the eigenfunctions of the covariance matrix of the process X_t.

7.8 Conditional expectation and probability [extra]

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $X:\Omega\to\mathbb{R}$ be a random variable (i.e., it is $\mathcal{F}$-measurable). Let $\mathfrak{G}\subseteq\mathcal{F}$ be a sub-σ-algebra. In general there is no guarantee that X is $\mathfrak{G}$-measurable!

> **Example 7.64.** Let $\Omega:=\mathbb{R}$, $\mathcal{F}:=\mathfrak{B}(\mathbb{R})$ and $\mathbb{P}$ be the standard normal Gaussian distribution. Then $X:\Omega\to\mathbb{R}$ is given by $X(\omega)=\omega$ and hence $X\sim\mathcal{N}(0,1)$. Clearly, $|X|$ is also a random variable (it is also a measurable function) and now we let
>
> $$\mathfrak{G}:=\sigma(|X|).$$
>
> One may see that $\mathfrak{G}$ consists of all Borel subsets of $\mathbb{R}$ which are symmetric w.r.t. $x\mapsto -x$. We claim that X is not $\mathfrak{G}$-measurable. Indeed, take $(0,\infty)\in\mathfrak{B}(\mathbb{R})$. Then $X^{-1}((0,\infty))$ should be in $\mathfrak{G}$, but since $X(\omega)=\omega$, $X^{-1}((0,\infty))=(0,\infty)\notin\mathfrak{G}$! So X is not $\mathfrak{G}$-measurable.

The conditional expectation "solves" this problem by introduction a new random variable,

$$\mathbb{E}[X|\mathfrak{G}]:\Omega\to\mathbb{R} \qquad \text{(this is merely a matter of conventional notation)}$$

which is constructed to be $\mathfrak{G}$-measurable so that for all $G\in\mathfrak{G}$,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[X|\mathfrak{G}]\chi_G\right] &= \mathbb{E}[X\chi_G] \qquad (G\in\mathfrak{G}). \\ &\updownarrow \\ \int_G \mathbb{E}[X|\mathfrak{G}]\,\mathrm{d}\mathbb{P} &= \int_G X\,\mathrm{d}\mathbb{P} \qquad (G\in\mathfrak{G}).\end{aligned}$$

To see that such $\mathbb{E}[X|\mathfrak{G}]$ exits, let $\iota : \mathfrak{G} \to \mathfrak{F}$ be the natural injection. Then

$$\mathfrak{F} \ni F \mapsto \mathbb{E}[X\chi_F] =: \mu_X(F)$$

defines a finite measure on $(\Omega, \mathfrak{F})$ which is absolutely continuous w.r.t. $\mathbb{P}$ (do not confuse it with the marginal $\mathbb{P}_X$, which equals $\mathbb{P} \circ X^{-1}$ and is a measure on $\mathfrak{B}(\mathbb{R})$). Moreover, $\mu_X \circ \iota$ is absolutely continuous w.r.t. $\mathbb{P} \circ \iota$. We then define

$$\mathbb{E}[X|\mathfrak{G}] := \frac{\mathrm{d}\mu_X \circ \iota}{\mathrm{d}\mathbb{P} \circ \iota}.$$

This is automatically a random variable $\mathbb{E}[X|\mathfrak{G}] : \Omega \to \mathbb{R}$ which is $\mathfrak{G}$-measurable and $L^1(\mathbb{P} \circ \iota)$. Moreover, we have for all $G \in \mathfrak{G}$,

$$\mathbb{E}[\mathbb{E}[X|\mathfrak{G}]\chi_G] = \int_G \mathbb{E}[X|\mathfrak{G}]\,\mathrm{d}\mathbb{P} \equiv \int_G \mathbb{E}[X|\mathfrak{G}]\,\mathrm{d}\mathbb{P}\circ\iota = \int_G \frac{\mathrm{d}\mu_X \circ \iota}{\mathrm{d}\mathbb{P} \circ \iota}\mathrm{d}\mathbb{P}\circ\iota = \int_G \mathrm{d}\mu_X \circ \iota = \int_G \mathrm{d}\mu_X \equiv \mathbb{E}[X\chi_G].$$

Definition 7.65 (Conditional expectation w.r.t. a sub-σ-algebra)**.** The $\mathfrak{G}$-measurable random variable $\mathbb{E}[X|\mathfrak{G}] : \Omega \to \mathbb{R}$ is called "the conditional expectation of X given the σ-sub-algebra $\mathfrak{G}$".

Claim 7.66 (Uniqueness)*.* Note that once we find *any* RV which is the conditional expectation it is (almost-surely) unique.

Proof. Let Z be any other $\mathfrak{G}$-measurable random variable such that $\mathbb{E}[Z\chi_G] = \mathbb{E}[X\chi_G]$ for all $G \in \mathfrak{G}$. Our goal is to show that $Z = \mathbb{E}[X|\mathfrak{G}]$ almost-surely. Assume that

$$\mathbb{P}[Z \neq \mathbb{E}[X|\mathfrak{G}]] > 0.$$

Then either

$$\mathbb{P}[Z > \mathbb{E}[X|\mathfrak{G}]] + \mathbb{P}[Z < \mathbb{E}[X|\mathfrak{G}]] \quad > \quad 0.$$

Consider $D := Z > \mathbb{E}[X|\mathfrak{G}]$. It is $\mathfrak{G}$-measurable since both variables are. Hence, by hypothesis,

$$\begin{aligned} \mathbb{E}[Z\chi_D] &= \mathbb{E}[\mathbb{E}[X|\mathfrak{G}]\chi_D] \\ &\downarrow \\ \mathbb{E}[Z\chi_D] - \mathbb{E}[\mathbb{E}[X|\mathfrak{G}]\chi_D] &= 0 \\ &\downarrow \\ \mathbb{E}[(Z - \mathbb{E}[X|\mathfrak{G}])\chi_D] &= 0 \\ &\downarrow \\ \mathbb{P}[D] &= 0 \end{aligned}$$

the last step being true since on D, $(Z - \mathbb{E}[X|\mathfrak{G}]) > 0$ by definition. Similarly we can show that the other set has zero probability. □

As a result, any way we have to calculate the conditional expectation works.

Remark 7.67 (Functional analytic perspective)*.* Since $\mathfrak{G} \subseteq \mathfrak{F}$, one may think of

$$L^2(\Omega, \mathfrak{G}, \mathbb{P})$$

as a Hilbert subspace of $L^2(\Omega, \mathfrak{F}, \mathbb{P})$ naturally by considering maps which are $\mathfrak{G}$-measurable as a subset of those which are $\mathfrak{F}$-measurable. Now, if $X \in L^2(\Omega, \mathfrak{F}, \mathbb{P})$ then $\mathbb{E}[X|\mathfrak{G}]$ is the projection of X onto the closed subspace $L^2(\Omega, \mathfrak{G}, \mathbb{P})$.

We also find it useful to have the

Definition 7.68 (Conditional expectation w.r.t. another random variable)**.** Given a random variable $Y : \Omega \to \mathbb{R}$, a sub-$\sigma$-algebra of $\mathfrak{F}$ is the smallest one generated by Y, $\sigma(Y)$. Then

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)].$$

Once we have the conditional expectation, the conditional probability is merely a special case:

$$\mathbb{P}[A|\mathfrak{G}] \equiv \mathbb{E}[\chi_A|\mathfrak{G}] \qquad (A \in \mathfrak{F}) .$$

Hence the conditional probability is still a random variable.

Example 7.69. Going back to Example 7.64, one may verify

$$\mathbb{E}[X|\mathfrak{G}] = \frac{1}{2}(X + X(-\cdot)) = 0 .$$

So this example is not very interesting. If instead we take

$$\mathfrak{G} := \sigma(\operatorname{sgn}(X)) = \{ \varnothing, \mathbb{R}, [0,\infty), (-\infty, 0) \}$$

then X is again not $\mathfrak{G}$-measurable, since, $X^{-1}(\{ 1 \}) = \{ 1 \} \notin \mathfrak{G}$. Now however we may verify that

$$\mathbb{E}[X|\mathfrak{G}] = \sqrt{\frac{2}{\pi}}\operatorname{sgn}(X) .$$

Remark 7.70. One should not confuse $\mathbb{P}[A|\mathfrak{G}]$ with the more naive conditional probability (7.2) given above,

$$\mathbb{P}[A|G] \equiv \frac{\mathbb{P}[A \cap G]}{\mathbb{P}[G]} \qquad (A, G \in \mathfrak{F} : \mathbb{P}[G] > 0) .$$

There is. however, a connection between the two: If $G \in \mathfrak{G}$

$$\mathbb{P}[A|G]\,\mathbb{P}[G] = \int_G \chi_A \mathrm{d}\mathbb{P} = \int_G \mathbb{P}[A|\mathfrak{G}]\,\mathrm{d}\mathbb{P} .$$

Example 7.71. One should think of $\mathbb{E}[X|\mathfrak{G}]$ as averaging over only the information NOT contained in $\mathfrak{G}$. Thus, the notation is actually confusing, as

$$\mathbb{E}[X|\mathfrak{F}] = X$$

whereas if $\mathfrak{G} = \{ \varnothing, \Omega \}$,

$$\mathbb{E}[X|\mathfrak{G}] = \mathbb{E}[X] .$$

Example 7.72 (Marginals forget, conditionals refine)**.** Contrast the notion of conditional expectation with *marginals*. Say we have two random variables $X, Y : \Omega \to \mathbb{R}$. Then $\mathbb{P}_X$ integrates "out" the information of Y, whereas $\mathbb{E}[X|Y]$ integrates out all information of X which does not depend on Y. To see this more clearly, let us consider the following concrete example: Let $\Omega = \mathbb{R}^2$ and (X, Y) be a two dimensional Gaussian whose density is given so that $\mathbb{E}[X] = \mu_X$, $\mathbb{E}[Y] = \mu_Y$, $\mathbb{Var}[X] = \sigma_X^2$, $\mathbb{Var}[Y] = \sigma_Y^2$ and $\mathbb{Cov}[X, Y] =: \sigma_X \sigma_Y > 0$ for some $\rho > 0$. Then we may calculate the marginal $\mathbb{P}_X$ which has one dimensional Gaussian density with $\mathbb{E}[X] = \mu_X$ and $\mathbb{Var}[X] = \sigma_X^2$. However, the conditional expectation is given by

$$\mathbb{E}[X|Y] \quad = \quad \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y) .$$

We have

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X] .$$

Claim 7.73 (Linearity)*.* $\mathbb{E}[\alpha X + \beta Y|\mathfrak{G}] = \alpha\mathbb{E}[X|\mathfrak{G}] + \beta\mathbb{E}[Y|\mathfrak{G}]$.

Claim 7.74 (Iterated property)*.* Let $\mathfrak{H} \subseteq \mathfrak{G} \subseteq \mathfrak{F}$. Then

$$\mathbb{E}[\mathbb{E}[X|\mathfrak{G}]\,|\mathfrak{H}] = \mathbb{E}[X|\mathfrak{H}] .$$

Claim 7.75 (Pull out)*.* If Y is $\mathfrak{G}$-measurable and X is not, then

$$\mathbb{E}[XY|\mathfrak{G}] = \mathbb{E}[X|\mathfrak{G}]\,Y .$$

Claim 7.76 (Law of total variance).

$$\mathbb{Var}\left[X\right] = \mathbb{E}\left[\mathbb{Var}\left[X|\mathfrak{G}\right]\right] + \mathbb{Var}\left[\mathbb{E}\left[X|\mathfrak{G}\right]\right] .$$

Claim 7.77. If X is independent of $\mathfrak{G}$ (i.e., if X is independent of χ_G for any $G \in \mathfrak{G}$) then $\mathbb{E}\left[X|\mathfrak{G}\right] = \mathbb{E}\left[X\right]$.

Claim 7.78. If X happens to be $\mathfrak{G}$-measurable, then $\mathbb{E}\left[X|\mathfrak{G}\right] = X$.

Most of the theorems for the Lebesgue integral hold for the conditional expectation, as one may verify: positivity, monotonicity, Jensen, etc.

A The extended real line

We shall frequently use the symbol $[-\infty,\infty]$ or $\overline{\mathbb{R}}$ to denote *the extended real line*. As a set it is given by

$$\overline{\mathbb{R}} \quad := \quad \mathbb{R}\cup\{\,\pm\infty\,\}$$

and topologically we add the neighborhoods of $\pm\infty$ as those sets which contain the basic open sets

$$(a,\infty]$$

and

$$[-\infty,a)$$

respectively.

B Elementary families

Definition B.1 (elementary family)**.** Let X be a non-empty set. An *elementary family* $\mathcal{E}$ is a subset $\mathcal{E}\subseteq\mathcal{P}(X)$ such that

- $\varnothing\in\mathcal{E}$.
- (*closed under intersection*) If $E,F\in\mathcal{E}$ then $E\cap F\in\mathcal{E}$.
- If $E\in\mathcal{E}$ then $X\setminus E$ is a finite disjoint union of members of $\mathcal{E}$.

Claim B.2. If $\mathcal{E}$ is an elementary family then the collection $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{E}$ is an *algebra*, i.e., it contains X, it is closed under complements, and it is closed under finite unions.

Proof. Since $\varnothing\in\mathcal{E}$, then X is a finite disjoint union of members of $\mathcal{E}$, and so it is in $\mathcal{A}$ as desired.

Next, assume that $A\in\mathcal{A}$. We want to show that $X\setminus A\in\mathcal{A}$. We know that $A=\bigcup_{j=1}^{n}A_j$ where $A_j\in\mathcal{E}$ and $A_j\cap A_k=\varnothing$ if $j\neq k$. Hence

$$A^c=\left(\bigcup_{j=1}^{n}A_j\right)^c=\bigcap_{j=1}^{n}A_j^c\,.$$

But since $A_j\in\mathcal{E}$, its complement is by assumption equal to $\bigcup_{l=1}^{n_j}B_{j,l}$ with $B_{j,l}\cap B_{j,l'}=\varnothing$ for $l\neq l'$ and $B_{j,l}\in\mathcal{E}$. Hence

$$\begin{aligned} A^c &= \bigcap_{j=1}^{n}\left(\bigcup_{l=1}^{n_j}B_{j,l}\right) \\ &= \bigcup_{l}B_{1,l_1}\cap\cdots\cap B_{n,l_n} \end{aligned}$$

so A^c is also the finite union of disjoint members of $\mathcal{E}$.

Next, we want to establish that $\mathcal{A}$ is closed under finite unions. To that end, let $A,B\in\mathcal{A}$. Then

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

now if

$$A=\bigcup_{j=1}^{n}A_j$$

where $A_j\cap A_{j'}=\varnothing$ for $j\neq j'$ and $A_j\in\mathcal{E}$ and similarly,

$$B \quad = \quad \bigcup_{j=1}^{m}B_j\,.$$

Then clearly $(A \cap B^c) \cup B$ is a finite disjoint union of elements from $\mathcal{E}$. □

C Banach spaces

Definition C.1 (Norm)**.** A vector space V is called *a normed vector space* iff there is a map

$$\|\cdot\| : V \to [0, \infty)$$

which obeys the following axioms:

1. Absolute homogeneity:
$$\|\alpha\psi\| = |\alpha| \, \|\psi\| \qquad (\alpha \in \mathbb{C}, \psi \in V) \, .$$
2. Triangle inequality:
$$\|\psi + \varphi\| \leq \|\psi\| + \|\varphi\| \qquad (\psi, \varphi \in V) \, .$$
3. Injectivity at zero: If $\|\psi\| = 0$ for some $\psi \in V$ then $\psi = 0$.

Example C.2. Of course the first example of a normed vector space is simply $\mathbb{C}^n$, with the Euclidean norm:

$$\mathbb{C}^n \ni z \mapsto \|z\| \equiv \sqrt{\sum_{j=1}^{n} |z_j|^2} \, .$$

To show this is a norm we only need to establish the triangle inequality (the other two properties being easy). To that end, From the Cauchy-Schwarz inequality:

$$|\langle z, w \rangle_{\mathbb{C}^n}| \leq \|z\| \|w\|$$

we get

$$\begin{aligned} \|z + w\|^2 &\equiv \langle z + w, z + w \rangle \\ &= \|z\|^2 + \|w\|^2 + 2 \, \mathbb{R}\mathrm{e} \, \{ \langle z, w \rangle \} \\ &\leq \|z\|^2 + \|w\|^2 + 2 \, |\langle z, w \rangle| \\ &\overset{\text{C.S.}}{\leq} \|z\|^2 + \|w\|^2 + 2\|z\|\|w\| \\ &= (\|z\| + \|w\|)^2 \, . \end{aligned}$$

Hence we merely need to show the Cauchy-Schwarz inequality. To that end, if $w = 0$ there is nothing to prove. So define

$$\tilde{z} := z - \frac{\langle z, w \rangle}{\|w\|^2} w \, .$$

By construction, $\langle \tilde{z}, w \rangle = 0$ so

$$\|z\|^2 = \left\| \tilde{z} + \frac{\langle z, w \rangle}{\|w\|^2} w \right\|^2 = \|\tilde{z}\|^2 + \frac{|\langle z, w \rangle|^2}{\|w\|^4} \|w\|^2 \geq \frac{|\langle z, w \rangle|^2}{\|w\|^4} \|w\|^2 \, .$$

Remark C.3. Be careful that in the foregoing example we have used the inner-product structure of $\mathbb{C}^n$, but more generally, a norm need not be associated with an inner product.

Definition C.4 (Inner product space)**.** An inner-product space is a vector space $\mathcal{H}$ along with a sesquilinear map

$$\langle\cdot,\cdot\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$$

such that

1. Conjugate symmetry:
$$\langle\psi,\varphi\rangle = \overline{\langle\varphi,\psi\rangle} \qquad (\varphi,\psi\in\mathcal{H})\ .$$
2. Linearity in second argument:
$$\langle\psi,\alpha\varphi+\tilde{\varphi}\rangle = \alpha\langle\psi,\varphi\rangle + \langle\psi,\tilde{\varphi}\rangle \qquad (\varphi,\tilde{\varphi},\psi\in\mathcal{H},\alpha\in\mathbb{C})\ .$$
3. Positive-definite:
$$\langle\psi,\psi\rangle > 0 \qquad (\psi\in\mathcal{H}\setminus\{\,0\,\})\ .$$

Example C.5. Of course $\mathbb{C}^n$ with

$$\langle z,w\rangle_{\mathbb{C}^n} \equiv \sum_{j=1}^{n}\overline{z_j}w_j$$

is an inner-product space.

To every inner product one immediately may associate a norm, via

$$\|\psi\| := \sqrt{\langle\psi,\psi\rangle} \qquad (\psi\in\mathcal{H})\ .$$

The converse, however, hinges on the norm obeying the parallelogram law

Claim C.6. If a norm satisfies the parallelogram law:

$$\|\psi+\varphi\|^2 + \|\psi-\varphi\|^2 \leq 2\|\psi\|^2 + 2\|\varphi\|^2 \qquad (\varphi,\psi\in\mathcal{H})$$

then

$$\langle\psi,\varphi\rangle := \frac{1}{4}\left[\|\psi+\varphi\|^2 - \|\psi-\varphi\|^2 + \mathrm{i}\|\mathrm{i}\psi-\varphi\|^2 - \mathrm{i}\|\mathrm{i}\psi+\varphi\|^2\right]$$

defines an inner product whose associated norm is $\|\cdot\|\equiv\sqrt{\langle\cdot,\cdot\rangle}$. Conversely if the parallelogram law is violated then *no* inner-product may be defined compatible with that norm.

Proof. Left as an exercise to the reader. □

Example C.7 (Normed vector space which is not an inner product space)**.** Consider the space $\mathbb{C}^n$, but now with the L^1 norm

$$\|z\|_1 := \sum_{j=1}^{n}|z_j|\ .$$

Convince yourself that it is indeed a norm, and furthermore, that it violates the parallelogram law and hence cannot be associated with any inner product.

Another example we will see later is that the space of bounded linear operators on a Hilbert space is a normed vector space which is not an inner-product space.

We will continue with inner product spaces a little later, but for now we focus on *normed* vector spaces.

To any norm $\|\cdot\|$ a metric is associated via

$$\begin{aligned} d : V^2 &\to [0,\infty) \\ (\psi,\varphi) &\mapsto \|\psi-\varphi\|\,. \end{aligned}$$

Hence every normed vector space is also a metric space automatically. Recall that a metric space is termed *complete* if every Cauchy sequence on it converges.

Definition C.8 (Banach space)**.** If a normed vector space $(V, \|\cdot\|)$ is complete when regarded as a metric space, then we refer to it as a *Banach* space.

Example C.9. It is clear that $\mathbb{C}^n$ as a TVS is also a Banach space with the Euclidean norm.

Example C.10 (Counter-example)**.** Let $X := \{ f : [0,1] \to \mathbb{C} \mid f \text{ is continuous} \}$. On X define pointwise addition and multiplication, which makes it into a VS. We furthermore define on it the L^2-norm

$$\|f\|_2 := \sqrt{\int_{t\in[0,1]} |f(t)|^2 \, \mathrm{d}t} \, .$$

One shows that on X, $\|\cdot\|_2$ is indeed a norm. However, one may find Cauchy-sequences in X which converge to discontinuous functions (i.e. do not converge in X) and thus X is incomplete. Contrast this with $(X, \|\cdot\|_\infty)$ which *is* a Banach space.

Here is an L^2-Cauchy sequence of continuous functions converging to a discontinuous function:

$$f_n(t) := \chi_{\left[\frac{1}{2}+2^{-n},1\right]}(t) + \chi_{\left[\frac{1}{2}-2^{-n},\frac{1}{2}+2^{-n}\right]}(t) \left(2^{n-1}t - 2^{n-2} + \frac{1}{2}\right) \qquad (t \in [0,1], n \in \mathbb{N}) \, .$$

One shows that

$$\|f_n - f_m\|_2 \quad \leq \quad 2^{-n} \qquad (m \geq n)$$

and the sequence is hence Cauchy. But alsom, $\left\|\chi_{\left[\frac{1}{2},1\right]} - f_n\right\|_2 \to 0$ and $\chi_{\left[\frac{1}{2},1\right]} \notin X$.

Definition C.11 (Dense subsets)**.** If $(V, \|\cdot\|)$ is a Banach space and $S \subseteq V$ then we say S is *dense* in V iff for any $\psi \in V$ and any $\varepsilon > 0$ there exists some $\varphi \in S$ such that

$$d(\psi, \varphi) < \varepsilon \, .$$

This definition agrees with the topological one (S is dense iff $\overline{S} = V$).

Definition C.12 (Separable spaces)**.** If $(V, \|\cdot\|)$ is a Banach space which contains a countable, dense subset then V is called *separable.*

C.1 The operator norm

Given any two Banach spaces X, Y, we may consider a *continuous linear* map

$$A : X \to Y \, .$$

Such maps are automatically bounded: If A is continuous then A maps bounded sets of X to bounded sets of Y. We rephrase this as saying: If $A : X \to Y$ is continuous, then

$$A(B_r(0_X)) \quad \subseteq \quad B_M(0_Y) \, .$$

In other words,

$$\sup_{\|x\|_X \leq r} \|Ax\|_Y \quad < \quad \infty \, .$$

An extremely useful notion in this regard for continuous linear maps is that of the

Definition C.13 (The operator norm)**.** Given a linear map $A : X \to Y$ between Banach spaces, we define its *operator norm* as

$$\|A\|_{\mathcal{B}(X\to Y)} \quad := \quad \sup(\{ \|Ax\|_Y \mid x \in X : \|x\| \leq 1 \})$$

and $\mathcal{B}(X \to Y)$ as the space of all *bounded linear maps.* I.e., the operator norm gives us the maximal scaling of the unit ball in the domain.

Claim C.14. The "operator norm" is indeed a norm.

Proof. Absolute homogeneity is clear. Now if $\|A\|_{\mathcal{B}(X\to Y)} = 0$ then $\|Ax\|_Y = 0$ for all $\|x\| \leq 1$, which implies that $Ax = 0$ for all x, and hence $A = 0$. Finally, the triangle inequality follows by that of $\|\cdot\|_Y$:

$$\|(A+B)\,x\|_Y \leq \|Ax\|_Y + \|Bx\|_Y \qquad (\|x\| \leq 1)\ .$$

Take now the supremum over $\|x\| \leq 1$ of both sides to obtain

$$\begin{aligned} \sup_{\|x\|\leq 1} \|(A+B)\,x\|_Y &\leq \sup_{\|x\|\leq 1} \left[\|Ax\|_Y + \|Bx\|_Y\right] \\ &\leq \left(\sup_{\|x\|\leq 1} \|Ax\|_Y\right) + \sup_{\|x\|\leq 1} \|Bx\|_Y\ . \end{aligned}$$

□

Summarizing the above succinctly, if $A : X \to Y$ is linear and continuous, then

$$\|A\|_{\mathcal{B}(X\to Y)} < \infty\ .$$

Claim C.15. If $A : X \to Y$ is a linear map between two Banach spaces and if $\|A\|_{\mathcal{B}(X\to Y)} < \infty$ then A is continuous.

Proof. Given $x \in X$ and $\varepsilon > 0$, we show continuity at x as follows: for any $\tilde{x} \in B_{\frac{\varepsilon}{\|A\|}}(x)$, we have (using Lemma C.16 right below)

$$\begin{aligned} \|Ax - A\tilde{x}\| &= \|A\,(x-\tilde{x})\| \\ &\leq \|A\|\|x-\tilde{x}\| \\ &< \varepsilon\ . \end{aligned}$$

□

Lemma C.16. *If $A : X \to Y$ is a bounded linear map between two Banach spaces then*

$$\|Ax\|_Y \leq \|A\|_{\mathcal{B}(X\to Y)}\|x\|_X\ .$$

Proof. We write thanks to the homogeneity of the norm,

$$\begin{aligned} \|Ax\|_Y &= \frac{\|Ax\|_Y}{\|x\|_X}\|x\|_X \\ &= \left\|A\frac{x}{\|x\|_X}\right\|_Y \|x\|_X\ . \end{aligned}$$

But since

$$\left\|\frac{x}{\|x\|_X}\right\|_X = 1$$

we must have

$$\left\|A\frac{x}{\|x\|_X}\right\|_Y \leq \|A\|_{\mathcal{B}(X\to Y)}\ .$$

□

Lemma C.17. *The operator norm is submultiplicative: If $A, B : X \to X$ then*

$$\|AB\|_{\mathcal{B}(X)} \leq \|A\|_{\mathcal{B}(X)}\|B\|_{\mathcal{B}(X)}\ .$$

Proof. We have thanks to the above

$$\|ABx\| \leq \|A\|_{\mathcal{B}(X)} \|Bx\|$$

taking the supremum over $\|x\| \leq 1$ on both sides we obtain the result. □

Claim C.18 (R&S Thm. III.2). If X, Y are two Banach spaces then $\mathcal{B}(X, Y)$ together with the operator norm is itself a Banach space.

Proof. Thanks to Claim C.14 we know that $\left(\mathcal{B}(X, Y), \|\cdot\|_{\mathcal{B}(X,Y)}\right)$ is indeed a normed vector space (with pointwise addition and scalar multiplication). To show it is a Banach space we need to show it is complete. Let $\{ A_n \}_n$ be Cauchy. Then that means that $\|A_n - A_m\|_{\mathcal{B}(X,Y)}$ is small as n, m are large. This implies that for any $x \in X$,

$$\|(A_n - A_m)x\|_Y = \|A_n x - A_m x\|$$

is small. I.e., the sequence $\{ A_n x \}_x$ is Cauchy in Y. Since Y itself is a Banach space (and is hence complete) that means it converges to some $y \in Y$. Define a new operator, B, via

$$X \ni x \mapsto \lim_{n\to\infty} A_n x \in Y$$

which is clearly linear too since the limit is linear.

From the triangle inequality we have

$$\|A_n - A_m\| \geq |\|A_n\| - \|A_m\||$$

so that $\{ \|A_n\| \}_n$ is a Cauchy sequence of real numbers, and so converges to some $\alpha \in \mathbb{R}$. Hence, by definition of B,

$$\begin{aligned} \|Bx\|_Y &= \lim_{n\to\infty} \|A_n x\|_Y \\ &\leq \lim_{n\to\infty} \|A_n\|_{\mathcal{B}(X\to Y)} \|x\|_X \\ &= \alpha \|x\|_X . \end{aligned}$$

Hence B is bounded, and so continuous. We want to show that $\lim_n A_n = B$ *in operator norm*. We have, by definition of B,

$$\|(B - A_m)x\|_Y = \lim_{n\to\infty} \|(A_n - A_m)x\|_Y$$

so that for $\|x\| \leq 1$ we have

$$\|(B - A_m)x\|_Y \leq \lim_{n\to\infty} \|A_n - A_m\|_{\mathcal{B}(X\to Y)}$$

which implies

$$\|B - A_m\|_{\mathcal{B}(X\to Y)} \leq \lim_{n\to\infty} \|A_n - A_m\|_{\mathcal{B}(X\to Y)} .$$

The right hand side however becomes arbitrarily small for large m. □

Definition C.19. A linear map $A : X \to Y$ between Banach spaces is called an *isometry* iff $\|Ax\|_Y = \|x\|_X$ for any $x \in X$.

Claim C.20. A closed vector subspace of a Banach space is itself a Banach space.

D Hilbert spaces

Definition D.1 (Inner product space)**.** An inner-product space is a $\mathbb{C}$-vector space $\mathcal{H}$ along with a sesquilinear map

$$\langle\cdot,\cdot\rangle : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$$

such that

1. Conjugate symmetry:
$$\langle\psi,\varphi\rangle = \overline{\langle\varphi,\psi\rangle} \qquad (\varphi,\psi\in\mathcal{H})\ .$$

2. Linearity in second argument:
$$\langle\psi,\alpha\varphi+\tilde{\varphi}\rangle = \alpha\langle\psi,\varphi\rangle + \langle\psi,\tilde{\varphi}\rangle \qquad (\varphi,\tilde{\varphi},\psi\in\mathcal{H},\alpha\in\mathbb{C})\ .$$

3. Positive-definite:
$$\langle\psi,\psi\rangle > 0 \qquad (\psi\in\mathcal{H}\setminus\{\,0\,\})\ .$$

To every inner product one immediately may associate a norm, via

$$\|\psi\| := \sqrt{\langle\psi,\psi\rangle} \qquad (\psi\in\mathcal{H})\ .$$

The converse, however, hinges on the norm obeying the parallelogram law as we have seen in Claim 5.32.

Claim D.2. Once we have an inner-product, we immediately have the Cauchy-Schwarz inequality

$$|\langle\varphi,\psi\rangle| \leq \|\varphi\|\|\psi\| \qquad (\varphi,\psi\in\mathcal{H})\ .$$

Definition D.3 (Hilbert space)**.** A Hilbert space $\mathcal{H}$ is a inner-product space with $\langle\cdot,\cdot\rangle$ such the induced norm $\|\cdot\|$ from this inner product makes $\mathcal{H}$ into a Banach space (i.e. a complete metric space w.r.t. to the metric induced by $\|\cdot\|$).

Hence we identify a Hilbert space as a Banach space whose norm satisfies the parallelogram identity.

One of the central notions available to us now in Hilbert space, which was not available before, is the notion of *orthogonality* of vectors:

Definition D.4 (Orthogonality)**.** Two vectors $\varphi,\psi\in\mathcal{H}$ in a Hilbert space are dubbed *orthogonal* iff

$$\langle\varphi,\psi\rangle_{\mathcal{H}} = 0\,.$$

A collection $\{\,\varphi_i\,\}_i$ is called *orthonormal* iff

$$\langle\varphi_i,\varphi_j\rangle = \delta_{ij}\ .$$

The following two claims involving orthogonality will be useful. Their proof will be a homework assignment.

Claim D.5. $\varphi\perp\psi$ iff

$$\|\varphi\| \leq \|z\psi+\varphi\| \qquad (\forall z\in\mathbb{C})\ .$$

Proof. We have

$$0 \leq \|z\psi+\varphi\|^2 = |z|^2\|\psi\|^2 + \|\varphi\|^2 + 2\,\mathbb{Re}\{\langle z\psi,\varphi\rangle\}\ . \tag{D.1}$$

Hence if $\langle\psi,\varphi\rangle = 0$ we have $\|z\psi+\varphi\|^2 \geq \|\varphi\|^2$.

Conversely, if $\psi = 0$ we are finished. Otherwise, let $z := -\frac{\langle\psi,\varphi\rangle}{\|\psi\|^2}$. Plugging this into (D.1) we find

$$\begin{aligned} 0 \leq \|z\psi+\varphi\|^2 &= \|\varphi\|^2 + |z|^2\|\psi\|^2 + 2\,\mathbb{Re}\{\langle z\psi,\varphi\rangle\} \\ &= \|\varphi\|^2 + \frac{|\langle\psi,\varphi\rangle|^2}{\|\psi\|^2} + 2\,\mathbb{Re}\left\{\left\langle -\frac{\langle\psi,\varphi\rangle}{\|\psi\|^2}\psi,\varphi\right\rangle\right\} \\ &= \|\varphi\|^2 - \frac{|\langle\psi,\varphi\rangle|^2}{\|\psi\|^2} \end{aligned}$$

which is coincidentally a proof of the Cauchy-Schwarz inequality. But this also shows that

$$\|z\psi+\varphi\|^2 < \|\varphi\|^2$$

for *one* z if $\langle\varphi,\psi\rangle \neq 0$. □

Claim D.6 (Cauchy-Schwarz). For any $f, g \in L^2(\mu)$ we have

$$\left|\langle f,g\rangle_{L^2(\mu)}\right| \leq \|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)} .$$

Theorem D.7. *Every nonempty closed convex set $E \subseteq \mathcal{H}$ contains a* unique *element of minimal norm.*

Proof. Let

$$d := \inf\left(\{\, \|x\| \mid x \in E \,\}\right) .$$

Let $\{\, x_n \,\}_n \subseteq E$ so that $\{\, \|x_n\| \,\}_n \to d$. Since E is convex,

$$\frac{1}{2}\left(x_n + x_m\right) \in E$$

and hence

$$\|x_n + x_m\|^2 = 4\left\|\frac{1}{2}\left(x_n + x_m\right)\right\|^2 \geq 4d^2 .$$

Then the parallelogram law Claim 5.32

$$\|x_n + x_m\|^2 + \|x_n - x_m\|^2 = 2\|x_n\|^2 + 2\|x_m\|^2$$

has its right hand side tend to $4d^2$ also, so $\|x_n - x_m\|^2 \to 0$ and hence $\{\, x_n \,\}_n$ is Cauchy, which hence converges to some $x \in E$ (as E is closed) and we have $\|x\| = d$.

For uniqueness, if $y \in E$ with $\|y\| = d$, then

$$\{\, x, y, x, y, \dots \,\}$$

must converge by the above, so $y = x$. □

Theorem D.8. *Let M be a closed subspace of $\mathcal{H}$. Then*

$$M^\perp \equiv \{\, \varphi \in \mathcal{H} \mid \varphi \perp \psi \forall \psi \in M \,\}$$

is also a closed subspace of $\mathcal{H}$ and $M \cap M^\perp = \{\, 0 \,\}$, and

$$\mathcal{H} = M \oplus M^\perp .$$

Proof. Since $\langle\varphi,\cdot\rangle$ is linear, $M^\perp$ is linear. Also,

$$M^\perp = \bigcap_{\varphi \in M} \langle\varphi,\cdot\rangle^{-1}\left(\{\, 0 \,\}\right) \tag{D.2}$$

and $\langle\varphi,\cdot\rangle$ is continuous by the Cauchy-Schwarz inequality, so $M^\perp$ is closed. Next, if $\varphi \in M \cap M^\perp$ then in particular $\langle\varphi,\varphi\rangle = 0$ so $\varphi = 0$.

Finally, let $\varphi \in \mathcal{H}$. The set $\varphi - M$ is a convex, closed subset and hence by Theorem D.7 we get some $\psi \in M$ such that $\|\varphi - \psi\|$ is minimal. Let

$$\eta := \varphi - \psi .$$

Then

$$\|\eta\| \leq \|\eta + \xi\| \qquad (\xi \in M)$$

by the minimizing property. So by Claim D.5, $\eta \in M^\perp$. But

$$\varphi = \psi + \eta \in M + M^\perp \,.$$

□

An important structural feature of the Hilbert space structure for us will be the Riesz representation theorem (not to be confused with the *related* Kakutani-Markov-Riesz representation theorem appearing in Theorem 2.84).

Definition D.9 (Bounded linear functionals)**.** A *bounded linear functional* on a Hilbert space $\mathscr{H}$ is a $\mathbb{C}$-linear map

$$\Lambda : \mathscr{H} \to \mathbb{C}$$

whose *operator norm* is finite:

$$\|\Lambda\|_{\mathrm{op}} \equiv \sup \left(\left\{ |\Lambda\varphi| \;\middle|\; \varphi \in \mathscr{H} : \|\varphi\|_{\mathscr{H}} = 1 \right\} \right) \overset{!}{<} \infty \,.$$

The space of all bounded linear functionals on a Hilbert space $\mathscr{H}$ is denoted by $\mathscr{H}^*$: it is the dual space.

Unlike for Banach space, the fact these maps act on a Hilbert space exhibits $\mathcal{B}(\mathscr{H} \to \mathbb{C}) \equiv \mathscr{H}^*$, the dual, as isomorphic to $\mathscr{H}$ itself. I.e., Hilbert spaces are self-dual, or reflexive. This is the Riesz theorem:

Theorem D.10. *There is an anti-$\mathbb{C}$-linear isometric bijection $K : \mathscr{H} \to \mathscr{H}^*$ given by*

$$\varphi \mapsto \langle \varphi, \cdot \rangle \,.$$

In particular, every bounded linear map $\Lambda : \mathscr{H} \to \mathbb{C}$ is the result of an inner product with *some* vector $\varphi_\Lambda \in \mathscr{H}$:

$$\Lambda = \langle \varphi_\Lambda, \cdot \rangle_{\mathscr{H}} \,.$$

Proof. Clearly K is anti-$\mathbb{C}$-linear. To show it is isometric, we have by the Cauchy-Schwarz inequality

$$\begin{aligned} \|K(\varphi)\|_{\mathrm{op}} &\equiv \sup(\{ (K(\varphi))(\psi) \mid \|\psi\| = 1 \}) \\ &\equiv \sup(\{ |\langle \varphi, \psi \rangle| \mid \|\psi\| = 1 \}) \\ &\overset{\mathrm{CS}}{\leq} \sup(\{ \|\varphi\| \|\psi\| \mid \|\psi\| = 1 \}) \\ &= \|\varphi\| \,. \end{aligned}$$

But also,

$$\|\varphi\|^2 \equiv \langle \varphi, \varphi \rangle \equiv (K(\varphi))(\varphi) \leq \|K(\varphi)\| \|\varphi\|$$

where the last inequality was Lemma C.16. Hence $\|K(\varphi)\| = \|\varphi\|$ so K is an isometry. But an isometry is always injective, is it merely remains to show that K is surjective.

Let then $\lambda \in \mathscr{H}^*$. If $\lambda = 0$ then $K(0_{\mathscr{H}}) = 0 = \lambda$. Otherwise, since $\ker(\lambda)$ is a closed linear subspace, the proof above in Theorem D.8 says

$$\mathscr{H} = \ker(\lambda) \oplus \ker(\lambda)^\perp \,.$$

Let therefore $\eta \in \ker(\lambda)^\perp$ and $\eta \neq 0$. Since by linearity we have

$$[(\lambda\psi)\,\eta - (\lambda\eta)\,\psi] \in \ker(\lambda) \qquad (\psi \in \mathscr{H})$$

we have

$$\begin{aligned} 0 &= \langle \eta, [(\lambda\psi)\,\eta - (\lambda\eta)\,\psi] \rangle \\ &= (\lambda\psi)\,\langle \eta, \eta \rangle - (\lambda\eta)\,\langle \eta, \psi \rangle \end{aligned}$$

i.e.,

$$\lambda\psi = \left\langle \frac{\overline{(\lambda\eta)}}{\|\eta\|^2}\eta, \psi \right\rangle$$

or

$$\lambda = \left\langle \frac{\overline{(\lambda\eta)}}{\|\eta\|^2}\eta, \cdot \right\rangle .$$

□

E Urysohn's Lemma [Rudin]

Thanks to Olivia Kwon for contributing this section about the Urysohn lemma.

The purpose of this section is to prove Urysohn's Lemma, which will be used in proving Theorem 2.84.

Definition E.1. We denote the space of all continuous compactly supported function by $C_c(X)$.

Given compact set K, if $f \in C_c(X)$ such that $f(x) = 1$ for all $x \in K$ and $f(x) \in [0, 1]$ for all $x \in X$, we write

$$K \prec f.$$

Moreover, if V open such that $f \in C_c(X)$ has range in $[0, 1]$ and satisfies satisfying $\text{supp}(f) \subset V$, we write

$$f \prec V.$$

Remark E.2. Notice that $f \prec V$ is stronger statement than $0 \le f \le \chi_V$ which only implies that $\text{supp}(f) \subset \bar{V}$.

Proposition E.3. *Suppose X is locally compact Hausdorff space. Suppose V is open set containing a compact set K. Then, there exists some open set U with compact closure such that $K \subset U \subset \bar{U} \subset V$.*

Proof. Observe that every point of K has a neighborhood with compact closure by definition of locally compact space. Also, observe that finitely many of these neighborhoods cover K. Therefore, we have that K lies in an open set O with a compact closure.

If $V = X$, simply take $U = O$.

Suppose $V \neq X$. Let $C = V^c$.

For every point $p \in C$, we can construct an open set O_p such that $p \notin \bar{O}_p$ and $K \subset O_p$. To do so, for all $x \in K$, using the definition of Hausdorff and the fact that $K \cap C = \emptyset$, we find a neighborhood of x, G_x, such that $p \notin \bar{G}_x$. Because K is compact, K can be covered by finitely many such G_x's, say $G_{x_1}, \cdots, G_{x_N}$. Now define $O_p = \bigcup_{n=1}^{N} G_{x_n}$. One can check easily that this O_p has all the desired properties.

Define $F_p = C \cap \bar{O} \cap \bar{O}_p$ for each $p \in C$. Then, we have a collection of compact sets $\{F_p\}_{p \in C}$ with empty intersection, i.e. $\bigcap_{p \in C} F_p = \emptyset$.

We claim that we can find some finite subcollection $F_{p_1}, \cdots, F_{p_M}$ such that $\bigcap_{n=1}^{M} F_{p_n} = \emptyset$. This will finish the proof. Observe that $\bigcap_{n=1}^{M} F_{p_n} = \emptyset$ implies $\bar{O} \cap \bar{O}_{p_1} \cap \cdots \cap \bar{O}_{p_M} \subset V$. Therefore, let

$$U = O \cap O_{p_1} \cap \cdots \cap O_{p_M}.$$

Note that because $\bar{U} \subset \bar{O} \cap \bar{O}_{p_1} \cap \cdots \cap \bar{O}_{p_M} \subset V$ and because $K \subset U$, we are done.

Hence it only remains to show the existence of such $F_{p_1}, \cdots, F_{p_M}$. For every $p \in C$, let $W_p = F_p^c$. Fix F_1 in $\{F_p\}_{p \in C}$. Because no point of F_1 belong to every $F_p \in \{F_p\}_{p \in C}$, we see that $\{W_p\}$ is an open cover of F_1. Take a finite subcover using compactness, say $W_{p_1}, \cdots, W_{p_{M-1}}$. Let$F_{p_1} = W_{p_1}^c, \cdots, F_{p_{M-1}} = W_{p_{M-1}}^c$, and $F_1 = F_{p_M}$. Then, by construction, we get that: $\bigcap_{n=1}^{M} F_{p_n} = \emptyset$. □

Definition E.4. Let f be a real function on a topological space. If $\{x : f(x) > \alpha\}$ is open for every $\alpha \in \mathbb{R}$, we say f is *lower semicontinuous*. If $\{x : f(x) < \alpha\}$ is open for every $\alpha \in \mathbb{R}$, we say f is *upper semi-continuous*.

Remark E.5. Notice that a real function is continuous if and only if it is both upper and lower semicontinuous. Moreover, The supremum of lower semicontinuous functions is lower semicontinuous and the infimum of upper semi-continuous is upper semicontinuous.

Theorem E.6. *(Uryshon's Lemma) [Rudin's Proof] S Given X a locally compact Hausdorff space, V open in X and compact $K \subset V$, there exists some $f \in C_c(X)$ such that $K \prec f \prec V$.*

Proof. Put $r_1 = 0$, $r_2 = 1$, and $\{r_n\}_{n\geq 3}$ be enumeration of rationals in $(0,1)$. Because X is locally compact and Hausdorff, by applying Proposition Proposition E.3 we can find open set V_0 with compact closure such that

$$K \subset V_0 \subset \bar{V_0} \subset V.$$

Now applying it again on K and V_0, we can find V_1 such that

$$K \subset V_1 \subset \bar{V_1} \subset V_0 \subset \bar{V_0} \subset V.$$

For fixed $n \geq 2$, choose open sets with compact closures $V_{r_1}, \cdots, V_{r_n}$ such that $r_i < r_j$ implies $\bar{V_{r_j}} \subset V_{r_i}$ using above proposition again. Find r_j among $r_1, \cdots, r_n$ such that r_i is the biggest one smaller than r_{n+1}. Moreover, find r_j among $r_1, \cdots, r_n$ such that it is the smallest one bigger than r_{n+1}. We know the existence of r_i and r_j are guaranteed because $r_1 = 1$ and $r_2 = 1$. They are unique because rationals are well-ordered. Now using the proposition again, we can find $V_{r_{n+1}}$ such that

$$\bar{V_{r_j}} \subset V_{r_{n+1}} \subset \bar{V_{r_{n+1}}} \subset V_{r_i}.$$

Recursively, we can define $\{V_{r_n}\}_{n\in\mathbb{N}}$ such that $K \subset V_1, \bar{V_0} \subset V$, and for every rationals of $[0,1]$ r and s, each $\bar{V_r}$ compact and $s > r$ implies $\bar{V_s} \subset V_r$.

Now, for all $r \in \mathbb{Q}\cap[0,1]$, define $f_r, g_r : X \to \mathbb{R}$, $f_r, g_r \in C_c(X)$, by:

$$f_r(x) = \begin{cases} r \text{ if } x \in V_r \\ 0 \text{ else.} \end{cases} \quad \text{and} \quad g_r(x) = \begin{cases} 1 \text{ if } x \in \bar{V_r}, \\ r \text{ else.} \end{cases} .$$

Observe that for every r, f_r is lower semicontinuous while g_r is upper semicontinuous. Moreover, define

$$f = \sup_r f_r \quad \text{and} \quad g = \inf_r g_r.$$

By construction, because f is the supremum of lower semicontinuous functions, f is lower semicontinuous. Similarly, because g is the infimum of upper semicontinuous, g is upper semicontinuous.

We claim that $f = g$. This will finish the proof because if $f = g$, then f is both lower and upper semicontinuous implying that it is continuous. Moreover, it is visible that $f(x) = 1$ if $x \in K$ and $\text{supp}(f) \subset \bar{V_0} \subset V$, , so we get that $K \prec f \prec V$, as desired.

To show $f = g$, we first show that $f \leq g$ then show that $f < g$ is impossible.

Notice that $f_r(x) > g_s(x)$ is only possible if $r > s, x \in V_r$, and $x \notin \bar{V_s}$. However, $r > s$ implies $V_r \subset V_s$ and hence this is impossible. Therefore, for all r, s, we have that $f_r \leq g_s$ and thus limits, we get $f \leq g$.

Now suppose for the sake of contradiction that there exists some $x \in X$ such that $f(x) < g(x)$. This means that there are rationals r and s such that $f(x) < r < s < g(x)$. Since $f(x) < r$, we get that $x \notin V_r$. Since $g(x) > s$, we get that $x \in \bar{V_s}$. However, this contradicts the constructions of $\{V_r\}$ because $s > r$ implies $\bar{V_s} \subset V_r$. Therefore, this is impossible and we get that $f(x) = g(x)$ for all $x \in X$. □

Corollary E.7. *(Corollary to Uryshon's Lemma) Given $V_1, \cdots V_N$ open subsets of X and K compact such that $K \subset V_1 \cup \cdots \cup V_N$, there exists $h_1, \cdots, h_N \in C_c(X)$ such that $h_i \prec V_i$ for all $1 \leq i \leq N$ and $h_1(x) + \cdots + h_N(x) = 1$ for all $x \in K$.*

Proof. We know for each $x \in K$, there exists some i such that $x \in V_i$. Using Prop Proposition E.3, find an open neighborhood W_x such that $x \subset W_x$ and $\bar{W_x} \subset V_i$. By compactness, we can find finite points $x_1, \cdots, x_m$ such that $K \subset W_{x_1} \cup \cdots \cup W_{x_m}$. For every $1 \leq i \leq n$, let $H_i = \bigcup_{\bar{W_{x_j}} \subset V_i} \bar{W_{x_j}}$. By construction, each H_i is compact and contained in V_i.

Using Uryshon's Lemma, find g_i satisfying $H_i \prec g_i \prec V_i$. Now, define $h_1, \cdots, h_n \in C_c(X)$ by:

$$\begin{aligned} h_1 &= g_1 \\ h_2 &= (1-g_1)\cdot g_2 \\ h_3 &= (1-g_1)(1-g_2)\cdot g_3 \\ &\vdots \\ h_n &= (1-g_1)\cdots(1-g_{n-1})g_n. \end{aligned}$$

Notice by construction that each $h_i \prec V_i$. Moreover, we can easily compute, using induction, that

$$h_1 + h_2 + \cdots + h_n = 1 - (1-g_1)(1-g_2)\cdots(1-g_n).$$

Because $K \subset H_1 \cup \cdots \cup H_n$, we see that for all $x \in K$, there exists g_i such that $g_i(x) = 1$ and therefore we get that $h_1(x) + \cdots + h_n(x) = 1$ on K. □

F Glossary of mathematical symbols and acronyms

Sometimes it is helpful to include mathematical symbols which can function as valid grammatical parts of sentences. Here is a glossary of some which might appear in the text:

- $\mathrm{im}\,(f)$ is the *range* or *image* of a function: If $f : X \to Y$ then

$$\mathrm{im}\,(f) \equiv \{\, f(x) \in Y \mid x \in X \,\}\ .$$

- The bracket $\langle\cdot,\cdot\rangle_V$ means an inner product on the inner product space V. For example,

$$\langle u, v\rangle_{\mathbb{R}^2} \equiv u_1 v_1 + u_2 v_2 \qquad \left(u, v \in \mathbb{R}^2\right)$$

  and

$$\langle u, v\rangle_{\mathbb{C}^2} \equiv \overline{u_1} v_1 + \overline{u_2} v_2 \qquad \left(u, v \in \mathbb{C}^2\right)\ .$$

- Sometimes we denote an integral by writing the integrand without its argument. So if $f : \mathbb{R} \to \mathbb{R}$ is a real function, we sometimes in shorthand write

$$\int_a^b f$$

  when we really mean

$$\int_{t=a}^b f(t)\,\mathrm{d}t\,.$$

  This type of shorthand notation will actually also apply for contour integrals, in the following sense: if $\gamma : [a,b] \to \mathbb{C}$ is a contour with image set $\Gamma := \mathrm{im}\,(\gamma)$ and $f : \mathbb{C} \to \mathbb{C}$ is given, then the contour integral of f along γ will be denoted equivalently as

$$\int_\Gamma f \equiv \int_\Gamma f(z)\,\mathrm{d}z \equiv \int_{t=a}^b f(\gamma(t))\,\gamma'(t)\,\mathrm{d}t$$

  depending on what needs to be emphasized in the context. Sometimes when the contour is clear one simply writes

$$\int_{z_0}^{z_1} f(z)\,\mathrm{d}z$$

  for an integral along *any* contour from z_0 to z_1.

- iff means "if and only if", which is also denoted by the symbol $\iff$.
- WLOG means "without loss of generality".
- CCW means "counter-clockwise" and CW means "clockwise".
- $\exists$ means "there exists" and $\nexists$ means "there does not exist". $\exists!$ means "there exists *a unique*".

- $\forall$ means "for all" or "for any".
- : (i.e., a colon) may mean "such that".
- ! means negation, or "not".
- $\wedge$ means "and" and $\vee$ means "or".
- $\Longrightarrow$ means "and so" or "therefore" or "it follows".
- $\in$ denotes set inclusion, i.e., $a \in A$ means a is an element of A or a lies in A.
- $\ni$ denotes set inclusion when the set appears first, i.e., $A \ni a$ means A includes a or A contains a.
- Speaking of set inclusion, $A \subseteq B$ means A is contained within B and $A \supseteq B$ means B is contained within A.
- $\varnothing$ is the empty set $\{ \ \}$.
- $\aleph_0$ is the cardinality of $\mathbb{N}$: $\aleph_0 := |\mathbb{N}|$. $\mathfrak{c} := 2^{\aleph_0} = |\mathbb{R}|$ is the cardinality of the continuum.
- While $=$ means equality, sometimes it is useful to denote types of equality:
  - $a := b$ means "this equation is now the instant when a is defined to equal b".
  - $a \equiv b$ means "at some point above a has been defined to equal b".
  - $a = b$ will then simply mean that the result of some calculation *or* definition stipulates that $a = b$.
  - Concrete example: if we write $\mathrm{i}^2 = -1$ we don't specify anything about *why* this equality is true but writing $\mathrm{i}^2 \equiv -1$ means this is a matter of definition, not calculation, whereas $\mathrm{i}^2 := -1$ is the first time you'll see this definition. So this distinction is meant to help the reader who wonders *why* an equality holds.

F.1 Important sets

1. The unit circle

$$\mathbb{S}^1 \quad \equiv \quad \{ \, z \in \mathbb{C} \mid |z| = 1 \, \} \, .$$

2. The (open) upper half plane

$$\mathbb{H} \quad \equiv \quad \{ \, z \in \mathbb{C} \mid \mathbb{Im}\{z\} > 0 \, \} \, .$$

References

[Bis] Marek Biskup. math.ucla.edu. https://www.math.ucla.edu/~biskup/275d.1.24w/PDFs/ch2.pdf. [Accessed 21-04-2025].

[Bog06] Vladimir I Bogachev. *Measure Theory*. Springer, Berlin, Germany, 2007 edition, November 2006.

[Dur19] Rick Durrett. *Probability*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, England, 5 edition, April 2019.

[DZ09] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer, Berlin, Germany, 2 edition, November 2009.

[Fol99] Gerald B Folland. *Real analysis*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, Nashville, TN, 2 edition, March 1999.

[FR10] Patrick M Fitzpatrick and H L Royden. *Real Analysis*. Pearson, Upper Saddle River, NJ, 4 edition, January 2010.

[Hat01] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, England, December 2001.

[Jak06] Vojkan Jakšić. *Topics in Spectral Theory*, pages 235–312. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[Par67] K R Parthasarathy. *Probability measures on metric spaces.* 1967.

[Rud76] W. Rudin. *Principles of Mathematical Analysis.* International series in pure and applied mathematics. McGraw-Hill, 1976.

[Rud86] Walter Rudin. *Real and Complex Analysis (Higher Mathematics Series).* McGraw-Hill Education, 1986.

[Sch66] M Schilder. Some asymptotic formulas for wiener integrals. *Trans. Am. Math. Soc.*, 125(1):63–85, 1966.

[Sha23a] Jacob Shapiro. Lecture notes on complex analysis with applications. [Online; accessed 14-May-2023], May 2023.

[Sha23b] Jacob Shapiro. Lecture notes on functional analysis. [Online; accessed 30-Oct-2023], Oct 2023.

[Sim04] Barry Simon. *Functional integration and quantum physics.* AMS Chelsea Publishing. American Mathematical Society, Providence, RI, December 2004.

[Sim15] Barry Simon. *Real Analysis.* American Mathematical Society, Providence, RI, October 2015.

[SS05] Elias M Stein and Rami Shakarchi. *Real analysis.* Princeton lectures in analysis. Princeton University Press, Princeton, NJ, March 2005.

[Tes09] Gerald Teschl. *Mathematical methods in quantum mechanics : with applications to Schroedinger operators.* American Mathematical Society, Providence, R.I, 2009.

[Var84] S R S Varadhan. *Large Deviations and Applications.* Society for Industrial and Applied Mathematics, January 1984.