Notes on Probability and Measure Theory

February 18, 2022

Contents

1 Overview

1.1 References

The primary reference here is [Ash et al., 2000]. The book is wonderful for statistical machine learning – it is rigorous, but also accessible (prerequisites are undergrad-level real analysis and mathematical probability). Most importantly, it is structured to build towards the kinds of applications in probability that we care about. (A point of contrast would be a book like that of Stein and Shakarchi, which tends to dwell heavily on things that are of higher interest to pure mathematicians — long existence proofs, Cantor sets and fractals, etc.)

Unless otherwise specified, all references to the “text" refers to this textbook. Likewise the symbol § refers to a Section of that textbook.

Other useful references are [Folland, 1999] and [Rudin, 1987].

1.2 Motivation for topic

What are some motivations for measure theory?

- Measure theory underpins some of the most interesting research in Bayesian statistics and probabilistic machine learning (see work from Stephen G. Walker, Michael Jordan, Tamara Broderick, David Dunson, and so on). Thus, fluency with measure theory opens doors to a higher level of research consumption.
- Measure theory underpins research on stochastic processes (as used in Bayesian nonparametrics) and stochastic differential equations (useful for continuous-time time series models, a current topic of active research interest in machine learning).
- Measure theory is convenient in unifying various kinds of random variables.[1]
- Lebesgue integration provide nice limit theorems, e.g. clarifying when one can interchange integrals and limits (such as derivatives).
- Lebesgue integrals can be seen as the completion of Riemann integrals (in the same way that the real numbers complete the rationals).
- Abstract Lebesgue integration allows one to integrate over spaces more general than the reals.

1.3 Motivation for notes

It is hard to beat directly consulting a textbook (such as [Ash et al., 2000]) written by a seasoned mathematician who is an excellent pedagogue. However, we have created these notes nonetheless in an attempt to *support lecture and/or discussion*. With that goal in mind, we:

- Format the presentation to encourage easier absorption.[2]
- Add sketches to support intuition.
- Curate the text.[3]
- Provide additional detail in proofs. Sometimes alternate paths have been given that seemed “nicer" to me.

[1]For example, it allows one to work with discrete and absolutely continuous random variables in a unified way. For example, the exponential family includes both types of random variables.

[2]E.g., we exploit space to organize the presentation, whereas a textbook will often provide proofs in paragraph form. We sometimes refactor presentations into more modular subsections.

[3]We highlight some of the main themes (and cores of proofs), offloading additional detail to the text.

- Add remarks for color (illustrating the need for propositions, the utility of theorems, or connections between things).
- Incorporate supporting material from worked homework problems and outside sources.[4]

2 § 1.1: Some notes on set theory

2.1 Limits of sequences of sets

Definition 2.1.1. The **upper limit** of a sequence of sets is given by

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$$

Alternatively,

$$x \in \limsup A_n \text{ iff } x \in A_n \text{ for infinitely many } n$$

△

Definition 2.1.2. The **lower limit** of a sequence of sets is given by

$$\liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k$$

Alternatively,

$$x \in \liminf A_n \text{ iff } x \in A_n \text{ eventually (for all but finitely many } n \text{)}$$

△

Discussion 2.1.1. Discuss why the two characterizations of upper limit and lower limit are equivalent. △

Definition 2.1.3. If $\liminf A_n = \limsup A_n = A$, then A is called the **limit** of the sequence $A_1, A_2, ...$. △

Now we present a particular kind of limit that will be useful when we discuss continuity of measure.

Definition 2.1.4. If $A_1 \subset A_2 \subset ...$ and $\cup_{n=1}^{\infty} A_n = A$, we say that the A_n form a **increasing** sequence of sets with limit A or that the A_n increase to A; we write $A_n \uparrow A$. If $A_1 \supset A_2 \supset ...$ and $\cap_{n=1}^{\infty} A_n = A$, we say that the A_n form a **decreasing** sequence of sets with limit A or that the A_n decrease to A; we write $A_n \downarrow A$. △

One can verify that this definition is consistent with the definition of limits, i.e.

$$\text{If } A_n \uparrow A \text{ or } A_n \downarrow A \text{ then } \liminf A_n = \limsup A_n = A.$$

As shown in Figure 2, limits of increasing and decreasing sequences are very special kinds of limits.

[4] This will happen increasingly often as the notes evolve.

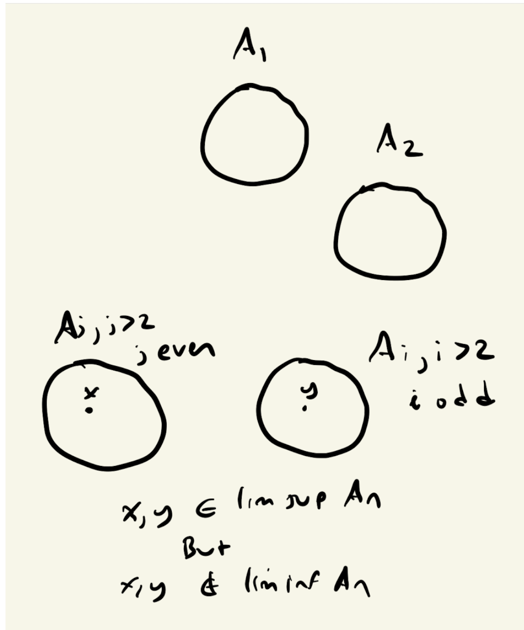


Figure 1: A sequence of sets with empty lower limit and non-empty upper limit.

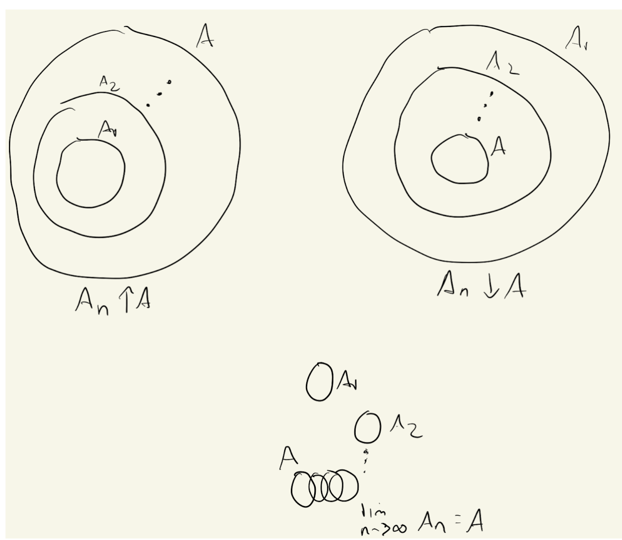


Figure 2: An increasing and decreasing sequence of sets, followed by a sequence of sets which is neither, but which has a limit.

2.2 Representing unions as disjoint unions

Remark 2.2.1. If $A_1, A_2, ...$ are subsets of some set Ω, then

$$\bigcup_{n=1}^{\infty} A_n = \dot{\bigcup_{n=1}^{\infty}} \left(A_n \cap A_{n-1}^c \cap ... \cap A_1^c \right) \tag{2.2.1}$$

In other words, any union can be re-represented as a disjoint union. This is useful because measures are countably additive on disjoint sets, so we prefer to work with collections of disjoint sets. △

Remark 2.2.2. If $A_n \uparrow A$, then (2.2.1) becomes

$$\bigcup_{n=1}^{\infty} A_n = \dot{\bigcup_{n=1}^{\infty}} \Big(A_n - A_{n-1} \Big) \tag{2.2.2}$$

This is because $A_{n-1} \subset A_n$, so $A_{n-1}^c \supset A_n^c$ by contraposition. △

3 § 1.2: Fields, σ-fields, measures

3.1 § 1.2.1-1.2.2: Fields and σ-fields

Probability measures, and measures more generally, cannot be defined on all subsets of many spaces that we would like to deal with. For instance, non-measurable sets can be shown to exist even for Lebesgue measure on the unit interval. Proposition 1.2.6 of [Rosenthal, 2006] shows that there is no definition of $P(A)$ that is defined for all subsets $A \subseteq [0, 1]$ satisfying all three conditions below[5]

1. $P([a, b]) = b - a, \quad 0 \le a \le b \le 1.$
2. $P(\dot\bigcup_{n=1}^{\infty} A_n) = \displaystyle\sum_{n=1}^{\infty} A_n$ for $A_1, A_2, ...$ disjoint subsets of $[0, 1]$.
3. $P(A \bigoplus r) = P(A), \quad 0 \le r \le 1$, where $A \bigoplus r$ denotes the *r-shift* of A, i.e.

$$A \bigoplus r := \{a + r : a \in A, a + r \le 1\} \cup \{a + r - 1 : a \in A, a + r > 1\}$$

In Sec. 5.8, we provide more information about a set that is not Lebesgue measurable.

The solution to this problem is to define measures on a restricted domain, σ-fields.

3.1.1 σ-fields

Definition 3.1.1. Let $\mathcal{F}$ be a collection of subsets of a set Ω. Then $\mathcal{F}$ is called a **sigma-field** (or *sigma-algebra*) if it satisfies

a) $\Omega \in \mathcal{F}$
b) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
c) If $A_1, A_2, ... \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

that is, if $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and countable unions. △

Remark 3.1.1. It follows that σ-fields are closed under countable intersections, since

$$\cap_{i=1}^{\infty} A_i \overset{\text{DeMorgan's Law}}{=} \cup_{i=1}^{\infty} A_i^c$$

△

Example 3.1.1. $\mathcal{F} = \{\emptyset, \Omega\}$ is the smallest σ-field on Ω. △

Example 3.1.2. $\mathcal{F} = 2^{\Omega}$, i.e. the set of all subsets of Ω, is the largest σ-field on Ω. △

Example 3.1.3. If $A \in \Omega$ is non-empty, then $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ is the smallest σ-field containing A. △

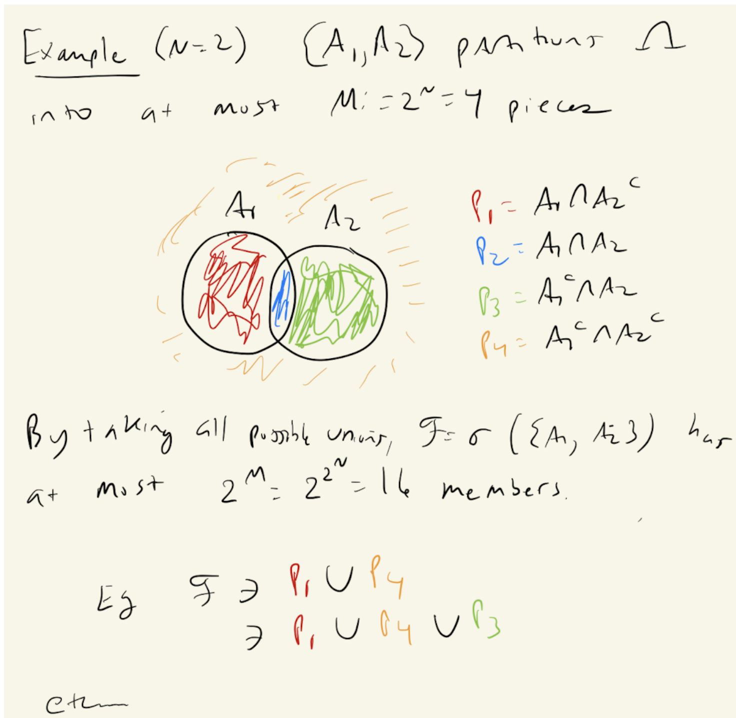
[5] In Proposition 5.8.1, we make a similar observation, along with a proof: there cannot be a measure defined on all subsets of the reals that is both translation invariant and has a finite value on all bounded intervals.

Notation 3.1.1. If $\mathcal{C}$ is a class of sets, the smallest σ-field containing the sets of $\mathcal{C}$ is written as $\sigma(\mathcal{C})$. This is sometimes called the *minimal σ-field over C* or the *σ-field generated by C*. △

Problem 3.1.1. (*[Ash et al., 2000] Problem 1.2.8*) Let $A_1, ..., A_n$ be subsets of Ω. Describe $\mathcal{F} := \sigma(\{A_1, ..., A_n\})$, the smallest σ-field containing $A_1, ..., A_n$. Also describe the number of sets in $\mathcal{F}$.

Solution. We can derive the strict upper bound $|\mathcal{F}| \leq 2^{2^n}$. For a complete answer, see GoodNotes.

The gist is that the collection $\{A_1, ..., A_n\}$ partitions Ω into up to $M = 2^N$ pieces, and the minimal sigma field contains all possible finite unions of these pieces, so has at most 2^M elements.



□

3.1.2 Fields

Fields are more general than σ-fields. Measures are sometimes constructed by being defined on fields, and then extended to σ-fields. Indeed, we will see this strategy with Lebesgue measure.

Definition 3.1.2. Let $\mathcal{F}$ be a collection of subsets of a set Ω. Then $\mathcal{F}$ is called a **field** (or *algebra*) if satisfies Definition 3.1.1 after replacing condition c) with

c') If $A_1, ...A_n \in \mathcal{F}$ then $\cup_{i=1}^n A_i \in \mathcal{F}$.

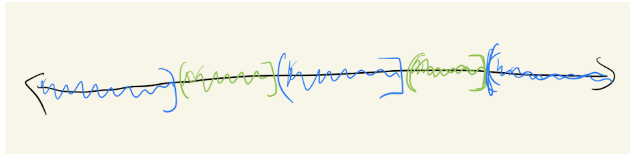
that is, if $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and *finite* unions. △

Example 3.1.4. What is an example of a collection that is a *field*, but not a *σ-field*?

Let $\Omega = \mathbb{R}$ and $\mathcal{F}_0 = \{$finite disjoint unions of right semi-closed intervals $(a, b], a \neq b\}$. Then $\mathcal{F}_0$ is a field, as can be easily verified.[6]

But $\mathcal{F}_0$ is <u>not</u> a σ-field. Note that if $A_n = (-\frac{1}{n}, 0]$, then $\bigcap_{n=1}^{\infty} A_n = \{0\} \notin \mathcal{F}_0$. △

[6] By convention, we also count (a, ∞) as right semi-closed for $-\infty \leq a < \infty$, which is necessary for the σ-field to be closed under complements.



Remark 3.1.2. If $\mathcal{F}$ is a field, a countable union of sets in $\mathcal{F}$ can be expressed as the limit of an increasing sequence of sets in $\mathcal{F}$, and conversely. For if $A_n \in \mathcal{F}$ and $A_n \uparrow A$, then A is a countable union of sets in $\mathcal{F}$ by definition. Conversely, if $A = \cup_{n=1}^{\infty} A_n$, then set $B_N := \cup_{n=1}^{N} A_n$ and $B_N \uparrow A$. This shows that a σ-field can also be described as a field that is closed under limits of increasing sequences. More generally, if $\mathcal{G}$ is the collection of all limits of increasing sequences of sets in some field $\mathcal{F}_0$, we can also describe $\mathcal{G}$ as the collection of all countable unions of sets in $\mathcal{F}_0$. $\triangle$

3.1.3 "Good sets" strategy

Ash says that there is a type of reasoning that occurs so often in problems involving σ-fields that it deserves explicit mention. It is called the *good sets strategy*. Suppose you want to show that all members of a σ-algebra $\mathcal{F}$ have some property P. Define "good sets" as those that satisfy the property

$$\mathcal{G} := \{G \in \mathcal{F} : G \text{ has property } P\}$$

The strategy is then to simply

1. Show $\mathcal{G}$ contains some class $\mathcal{C}$ such that $\mathcal{F} = \sigma(\mathcal{C})$
2. Show $\mathcal{G}$ is a σ-algebra

Then you're done!

Why does this work?

$$\begin{aligned}
& \mathcal{C} \subset \mathcal{G} && \text{by 1} \\
\implies\ & \sigma(\mathcal{C}) \subset \sigma(\mathcal{G}) && \\
\implies\ & \mathcal{F} \subset \mathcal{G} && \text{by 1,2}
\end{aligned}$$

Yet $\mathcal{G} \subset \mathcal{F}$ by definition of $\mathcal{G}$.

So $\mathcal{G} = \mathcal{F}$.

So all sets in $\mathcal{F}$ are good.

Some example applications:

- In the text, Ash uses this strategy (see pp.5) to show that if $\mathcal{C}$ is a class of subsets of Ω, and $A \in \Omega$, then

$$\underbrace{\sigma_{\Omega}(\mathcal{C}) \cap A}_{\text{take minimal sigma field first, then intersect}} \quad = \quad \underbrace{\sigma_A(\mathcal{C} \cap A)}_{\text{intersect first, then take minimal sigma-field}}$$

- See my handwritten homework exercise for § 1.2, Problem 6.

Remark 3.1.3. Later, we will cover the Monotone Class Theorem (see Theorem 4.1.2), which provides an alternate mechanism for executing the Good Sets Strategy. See Remark 4.1.2. $\triangle$

3.1.4 Borel Sets

An important example of a σ-field is the Borel Sets $\mathcal{B}(\mathbb{R})$, defined as the smallest σ-field of subsets of $\mathbb{R}$ containing all intervals $(a, b] \subset \mathbb{R}$.

We may alternately characterize $\mathcal{B}(\mathbb{R})$ as the smallest σ-field containing

a) all intervals $(a, b],\ a, b \in \mathbb{R}$

b) all intervals $(a, b),\ a, b \in \mathbb{R}$

c) all intervals $[a, b),\ a, b \in \mathbb{R}$

d) all intervals $[a, b],\ a, b \in \mathbb{R}$.

e) all intervals $(a, \infty),\ a \in \mathbb{R}$.

f) all intervals $[a, \infty),\ a \in \mathbb{R}$.

g) all intervals $(-\infty, b),\ b \in \mathbb{R}$.

h) all intervals $(-\infty, b],\ b \in \mathbb{R}$.

i) all open sets of $\mathbb{R}$.[7]

j) all closed sets of $\mathbb{R}$.[8]

To illustrate these equivalences, let us equate the first two conditions. That is, let us show that a σ-field contains all open intervals (a, b) iff it contains all right semi-closed intervals $(a, b]$. To see this, simply note

$$(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right) \tag{3.1.1a}$$

and

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right] \tag{3.1.1b}$$

Question 3.1.1. The text gives another description of the Borel sets $\mathcal{B}(\mathbb{R})$ as the smallest σ-field containing $\mathcal{F}_0$, the field of disjoint unions of right semi-closed intervals $(a, b]$. Can we make the same statement about the field of finite disjoint unions of left semi-closed intervals? △

The Borel sets are a large collection of sets. For instance, Remark 3.1.4 notes that the Cantor set is a Borel set.

Remark 3.1.4. (*The Cantor set is a Borel set*) The Cantor set must be a Borel set because it is closed. To see this more explicitly, note that in each step you "remove the middle third of each part".

$$K = \bigcap_{i=1}^{\infty} \bigcap_{j=1}^{3^{i-1}-1} \left[0, \frac{3j+1}{3^i}\right] \cup \left[\frac{3j+2}{3^i}, 1\right]$$

which is a countable number of intersections and unions of closed intervals, and hence Borel by characterization (d) above. △

[7] Recall that an open set is a countable union of open intervals.
[8] Recall that a set is open iff its complement is closed.

3.2 § 1.2.3-1.2.4: Measures

Definition 3.2.1. A **measure** on a σ-field $\mathcal{F}$ is a non-negative, extended real-valued function μ on $\mathcal{F}$ such that whenever $A_1, A_2, ...$ form a finite or countably infinite collection of disjoint sets in $\mathcal{F}$, we have countable additivity; that is,

$$\mu\left(\dot{\bigcup_n} A_n\right) = \sum_n \mu(A_n)$$

$\triangle$

Definition 3.2.2. A **probability measure** is a measure (Definition 3.2.1) where $\mu(\Omega) = 1$. $\triangle$

Remark 3.2.1. Ash additionally assumes that a measure does not take $\mu(A) = \infty$ for all $A \in \mathcal{F}$.[9] From this, we automatically obtain $\mu(\emptyset) = 0$. For $\mu(A) < \infty$ for some A, and by considering the sequence $A, \emptyset, \emptyset, ...$, we have that $\mu(\emptyset) = 0$ by countable additivity. $\triangle$

Example 3.2.1. Let Ω be any set. Fix $x_0 \in \Omega$. Let $\mathcal{F} = 2^\Omega$. For any $A \in \mathcal{F}$ define $\mu(A) = 1$ if $x_0 \in A$ and $\mu(A) = 0$ if $x_0 \notin A$. Then μ may be called the **unit mass** concentrated at x_0. $\triangle$

Example 3.2.2. Let $\Omega = \{x_1, x_2, ...\}$ be a finite or countably infinite set. Let $p_1, p_2, ...$ be non-negative reals. Let $\mathcal{F} = 2^\Omega$. Define

$$\mu(A) = \sum_{x_i \in A} p_i \quad \text{for all } A \in \mathcal{F}$$

Then μ is a measure on $\mathcal{F}$. We might call it the “point weighting" measure.

- If $p_i \equiv 1 \; \forall\, i$, then μ is called the **counting measure**.
- If $\sum_i p_i = 1$, then μ is a probability measure.

$\triangle$

Example 3.2.3. (*Lebesgue measure*) Define μ such that

$$\mu(a, b] = b - a \quad \forall\, a, b \in \mathbb{R} : b > a$$

As we will see in Section 4, this requirement determines μ on a large collection of sets, the Borel Sets $\mathcal{B}(\mathbb{R})$, which we defined in Section 3.1.4 as the smallest σ-field of subsets of $\mathbb{R}$ containing all intervals $(a, b] \subset \mathbb{R}$. $\triangle$

3.3 § 1.2.5-1.2.6: Properties of measures (and some more general set functions)

The text considers some generalizations of measures that can be obtained

1. by restricting the domain to a field (in other texts, such functions are called *pre-measures*)
2. by only assuming *finite* additivity
3. by allowing the range to be extended reals ($\bar{\mathbb{R}}$) instead of non-negative extended reals ($\bar{\mathbb{R}}_{\geq 0}$).

Remark 3.3.1. With respect to pre-measures, a countably additive function can be defined on a *field* (rather than σ-field) if the condition is taken to hold whenever a countable union *does* happen to still be in the field. Unless otherwise specified, I will assume in these notes by that countably additive functions are always defined on σ-fields, and finitely additive functions are defined on fields. $\triangle$

	Range	
	non-negative extended reals	**extended reals**
countably additive	μ measure	$\tilde{\mu}$ signed measure
finitely additive	μ_0	$\tilde{\mu}_0$

Table 1: Notation for generalizations of measure (For assumed domain in each case, see Remark 3.3.1.)

In Table 1, we introduce some notation to try to clarify more immediately when results hold. Note the relations[10]

$$\{\mu\} \subset \{\mu_0\}, \{\tilde{\mu}\} \subset \{\tilde{\mu}_0\}.$$

Remark 3.3.2. Being able to work with these generalizations will be important in Section 4 on extension of measures. In particular, it will help us show that we can construct the Lebesgue measure on the Borel sets. △

Example 3.3.1. Let $\mathcal{F}_0$ be the field of finite disjoint unions of right semi-closed intervals (see Definition C.1.1), and define the set function $\tilde{\mu}_0$ on $\mathcal{F}_0$ as follows[11]:

$$\begin{aligned}
\tilde{\mu}_0(-\infty, a] &= a, && a \in \mathbb{R} \\
\tilde{\mu}_0(a, b] &= b - a, && a, b \in \mathbb{R}, \quad a < b \\
\tilde{\mu}_0(b, \infty) &= -b, && b \in \mathbb{R} \\
\tilde{\mu}_0(\mathbb{R}) &= 0 \\
\tilde{\mu}_0\Big(\dot{\bigcup_{i=1}^{n}} I_i\Big) &= \sum_{i=1}^{n} \tilde{\mu}_0(I_i), && \text{if } I_1, ..., I_n \text{ are right semi-closed intervals}
\end{aligned}$$

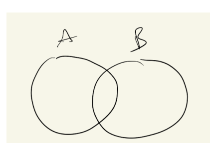
Then $\tilde{\mu}_0$ is finitely additive, but not countably additive on $\mathcal{F}_0$. (Why?) For a proof, see GoodNotes. △

Measure-like set functions have useful properties. Using the notation in Table 1, we rewrite Theorem 1.2.5 of the text:

Theorem 3.3.1. *Let $\tilde{\mu}_0$ be a finitely additive set function on the field $\mathcal{F}_0$. Then*

a) $\tilde{\mu}_0(\emptyset) = 0$

b) $\tilde{\mu}_0(A \cup B) + \tilde{\mu}_0(A \cap B) = \tilde{\mu}_0(A) + \tilde{\mu}_0(B)$ *for all* $A, B \in \mathcal{F}_0$.



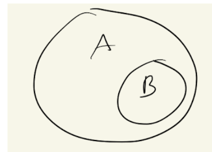
c) If $A, B \in \mathcal{F}_0$ *and* $B \subset A$, *then*

$$\tilde{\mu}_0(A) = \tilde{\mu}_0(B) + \tilde{\mu}_0(A - B) \quad \textit{(piece-and-difference decomposition)}$$

[9] Likewise, he assumes that signed measures do not take $\mu(A) = -\infty$ for for all $A \in \mathcal{F}$.

[10] So, for example, if something holds for $\tilde{\mu}_0$, it holds for μ. A simple mnemonic is that adding stuff to the notation generalizes the function.

[11] This example comes from Problem 4 in Section 1.2 of the text



[12]*So* $\tilde{\mu}_0(A) \geq \tilde{\mu}_0(B)$ *if* $\tilde{\mu}_0(A - B) \geq 0$. *More generally, for non-negative set functions, we have*

$$\mu_0(A) \geq \mu_0(B) \quad \textit{(monotonicity)}$$

d) *Subadditivity holds if* $\tilde{\mu}_0$ *is non-negative, i.e.*

$$\mu_0(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mu_0(A_i)$$

$$\mu(\cup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu(A_i)$$

Proof. We prove Theorem 3.3.1 (b). The rest is an exercise for the reader (or see the text).

First, we break things into disjoint pieces

$$A = \Big(A \cap B\Big) \biguplus \Big(A \cap B^c\Big) \qquad \implies \tilde{\mu}_0(A) = \tilde{\mu}_0(A \cap B) + \tilde{\mu}_0(A \cap B^c) \tag{1}$$

$$B = \Big(A \cap B\Big) \biguplus \Big(A^c \cap B\Big) \qquad \implies \tilde{\mu}_0(B) = \tilde{\mu}_0(A \cap B) + \tilde{\mu}_0(A^c \cap B) \tag{2}$$

$$A \cup B = \Big(A \cap B\Big) \biguplus \Big(A \cap B^c\Big) \biguplus \Big(A^c \cap B\Big) \qquad \implies \tilde{\mu}_0(A \cup B) = \tilde{\mu}_0(A \cap B) + \tilde{\mu}_0(A \cap B^c) + \tilde{\mu}_0(A^c \cap B) \tag{3}$$

Summing (1) and (2), we obtain

$$\tilde{\mu}_0(A) + \tilde{\mu}_0(B) = 2\tilde{\mu}_0(A \cap B) + \tilde{\mu}_0(A \cap B^c) + \tilde{\mu}_0(A^c \cap B).$$

We use (3) to simplify the RHS, and the result follows. □

Remark 3.3.3. In the proof of Theorem 3.3.1 (b), note that we use a common strategy – breaking sets into disjoint pieces so that we can apply the assumed (finite or countable) additivity of the set function. △

Remark 3.3.4. Is *finiteness* ($|\mu_g(A)| < \infty \ \forall \ A \in \mathcal{F}_g$) equivalent to *boundedness* ($\sup\{|\mu_g(A)| : A \in \mathcal{F}_g\} < \infty$)?

- $\mu_0, \widetilde{\mu}$? ✓
- $\tilde{\mu}_0$? ✗ (too general)

The fact that equivalence holds for signed measures $\widetilde{\mu}$ is surprising. Somehow countable additivity compensates for the signedness. See Section 2.1.3 of the text. △

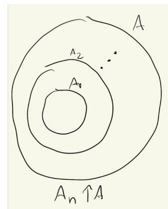
[12]If the "piece" satisfies $\tilde{\mu}_0(B) < \infty$, we have $\tilde{\mu}_0(A - B) = \tilde{\mu}_0(A) - \tilde{\mu}_0(B)$. One useful takeaway for piece-and-difference decompositions is that : *the finite measure of the difference is the difference of the finite measures*.

3.4 § 1.2.7-1.2.8: Continuity of countably additive set functions

Countably additive set functions have a basic continuity property. Continuity of measure is a special case.

Theorem 3.4.1. *Let $\widetilde{\mu}$ be a countably additive set function on the σ-field $\mathcal{F}$. Then*

a) (continuity from below) If $A_1, A_2, ... \in \mathcal{F}$ and $A_n \uparrow A$, then $\widetilde{\mu}(A_n) \to \widetilde{\mu}(A)$ as $n \to \infty$.



b) (continuity from above) If $A_1, A_2, ... \in \mathcal{F}$, $A_n \downarrow A$, and $\widetilde{\mu}(A_1)$ is finite, then $\widetilde{\mu}(A_n) \to \widetilde{\mu}(A)$ as $n \to \infty$.

Proof. We prove continuity from below, and leave continuity from above as an exercise to the reader (or see text).

First let us assume that all $\widetilde{\mu}(A_n)$ are finite (*). Then

$$\begin{aligned} A &= A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup ... && \text{by (2.2.2)} \\ \implies \widetilde{\mu}(A) &= \widetilde{\mu}(A_1) + \widetilde{\mu}(A_2 - A_1) + \widetilde{\mu}(A_3 - A_2) + ... && \text{(countable additivity)} \\ &= \widetilde{\mu}(A_1) + \widetilde{\mu}(A_2) - \widetilde{\mu}(A_1) + \widetilde{\mu}(A_3) - \widetilde{\mu}(A_2) + ... && \text{(Theorem 3.3.1 c), (*)} \\ &= \lim_{n\to\infty} \widetilde{\mu}(A_n) && \text{(telescoping difference)} \end{aligned}$$

Now suppose $\widetilde{\mu}(A_n) = \infty$ for some n. So write

$$\begin{aligned} A &= A_n \cup A - A_n && \text{(increasing sequence)} \\ \implies \widetilde{\mu}(A) &= \widetilde{\mu}(A_n) + \widetilde{\mu}(A - A_n) && \text{(countable additivity)} \\ &= \infty + \widetilde{\mu}(A - A_n) \end{aligned}$$

So $\widetilde{\mu}(A) = \infty$.[13] Replace A by A_k for any $k \geq n$ to also find $\widetilde{\mu}(A_k) = \infty$ for all $k \geq n$ and the result follows.

Finally suppose $\widetilde{\mu}(A_n) = -\infty$ for some n. Then the result follows in the same way as for $\widetilde{\mu}(A_n) = \infty$.

□

Remark 3.4.1. The logic of the proof of Theorem 3.4.1 under the finiteness assumption is as follows. First, we re-represent the union as a disjoint union (the form is particularly simple since the sets are increasing). This allows us to apply countable additivity. Then we apply the piece-and-difference decomposition (and the subtraction is defined under the finiteness assumption). △

Remark 3.4.2. In proving Theorem 3.4.1 for the case where $\mu(A_n) = \infty$ for some n, it is tempting to make the simpler argument

$$\begin{aligned} \mu(A) &\geq \mu(A_n) && \text{(monotonicity)} \\ \mu(A_k) &\geq \mu(A_n) && \text{(monotonicity)} \end{aligned}$$

[13] Note that we cannot have $\widetilde{\mu}(A - A_n) = -\infty$, because that would violate additivity.

for $k \geq n$. But recall from Theorem 3.3.1 that monotonicity only holds under non-negativity, and the theorem statement is more general, applying to *signed* set functions as well. $\triangle$

Remark 3.4.3. Theorem 3.4.1 still holds if $\mathcal{F}$ is only assumed to be a field, so long as the limit sets A belong to $\mathcal{F}$. $\triangle$

We have the result that finite additivity plus continuity equals countable additivity.

Theorem 3.4.2. *Let $\tilde{\mu}_0$ be a finitely additive set function on the field $\mathcal{F}_0$. Suppose either*

a) $\tilde{\mu}_0$ is continuous from below

b) $\tilde{\mu}_0$ is continuous from above at the empty set.

Then $\tilde{\mu}_0$ is countably additive.

Proof. We prove that the conclusion holds under (a) and leave doing the same for (b) as an exercise to the reader (or see text).

Given $A = \dot{\bigcup}_{n=1}^{\infty} A_n$, we define $P_n := \bigcup_{m \leq n} A_n$ and so $P_n \uparrow A$. So we have

$$\begin{aligned}
&\tilde{\mu}_0(P_n) \to \tilde{\mu}_0(A) && \text{(continuity from below)} \\
\implies\ &\tilde{\mu}_0\Big(\bigcup_{m \leq n} A_n\Big) \to \tilde{\mu}_0(A) && \text{(definition)} \\
\implies\ &\sum_{m=1}^{n} \tilde{\mu}_0(A_n) \to \tilde{\mu}_0(A) && \text{(finite additivity)}
\end{aligned}$$

Taking $n \to \infty$ gives countable additivity. $\square$

3.5 Borel-Cantelli Lemma

Lemma 3.5.1. ***Borel Cantelli Lemma*** *If $A_1, A_2, \ldots \in \mathcal{F}$ and $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, then $\mu(\limsup_{n\to\infty} A_n) = 0$.*

Proof. [14] Recall from Definition 2.1.1 that

$$\underbrace{\limsup_{n\to\infty} A_n}_{:= B} = \bigcap_{k=1}^{\infty} \underbrace{\bigcup_{n=k}^{\infty} A_n}_{:= B_k} = \{x : x \in A_n \text{ i.o. }\}$$

where we have also introduced some notation for convenience.

Now $B_{k+1} \subset B_k$ and $\cap_{k=1}^{\infty} B_k = B$, so $B_k \downarrow B$. Since also $\mu(B_1) < \infty$ (by hypothesis and monotonicity), then we can apply continuity from above (Theorem 3.4.1) to get

$$\begin{aligned}
\mu(B) &= \lim_{k\to\infty} \mu(B_k) \\
&= \lim_{k\to\infty} \mu\Big(\bigcup_{n=k}^{\infty} A_n\Big) && \text{def. } B_k \\
&\leq \lim_{k\to\infty} \sum_{n=k}^{\infty} \mu(A_n) && \text{subadditivity} \\
&= 0 && \text{convergent series have vanishing tails}
\end{aligned}$$

[14] [Ash et al., 2000] has a proof that does not require continuity of measure, although I currently personally enjoy its role here.

□

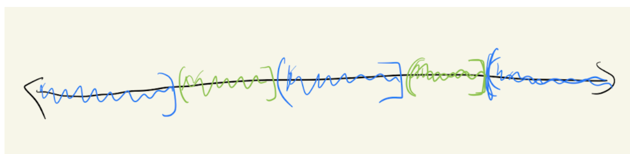
Remark 3.5.1. (*Intuition for Borel-Cantelli*) An attempt at a verbal description of the proof of Lemma 3.5.1 follows: Convergent series of real numbers have arbitrarily small tails. When the series is constructed of measures ($\sum_{n=1}^{\infty} \mu(A_n)$), this implies that the measure of the tail ($\cup_{n=k}^{\infty} A_n$) is arbitrarily small (by subadditivity). Now $\limsup_{n\to\infty} A_n$ is the set of points in $\{A_n\}$ i.o., so such points must be in *all* tails ($\cap_{k=1}^{\infty} \cup_{n=k}^{\infty} A_n$), and by continuity of measure (from above), the measure of such a set is the limit of the measure of the tail, i.e. 0. △

4 § 1.3: Extension of measures

4.1 Extension and approximation

In Example 3.2.3, we discussed the concept of length of a subset of $\mathbb{R}$; in particular, we mentioned extending the set function given on intervals by $\mu(a, b] = b - a$ to a larger class of subsets of $\mathbb{R}$.

As remarked in Example 3.1.4, if we define $\mathcal{F}_0 = \{$finite disjoint unions of right semi-closed intervals $(a, b], a < b\}$, then $\mathcal{F}_0$ is a field, as can be easily verified. And μ can easily be seen to be a finitely additive set function on $\mathcal{F}_0$.



However, $\mathcal{F}_0$ is not a σ-field. So how can we extend this function to a measure on a larger class of subsets? For instance, we would at least like to be able to measure intervals such as (a, b), $[a, b)$ or $[a, b]$ and points $\{x\}$. The challenges are:

- *We need to show that μ is countably additive.* We will do this in Section 5. Moreover, in that section, we will generalize our problem to set functions given by $\mu(a, b] = F(b) - F(a)$, where F is an increasing right-continuous function from $\mathbb{R}$ to $\mathbb{R}$.
- *We need to extend μ to $\sigma(\mathcal{F}_0)$, the minimal σ-field containing $\mathcal{F}_0$.* In other words, we need to extend μ to the Borel sets. We will handle the problem in this section more generally. In this section, we will deal with the problem of extending a measure on $\mathcal{F}_0$ to a measure on $\sigma(\mathcal{F}_0)$. We do so using Carathéodory's Theorem (Theorem 4.1.3). Along the way, we will use Theorem 4.1.1 and Theorem 4.1.2 to prove Theorem 4.1.3.

Theorem 4.1.1. *(Theorem 1.3.6 [Ash et al., 2000]) A finite measure on a field $\mathcal{F}_0$ can be extended to a measure on $\sigma(\mathcal{F}_0)$.*

Proof. See pp. 12-17 of [Ash et al., 2000]. □

Theorem 4.1.2. ***(Monotone Class Theorem)*** *Let $\mathcal{F}_0$ be a field of subsets of Ω and $\mathcal{C}$ be a class of subsets of Ω that is monotone (if $A_n \in \mathcal{C}$ and $A_n \uparrow A$ or $A_n \downarrow A$, then $A \in \mathcal{C}$). If $\mathcal{C} \supset \mathcal{F}_0$ then $\mathcal{C} \supset \sigma(\mathcal{F}_0)$, then minimal σ-field over $\mathcal{F}_0$.*

Proof. See pp. 18-19 of [Ash et al., 2000]. □

Remark 4.1.1. During the proof of Theorem 4.1.2, some key observations are made about the relationship between monotone classes and σ-fields:

a) A monotone class that is also field is a sigma-field. (See Remark 3.1.2.)

b) The smallest monotone class and smallest sigma-field over a field coincide.

$\triangle$

Remark 4.1.2. (*The utility of the Monotone Class Theorem*) The Monotone Class Theorem provides an alternate route towards executing on the Good Sets Strategy (Section 3.1.3.) Suppose you want to show that all members of a σ-algebra $\mathcal{F}$ have some property P. Define "good sets" as those that satisfy the property

$$\mathcal{G} := \{G \in \mathcal{F} : G \text{ has property } P\}$$

The strategy is then to simply

1. Show $\mathcal{G}$ contains some class $\mathcal{C}$ such that $\mathcal{F} = \sigma(\mathcal{C})$
2. Show $\mathcal{G}$ is a monotone class.

$\triangle$

Remark 4.1.3. The strategy in Remark 4.1.2 is very much like induction. Step #1 is the "base" step and step #2 is the "induction" step. $\triangle$

For an example of where the strategy in Remark 4.1.2 is used, see the proof of uniqueness in the Caratheodory Extension Theorem (Theorem 4.1.3). It is also used extensively to show that Borel sets have some property; see Section 5.7.

Theorem 4.1.3. ***(Carathéodory Extension Theorem)*** *Let μ be a measure on the field $\mathcal{F}_0$ of subsets of Ω, and assume that μ is σ-finite on $\mathcal{F}_0$, so that Ω can be decomposed as $\cup_{n=1}^{\infty} A_n$ where $A_n \in \mathcal{F}_0$ and $\mu(A_n) < \infty$ for all n. Then μ has a unique extension to a measure on $\mathcal{F} := \sigma(\mathcal{F}_0)$, the minimal σ-field over $\mathcal{F}_0$.*

Proof. (We follow the argument of [Ash et al., 2000], but add some detail.) First we prove existence. [Without loss of generality, we assume the A_n are disjoint. This is possible because we can use (2.2.1) to re-express the countable union as a disjoint countable union: $\Omega = \cup_{i=1}^{\infty} A_i = \uplus_{i=1}^{\infty} B_i$, where $B_i := A_i \cap A_{i-1}^c \dots \cap A_1^c$.]

If we define $\mu_n(A) = \mu(A \cap A_n)$ for each $A \in \mathcal{F}_0$, then we can decompose μ into a countable sum of finite measures:

- μ_n is a measure on $\mathcal{F}_0$. [Its countable additivity is inherited from μ. If $\uplus_{i=1}^{\infty} A_i$ is a disjoint union, then so is $\uplus_{i=1}^{\infty} (A_i \cap A_n)$, and $\mu(\uplus_{i=1}^{\infty} (A_i \cap A_n)) = \sum_{i=1}^{\infty} \mu(A_i \cap A_n)$ since $A_i \cap A_n$ are in $\mathcal{F}_0$.]
- μ_n is finite. [True because $\mu_n(A) = \mu(A \cap A_n) \overset{\text{monotonicity}}{\leq} \mu(A_n) < \infty$.]
- $\mu = \sum_{n=1}^{\infty} \mu_n$. [True because $\mu(A) = \mu(A \cap \Omega) = \mu(A \cap (\uplus_{n=1}^{\infty} A_n)) = \mu(\uplus_{n=1}^{\infty} (A \cap A_n)) = \sum_{n=1}^{\infty} \mu(A \cap A_n) = \mu_n(A)$.]

Now by Theorem 4.1.1, we can extend each μ_n to a measure μ_n^* on $\mathcal{F}$. Thus $\mu^* := \sum_{n=1}^{\infty} \mu_n^*$ extends μ to $\mathcal{F}$. Moreover, μ^* is still a measure since the order of summation in a double series of nonnegative terms can be reversed. [Countable additivity still holds since:

$$\begin{aligned}
\mu^*(\uplus_{i=1}^{\infty} A_i) &= \sum_{n=1}^{\infty} \mu_n^*(\uplus_{i=1}^{\infty} A_i) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \mu_n^*(A_i) && \mu_n^* \text{ is measure, so countably additive} \\
&= \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} \mu_n^*(A_i) && \text{reverse order of summation for double series with non-negative terms} \\
&= \sum_{i=1}^{\infty} \mu^*(A_i) && \text{def. of } \mu^*
\end{aligned}$$

].

Now we prove uniqueness. That is, we prove that if λ is a measure on $\mathcal{F}$ and $\lambda = \mu^*$ on $\mathcal{F}_0$, then $\lambda = \mu^*$ on $\mathcal{F}$. To see this, as before, we decompose the measure into a sum of finite measures: $\lambda = \sum_{n=1}^{\infty} \lambda_n$ where $\lambda_n := \lambda(A_n \cap A)$. Now by assumption $\lambda_n = \mu_n^*$ on $\mathcal{F}_0$. Where are they equal on $\mathcal{F}$? Let us define the "good sets" (recall Section 3.1.3)

$$\mathcal{G} := \{A \in \mathcal{F} : \lambda_n(A) = \mu_n^*(A)\}$$

Now we can show $\mathcal{G} = \mathcal{F}$ – that is, *all* sets in the σ-field are good sets – by observing

- $\mathcal{G}$ is a monotone class. [This is true by continuity from below (see Theorem 3.4.1). In particular, a countable union can be considered the limit of an increasing sequence of partial unions (See Remark 3.1.2.) As a result, the measure of the limiting set is determined, as the limit of the the measure of the sets in that sequence.]
- $\mathcal{G} \supset \mathcal{F}_0$. [This is true by construction.]

And so by Monotone Class Theorem (Theorem 4.1.2), we have $\mathcal{G} \supset \mathcal{F}$. But by construction $\mathcal{G} \subset \mathcal{F}$, and so $\mathcal{G} = \mathcal{F}$. Therefore $\lambda_n = \mu_n^*$ for each n.

So

$$\lambda \overset{\text{decomposition}}{=} \sum_n \lambda_n = \sum_n \mu_n^* \overset{\text{recomposition}}{=} \mu^*,$$

proving uniqueness. □

Remark 4.1.4. The proof of Theorem 4.1.3 reveals the appeal of σ-finite measures – they can be decomposed as the countable sum of finite measures (and the order of summation of double series can be reversed for nonnegative series, so countable additivity still holds). △

In Remark 4.1.1 (b), we observed that minimal σ-fields over a field can be characterized as the minimal monotone classes over a field – so we merely need to close the field over increasing and decreasing sequences of sets. This idea suggests that if $\mathcal{F}_0$ is a field and $\mathcal{F} = \sigma(\mathcal{F}_0)$, sets in $\mathcal{F}$ can be approximated in some sense by sets in $\mathcal{F}_0$. The following result formalizes this notion.

Theorem 4.1.4. ***(Approximation Theorem)*** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $\mathcal{F} = \sigma(\mathcal{F}_0)$ where $\mathcal{F}_0$ is a field of subsets of Ω. Let μ be σ-finite on $\mathcal{F}_0$. Then for every $A \in \mathcal{F}$ and fixed $\epsilon > 0$, there is a set $B \in \mathcal{F}_0$ such that $\mu(A \triangle B) < \epsilon$.*

Example 4.1.1. This interesting example (from [Ash et al., 2000] pp. 20) provides a counterexample to the theorems when $\mathcal{F}_0$ is not σ-finite.

△

1.3.12 Example. Let Ω be the rationals, $\mathscr{F}_0$ the field of finite disjoint unions of right-semiclosed intervals $(a, b] = \{\omega \in \Omega\colon\ a < \omega \le b\}$, a, b rational [counting (a, ∞) and Ω itself as right-semiclosed; see 1.2.2]. Let $\mathscr{F} = \sigma(\mathscr{F}_0)$. Then:

(a) $\mathscr{F}$ consists of all subsets of Ω.

(b) If $\mu(A)$ is the number of points in A (μ is counting measure), then μ is σ-finite on $\mathscr{F}$ but not on $\mathscr{F}_0$.

(c) There are sets $A \in \mathscr{F}$ of finite measure that cannot be approximated by sets in $\mathscr{F}_0$, that is, there is no sequence $A_n \in \mathscr{F}_0$ with $\mu(A \,\Delta\, A_n) \to 0$.

(d) If $\lambda = 2\mu$, then $\lambda = \mu$ on $\mathscr{F}_0$ but not on $\mathscr{F}$.

Thus both the approximation theorem and the Carathéodory extension theorem fail in this case.

1.3 EXTENSION OF MEASURES **21**

PROOF. (a) We have $\{x\} = \bigcap_{n=1}^{\infty} (x - (1/n), x]$, and therefore all singletons are in $\mathscr{F}$. But then all sets are in $\mathscr{F}$ since Ω is countable.

(b) Since Ω is a countable union of singletons, μ is σ-finite on $\mathscr{F}$. But every nonempty set in $\mathscr{F}_0$ has infinite measure, so μ is not σ-finite on $\mathscr{F}_0$.

(c) If A is any finite nonempty subset of Ω, then $\mu(A \,\Delta\, B) = \infty$ for all nonempty $B \in \mathscr{F}_0$, because any nonempty set in $\mathscr{F}_0$ must contain infinitely many points not in A.

(d) Since $\lambda\{x\} = 2$ and $\mu\{x\} = 1, \lambda \ne \mu$ on $\mathscr{F}$. But $\lambda(A) = \mu(A) = \infty, A \in \mathscr{F}_0$ (except for $A = \emptyset$). $\square$

4.2 Completion of measure spaces

Definition 4.2.1. A measure μ on a σ-field $\mathcal{F}$ is said to be *complete* iff whenever $A \in F$ and $\mu(A) = 0$, we have $B \in F$ for all $B \subset A$. $\triangle$

Definition 4.2.2. The *completion* of a measure space $(\Omega, \mathcal{F}, \mu)$ is given by $(\Omega, \mathcal{F}_\mu, \mu)$, where

$$\mathcal{F}_\mu := \{A \cup S : A \in \mathcal{F}, S \subset N \text{ for some } N \in \mathcal{F} \text{ with } \mu(N) = 0\}$$

and where μ is extended to $\mathcal{F}_\mu$ by setting $\mu(A \cup S) = \mu(A)$. $\triangle$

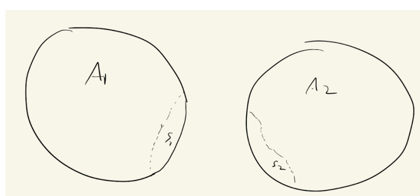
Remark 4.2.1. Let us show that Definition 4.2.2 is a valid definition by showing that

1. *$\mathcal{F}_\mu$ is a σ-field.*
2. *μ is a measure on $\mathcal{F}_\mu$.*
3. *The completion is complete.*

We justify these in turn:

1. $\mathcal{F}_\mu$ is closed under countable unions, since
$$\cup_{i=1}^\infty (A_i \cup S_i) = \underbrace{(\cup_{i=1}^\infty A_i)}_{\in \mathcal{F}} \cup \underbrace{(\cup_{i=1}^\infty S_i)}_{\text{has measure 0}}$$
where the term on the right has measure 0 because $\cup_{i=1}^\infty S_i \subset \cup_{i=1}^\infty N_i \in \mathcal{F}$, and $\mu(\cup_{i=1}^\infty N_i) = \sum_{i=1}^\infty \mu(N_i) = 0$.
F_μ is also closed under complements, since $S \subset N \implies N^c \subset S^c$, and so
$$(A \cup S)^c = (A^c \cap S^c) = \underbrace{(A^c \cap N^c)}_{\in \mathcal{F}} \cup \underbrace{(A^c \cap S^c - N^c)}_{\text{has measure 0}}$$
where the term on the right has measure 0 by monotonicity, because $A^c \cap S^c - N^c \subset S^c - N^c = S^c \cap (M^c)^c = S^c \cap N \subset N$.

2. First, we show that countable additivity holds in $\mathcal{F}_\mu$.
$$\mu(\uplus_{i=1}^\infty (A_i \cup S_i)) \overset{\text{see below}}{=} \mu(\uplus_{i=1}^\infty A_i) \overset{\mu \text{ countably additive on } \mathcal{F}}{=} \sum_{i=1}^\infty \mu(A_i) \overset{\text{construction of extension}}{=} \sum_{i=1}^\infty \mu(A_i \cup S_i)$$
The first equality holds because we can re-represent a disjoint union $\uplus_{i=1}^\infty (A_i \cup S_i) = (\uplus_{i=1}^\infty A_i) \cup (\uplus_{i=1}^\infty S_i)$. Since $\uplus_{i=1}^\infty S_i \subset \underbrace{\uplus_{i=1}^\infty N_i}_{\text{has measure 0 in } \mathcal{F}}$, we have that $\mu((\uplus_{i=1}^\infty A_i) \cup (\uplus_{i=1}^\infty S_i)) = \mu(\uplus_{i=1}^\infty A_i)$.
Next, we show that μ is invariant to decompositions: if $A_1 \cup S_1 = A_2 \cup S_2$, then $\mu(A_1 \cup S_1) = \mu(A_2 \cup S_2)$, or more simply $\mu(A_1) = \mu(A_2)$.



We have
$$\mu(A_1) \overset{\text{countable additivity}}{=} \mu(A_1 \cap A_2) + \mu(A_1 \cap A_2^c) \overset{\text{see below}}{=} \mu(A_1 \cap A_2) \overset{\text{monotonicity}}{\leq} \mu(A_2)$$
where the second equality holds since $A_1 \cap A_2^c \subset S_2$ (which, in turn, holds since $x \in A_1 \implies x \in A_2$ or $x \in S_2$, so $x \in A_1$ and $x \notin A_2 \implies x \in S_2$) .
By symmetry, $\mu(A_2) \leq \mu(A_1)$, so $\mu(A_1) = \mu(A_2)$.

3. By the definition of a complete measure, we need to show that if $B \in \mathcal{F}_\mu$ and $\mu(B) = 0$ then $C \in \mathcal{F}_\mu$ for all $C \subset B$.
Now $B \in \mathcal{F}_\mu \implies B = \underbrace{A}_{\in \mathcal{F}} \cup \underbrace{S}_{\subset N \in \mathcal{F} : \mu(N) = 0}$.
So our assumption $\mu(B) = 0$ gives us $\mu(A) = 0$, since $\mu(B) = \mu(A \cup S) \overset{\text{choice of extension}}{=} \mu(A) = 0$.
Now since we have assumed $C \subset B$ we have
$$\mu(C) \overset{\text{monotonicity}}{\leq} \mu(B) \overset{B \in \mathcal{F}_\mu}{=} \mu(A \cup S) \overset{\text{subadditivity}}{\leq} \mu(A) + \mu(S) \overset{\text{see above}}{=} 0 + \mu(S) = 0 + 0 = 0$$
Since μ is non-negative, this implies that $\mu(C) = 0$.
We can therefore write $C = \underbrace{\emptyset}_{\in \mathcal{F}} \cup \underbrace{C}_{\text{has measure 0}}$, so $C \in \mathcal{F}_\mu$.
Thus, μ on $\mathcal{F}_\mu$ is complete, since any subset of measure 0 is contained in $\mathcal{F}_\mu$.

△

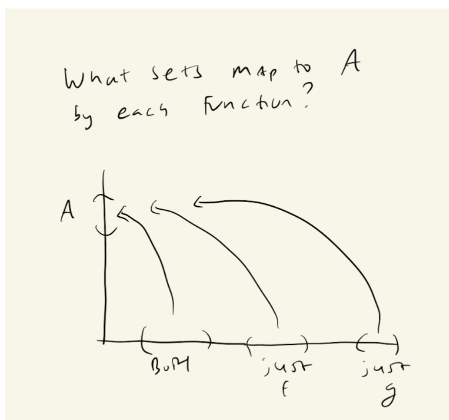
Example 4.2.1. Let us provide a simple example of where completing a measure space generates new measurable sets. Consider a measure space given by $(\Omega, \mathcal{F}, \mu)$ where $\Omega = \mathbb{R}$, $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ where we take $A = [0, 1]$ for concreteness, and where μ is defined by $\mu(A^c) = 1$ and $\mu(A) = 0$. The measure space is not complete, since no proper subset of A is contained in $\mathcal{F}$. (For example, $[0, \frac{1}{2}] \notin \mathcal{F}$.) If we complete the measure space, we obtain $(\Omega, \mathcal{F}_\mu, \mu)$ where $\mathcal{F}_\mu = \{\emptyset, A, A^c, \Omega, \text{ any subset of } [0, 1]\}$. △

Problem 4.2.1. Let $(\Omega, \mathcal{F}, \mu)$ be a complete measure space. If $f : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ and $g : \Omega \to \Omega'$, $g = f$ except on a subset of a set $A \in \mathcal{F}$ with $\mu(A) = 0$, show that g is measurable (relative to $\mathcal{F}$ and $\mathcal{F}'$)

Solution. For all $A \in \mathcal{F}'$,

$$\begin{aligned} g^{-1}(A) &= \Big\{x : x \in g^{-1}(A) \text{ and } x \in f^{-1}(A)\Big\} \qquad \bigcup \Big\{x : x \in g^{-1}(A) \text{ and } x \notin f^{-1}(A)\Big\} \\ &= \underbrace{\{x \in f^{-1}(A)\}}_{\in \mathcal{F} \text{ since } f \text{ measurable}} \setminus \underbrace{\Big\{x : x \in f^{-1}(A) \text{ and } x \notin g^{-1}(A)\Big\}}_{\in \mathcal{F} \text{ by completeness, as a subset of a set of measure 0}} \bigcup \underbrace{\Big\{x : x \in g^{-1}(A) \text{ and } x \notin f^{-1}(A)\Big\}}_{\in \mathcal{F} \text{ by completeness, as a subset of a set of measure 0}} \end{aligned}$$

So by closure properties of σ-fields, $g^{-1}(A) \in \mathcal{F}$.



□

5 § 1.4: Lebesgue-Stieltjes Measures and Distribution Functions

Definition 5.0.1. A *Lebesgue-Stieltjes measure* on $\mathbb{R}$ is a measure μ on $\mathcal{B}(\mathbb{R})$ such that $\mu(I) < \infty$ for each bounded interval I. △

Definition 5.0.2. A *distribution function* on $\mathbb{R}$ is a map $F : \mathbb{R} \to \mathbb{R}$ that is increasing [$a < b$ implies $F(a) \leq F(b)$] and right continuous [$\lim_{x \downarrow x_0} F(x) = F(x_0)$]. △

In this Section, we show that the formula $\mu(a, b] = F(b) - F(a)$ sets up a one-to-one correspondence between distribution functions and Lebesgue-Stieltjes measures.

5.1 § 1.4.2 Each Lebesgue-Stietljes measure uniquely determines a distribution function (up to an additive constant)

First, the easy part: we show that to every Lebesgue-Stieltjes measure, there is a unique distribution function (up to an additive constant).

Theorem 5.1.1. *Let μ be a Lebesgue-Stietljes measure on $\mathbb{R}$. Let $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ be defined (up to additive constant) by $F(b) - F(a) = \mu(a, b]$ for $a < b$. Then F is a distribution function.*

Proof. We must show that F is increasing and right continuous.

1. We have $F(b) - F(a) = \mu(a, b] \geq 0$, since μ is non-negative. So F is increasing.
2. By the continuity (from above) of measure (which can be applied since since Lebesgue-Stietljes measures are finite on any interval),
$$\lim_{b' \downarrow b} [F(b') - F(a)] = \lim_{b' \downarrow b} \mu(a, b'] = \mu(a, b]$$
Thus, rearranging,
$$\lim_{b' \downarrow b} F(b') = \mu(a, b] + F(a) = \Big(F(b) - F(a)\Big) + F(a) = F(b)$$
So F is right continuous.

□

5.2 § 1.4.3-1.4.4 Each distribution function (identified up to additive constant) uniquely determines a Lebesgue-Stietljes measure

Now the harder part. We need to show that every distribution function F (identified up to additive constant) uniquely determines a Lebesgue-Stieltjes measure.

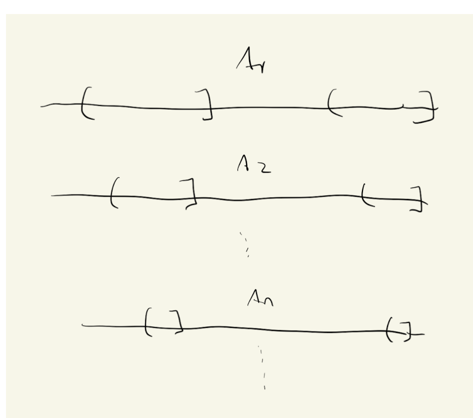
We will temporarily work with $\overline{\mathbb{R}}$, because it is a compact space, and then convert back to $\mathbb{R}$. In $\overline{\mathbb{R}}$, by a similar reasoning as we've seen before (e.g. see Section 4.1), it is straightforward to show that the formula $\mu(a, b] = F(a) - F(b), a, b \in \overline{\mathbb{R}}, a < b$ defines a finitely additive set function on $\mathcal{F}_0(\overline{\mathbb{R}})$, the field of disjoint unions of right semi-closed intervals of the extended reals.

The challenge will be to show that this set function is countably additive. If we can do that, then we can apply Carathéodory's Extension Theorem to extend the corresponding function μ on $\mathcal{F}_0(\mathbb{R})$ to $\mathcal{B}(\mathbb{R})$, as will be done in Theorem 5.2.1.

Lemma 5.2.1. *The set function μ is countably additive on $\mathcal{F}_0(\overline{\mathbb{R}})$.*

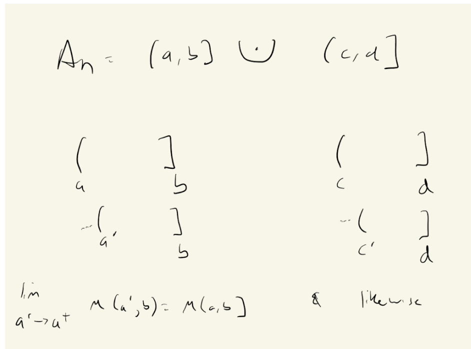
Proof. We assume $F(\infty) - F(-\infty) < \infty$, so that μ is finite. (We leave the case where $F(\infty) - F(-\infty) = \infty$ to the reader, or see the text.) Our strategy will be to show that μ is continuous from above, in which case we can apply Theorem 3.4.2 (b) to show that the set function is countably additive.

Let A_n be a sequence of sets in $\mathcal{F}_0(\overline{\mathbb{R}})$ such that $A_n \downarrow \emptyset$. Now each A_n is a finite union of disjoint r.s.c. intervals.



Suppose one such interval is $(a, b]$. By the right continuity of F, we can find intervals $(a', b]$ that approximate $(a, b]$ from the inside arbitrarily well, since by continuity from below

$$\mu(a', b] = F(b) - F(a') \to \mu(a, b] = F(b) - F(a) \text{ as } a' \downarrow a$$



Thus, we can find sets $B_n \in \mathcal{F}_0(\overline{\mathbb{R}})$ where $\mu(B_n)$ approximates $\mu(A_n)$ to any desired $\epsilon > 0$ that satisfy $B_n \subset \overline{B}_n \subset A_n$. By these inclusion properties and the decreasing nature of the sequence, we have:

a) $\cap_{n=1}^{\infty} \overline{B}_n = \emptyset$. [True because each $\overline{B}_n \subset A_n$, so $\cap_{n=1}^{\infty} \overline{B}_n \subset \cap_{n=1}^{\infty} A_n = \emptyset$.]

b) $\cap_{k=1}^{n} \overline{B}_k = \emptyset$ for sufficiently large n. [We have $\overline{\mathbb{R}} \overset{\text{item a)}}{=} (\overline{\mathbb{R}} - \cap_{n=1}^{\infty} \overline{B}_n) \overset{\text{DeMorgan (C.2.1)}}{=} \cup_{n=1}^{\infty} (\overline{\mathbb{R}} - \overline{B}_n)$. So $\{\overline{\mathbb{R}} - \overline{B}_n\}$ is an open cover of the compact space $\overline{\mathbb{R}}$. By the Heine-Borel theorem, there must be a finite subcover. So for sufficiently large n, we have $\cup_{k=1}^{n} (\overline{\mathbb{R}} - \overline{B}_k) = \overline{\mathbb{R}}$. Taking complements of both sides, and once again applying DeMorgan's law (C.2.1) to the relative complement, we find $\cap_{k=1}^{n} \overline{B}_k = \emptyset$.]

c) $\cap_{k=1}^{n} B_k = \emptyset$ for sufficiently large n. [This follow from item b) and the fact that each $B_k \subset \overline{B}_k$.]

So now we use a piece-and-difference decomposition (Theorem 3.3.1 (b)):

$$\begin{aligned}
A_n &= \left(\bigcap_{k=1}^{n} B_k \right) \dot{\bigcup} \left(A_n - \bigcap_{k=1}^{n} B_k \right) && \text{since } \cap_{k=1}^{n} B_k \subset B_n \subset A_n \\
\implies \mu(A_n) &= \mu(\cap_{k=1}^{n} B_k) \;+\; \mu(A_n - \cap_{k=1}^{n} B_k) && \text{countable additivity} \\
&= \cancelto{0}{\mu(\cap_{k=1}^{n} B_k)} \;+\; \mu(A_n - \cap_{k=1}^{n} B_k) && \text{for sufficiently large } n\text{, by item c) above} \\
&\le \mu(\cup_{k=1}^{n} (A_k - B_k)) && \text{monotonicity, since } A_n - \cap_{k=1}^{n} B_k \overset{\text{DeMorgan}}{=} \cup_{k=1}^{n} (A_n - B_k) \subset \cup_{k=1}^{n} (A_k - B_k) \\
&\le \sum_{k=1}^{n} \mu(A_k - B_k) && \text{finite subadditivity} \\
&= \sum_{k=1}^{n} \mu(A_k) - \mu(B_k) && \text{piece-and-difference decomposition; also uses finiteness} \\
&\le \epsilon \sum_{k=1}^{n} 2^{-k} && \text{Choose } B_k \text{ such that } \mu(A_k) - \mu(B_k) < \epsilon 2^{-k} \\
&< \epsilon.
\end{aligned}$$

So for sufficiently large n, we have $\mu(A_n) < \epsilon$ for any fixed $\epsilon > 0$. Thus, $\mu(A_n) \to 0$ for $A_n \downarrow \emptyset$, and so μ is continuous from above. So by Theorem 3.4.2 (b), μ is countably additive. $\square$

Remark 5.2.1. The proof of Lemma 5.2.1 is a very cool application of Heine-Borel! In trying to show continuity from above, we started out with an *infinite* intersection of sets. But in showing continuity, we needed to work with *finite* collection so that we could apply *finite* subadditivity, since that's all we had to use, by assumption. △

Theorem 5.2.1. *Let F be a distribution function on $\mathbb{R}$, and let $\mu(a, b] = F(b) - F(a), a < b$. Then there is a unique extension of μ to a Lebesgue-Stieltjes measure on $\mathbb{R}$.*

Proof. See text. □

Remark 5.2.2. The proof of Theorem 5.2.1 essentially directly applies Carathéodory's Extension Theorem, since we know from Lemma 5.2.1 that μ is countably additive on $\mathcal{F}_0(\mathbb{R})$, a field from which the Borel sets are generated. The only real additional work is a tedious technical detail to identify a μ-preserving correspondence between sets in $\mathcal{F}_0(\overline{\mathbb{R}})$ (over which we proved countable additivity) and sets in $\mathcal{F}_0(\mathbb{R})$ (which is the field we actually want to extend). △

5.3 § 1.4.5 Properties of Lebesgue-Stietljes measures

Before extension, we had $\mu(a, b] = F(b) - F(a)$ for $a < b$ where F is a distribution function. The set function μ was defined only on $\mathcal{F}_0(\mathbb{R})$, the field of disjoint unions of r.s.c interval. But after extension, μ is defined on $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_0(\mathbb{R}))$, which allows us to measure other types of intervals as well (by expressing those intervals as countable unions or intersections of r.s.c intervals; recall (3.1.1)).

Proposition 5.3.1. *Let μ be a Lebesgue-Stieltjes measure, and let F be its associated distribution function. Let $F(x^-) = \lim_{y \uparrow x} F(y)$. Then*

a) $\mu(a, b] = F(b) - F(a)$

b) $\mu(a, b) = F(b^-) - F(a)$

c) $\mu[a, b] = F(b) - F(a^-)$

d) $\mu[a, b) = F(b^-) - F(a^-)$

e) $\mu\{x\} = F(x) - F(x^-)$

f) $\mu(-\infty, x] = F(x) - F(-\infty)$

g) $\mu(-\infty, x) = F(x^-) - F(-\infty)$

h) $\mu(x, \infty) = F(\infty) - F(x)$

i) $\mu[x, \infty) = F(\infty) - F(x^-)$

j) $\mu(\mathbb{R}) = F(\infty) - F(-\infty)$

Proof. We prove some of these statements and leave the rest to the reader.

For (b), note that $(a, b) = \cup_{n=1}^{\infty}(a, b - \frac{1}{n}]$. So let $A_n = (a, b - \frac{1}{n}]$. Then by continuity from below,

$$\mu(a, b) = \lim_{n \to \infty} \mu(A_n) = \lim_{n \to \infty} \left[F(b - \frac{1}{n}) - F(a)\right] = F(b^-) - F(a)$$

For (c), note that $[a, b] = \cap_{n=1}^{\infty}(a - \frac{1}{n}, b]$. So by continuity from above (which applies since the sets in the intersection have finite measure),

$$\mu(a, b] = \lim_{n \to \infty} \left[F(b) - F(a - \frac{1}{n})\right] = F(b) - F(a^-)$$

For (e), note that $\{x\} = \cap_{n=1}^{\infty}(x - \frac{1}{n}, x]$. So the statement follows by the same argument as used in (c).

For (i), we can write $[x, \infty) = \cup_{n=1}^{\infty}[x, x+n)$. So by continuity from below,

$$\mu[x,\infty) = \lim_{n\to\infty} \mu[x, x+n) \stackrel{(d)}{=} \lim_{n\to\infty} \big[F\big((x+n)^-\big) - F(x^-)\big] = F(\infty) - F(x^-)$$

For (j), we can write $\mathbb{R} = \cup_{n=1}^{\infty}[-n, n]$. So by continuity from below,

$$\mu(\mathbb{R}) = \lim_{n\to\infty} \mu[-n, n] \stackrel{(c)}{=} \lim_{n\to\infty} \big[F(n) - F(-n)\big] = F(\infty) - F(-\infty)$$

□

Remark 5.3.1. (*Continuity at a point iffi measure zero at a point*)

1. Note that

$$\mu\{x\} = 0 \quad \Leftrightarrow \quad \text{F is continuous at } x \tag{5.3.1}$$

   which holds by Proposition 5.3.1 part e) and the fact that F is already right-continuous by definition.

2. The magnitude of the discontinuity corresponds with the measure of $\{x\}$.

For example, the measure corresponding to the distribution function in Figure 3 puts positive probability mass on the points $\{x_1\}, \{x_2\}, \{x_3\}$ and zero probability mass on all other points.

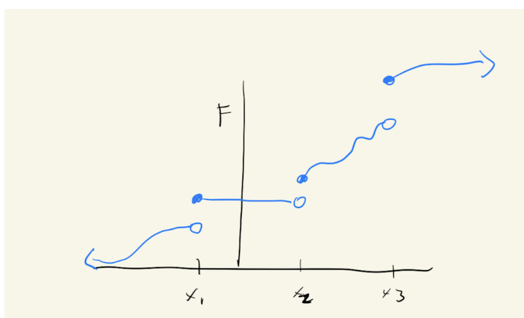


Figure 3: A distribution function with positive mass on points that is not concentrated on a countable set

△

Remark 5.3.2. The characterization of continuity in Remark 5.3.1 in terms of measure zero can be an interesting way to prove continuity, or prove the existence of functions with interesting properties. For instance, take a countable set $S = \{x_1, x_2, ...\}$ and non-negative weights $\{w_1, w_2, ...\}$. such that $\sum_i w_i < \infty$. Then define $\mu(A) = \sum_i \{w_i : x_i \in A\}$. Now μ is a Lebesgue-Stietljes measure (and is in fact a finite measure), since $\mu(I) < \infty$ for each bounded interval I. By taking S to be the rationals, we have proven the existence of an increasing function $F : \mathbb{R} \to \mathbb{R}$ that is continuous on the irrationals and discontinuous on the rationals [since each Lebesgue-Stieltjes measure determines a distribution function F (up to additive constant), and the set of continuities is given by (5.3.1)]. △

Remark 5.3.3. (*Lebesgue-Stieltjes measures of intervals for continuous distribution functions*) When a distribution function F is continuous rather than simply right continuous, the properties in

Proposition 5.3.1 reveal that the Lebesgue-Stieltjes measure of an interval does not depend upon whether the intervals are open or closed, i.e.

$$\mu(a,b] = \mu(a,b) = \mu[a,b) = \mu[a,b] = F(b) - F(a) \qquad \text{for } a \leq b \tag{5.3.2a}$$
$$\mu(-\infty,x) = \mu(-\infty,x] = F(x) - F(-\infty) \qquad \text{for } x \in \mathbb{R} \tag{5.3.2b}$$
$$\mu(x,\infty) = \mu[x,\infty) = F(\infty) - F(x) \qquad \text{for } x \in \mathbb{R} \tag{5.3.2c}$$

We will informally summarize this as $\mu(a,b] = \mu(a,b) = \mu[a,b) = \mu[a,b]$, where we may take $a, b \in \overline{\mathbb{R}}$ as long as we aren't closing the interval at $\pm\infty$. $\triangle$

Remark 5.3.4. Note that the properties in Proposition 5.3.1 hold even though differences (between a set and a subset) and measures don't commute outside of finite measures.[15] For instance, if we determine F from the equivalence class by setting $F(-\infty) = 0$, then property d) of Proposition 5.3.1 says

$$\mu[a,b) = \mu(-\infty,b) - \mu(-\infty,a).$$

But we couldn't make that statement by the piece-and-difference decomposition (see Theorem 3.3.1), since μ isn't necessarily finite. Thus, continuity of measure lets claim things that the piece-and-difference decomposition does not. $\triangle$

5.4 Examples of Lebesgue-Stieltjes measures on $\mathbb{R}$

Example 5.4.1. (*Lebesgue measure*) Under the identity distribution function ($F(x) = x$), we have $\mu(a,b] = F(b) - F(a)$. This is known as Lebesgue measure. Recall from Remark 5.3.3 that since F is continuous, we also have $\mu(a,b] = \mu(a,b) = \mu[a,b) = \mu[a,b]$. $\triangle$

Example 5.4.2. (*Generating Lebesgue-Stieltjes measures via integration*) We can generate a large class of measures on $\mathcal{B}(\mathbb{R})$ as follows. Let f be integrable (Riemann for now) on any finite interval, and define

$$F(b) - F(a) = \int_a^b f(t)\,dt$$

which determines F up to an additive constant. Then F is a distribution function (as it is both increasing and continuous), so it gives rise to a Lebesgue-Stieltjes measure $\mu(a,b] = F(b) - F(a)$. Lebesgue measure (Example 5.4.1) is a special case where $f \equiv 1$. Once again, Remark 5.3.3 reveals that by continuity of F, we have $\mu(a,b] = \mu(a,b) = \mu[a,b) = \mu[a,b]$. $\triangle$

A non-example. All Lebesgue-Stieltjes measures are sigma-finite. (To see this, simply set $\mathbb{R} = \cup_{n\in\mathbb{N}}(-n,n)$, and observe that $\mu(-n,n) < \infty$.). Here we provide an example of a sigma-finite measure that is not Lebesgue-Stieltjes. First, let μ be concentrated on S (i.e. $\mu(S^c) = 0$), where we set $S = \{1/n : n = 1, 2, ...\}$. Take $\mu\{1/n\} = 1/n$ for all n. Since $\mathbb{R} = \cup_{n=1}^{\infty} 1/n \cup S^c$, μ is sigma-finite. However,

$$\mu[0,1] \stackrel{\text{countable additivity}}{=} \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

and so μ is not a Lebesgue-Stieltjes measure.

5.5 Lebesgue measurable sets

Definition 5.5.1. The completion of Lebesgue measure relative to $\mathcal{B}(\mathbb{R})$ gives what is known as the *Lebesgue measurable sets*, denoted $\overline{\mathcal{B}}(\mathbb{R})$.[16] $\triangle$

[15] See Theorem 3.3.1.

[16] See Section 4.2 for the definition of the completion of a measure space.

Each Lebesgue measurable set is the union of a Borel set and a subset of a Borel set with Lebesgue measure zero.

Remark 5.5.1. Sometimes people use the term "Lebesgue measure" to refer to

$$\mu : \overline{\mathcal{B}}(\mathbb{R}) \to \mathbb{R}^+$$

as well as

$$\mu : \mathcal{B}(\mathbb{R}) \to \mathbb{R}^+$$

△

5.6 § 1.4.6 Lebesgue-Stieltjes Measures on $\mathbb{R}^n$

5.6.1 Overview

In $\mathbb{R}^n$, as with $\mathbb{R}$, is it possible to establish a one-to-one correspondence between Lebesgue-Stieltjes measures and distribution functions (up to some identification conditions). However, the details are quite tedious.

For our purposes, we will focus on

- Pointing out that, and motivating why, the definition of a distribution function must change in $\mathbb{R}^n$.
- Showing that if μ is a *finite* measure on the Borel sets of $\mathbb{R}^n$ and $F(x) = \mu(-\infty, x], x \in \mathbb{R}^n$, then F is a distribution function on $\mathbb{R}^n$ and $\mu(a, b]$ can be provided in terms of it. (The finite condition can be relaxed, but we omit this here.)
- Providing some examples of Lebesgue-Stieltjes distribution functions in $\mathbb{R}^n$.

5.6.2 Definitions

The definition of Lebesgue-Stieltjes measures on $\mathbb{R}^n$ parallels those on $\mathbb{R}$.

Definition 5.6.1. We define a *right semi-closed interval* (or right semi-closed box) in $\mathbb{R}^n$ as

$$(a, b] := (a_1, b_1] \times ... \times (a_n, b_n] = \{x \in \mathbb{R}^n : a_1 < x_1 \leq b_1, ..., a_n < x_n \leq b_n\}$$

△

Definition 5.6.2. The *vertices* of a right semi-closed interval in $\mathbb{R}^n$ are given by

$$V(a, b] = \{a_1, b_1\} \times ... \times \{a_n, b_n\}$$

△

Definition 5.6.3. The *Borel sets* of $\mathbb{R}^n$, denoted $\mathcal{B}(\mathbb{R}^n)$, are those sets which are members of the smallest sigma field containing all right semi-closed intervals $(a, b], a, b \in \mathbb{R}^n$. △

Definition 5.6.4. A *Lebesgue-Stieltjes measure* on $\mathbb{R}^n$ is a measure μ on $\mathcal{B}(\mathbb{R}^n)$ such that $\mu(I) < \infty$ for each bounded interval I. △

5.6.3 From (finite) measures on $\mathcal{B}(\mathbb{R}^n)$ to distribution functions

Recall that in $\mathbb{R}$, we observed the following relation between distribution functions and Lebesgue-Stieltjes measures on right semi-closed intervals

$$\mu(a,b] = F(b) - F(a), \quad a, b \in \mathbb{R}, a < b \tag{5.6.1}$$

In particular, we observed that given μ, we could construct an F (up to additive constant) via the above relationship. If we defined $F(-\infty) = 0$, then we could construct F from μ directly via

$$F(x) = \mu(-\infty, x] = \mu(\omega \in \mathbb{R} : \omega \leq x)$$

We would like to to do the same for $\mathbb{R}^n$. However, note that the equation

$$\mu(a,b] = F(b) - F(a), \quad a, b \in \mathbb{R}^n, a < b \tag{5.6.2}$$

does *not* hold anymore! To see this, let us define $F : \mathbb{R}^n \to \mathbb{R}$ via

$$F(x) = \mu(-\infty, x] = \mu(\omega \in \mathbb{R}^n : \omega_1 \leq x_1, ..., \omega_n \leq x_n)$$

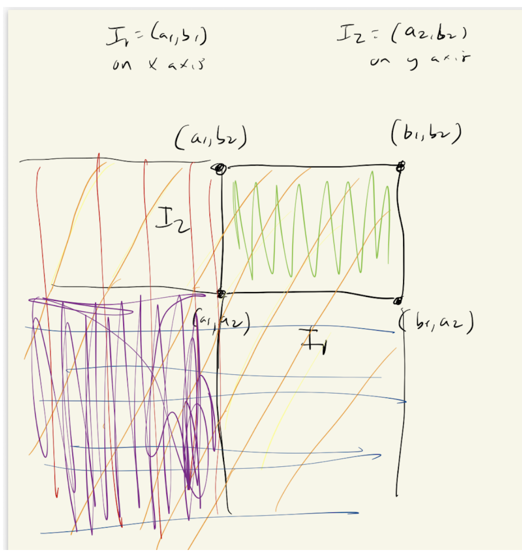


Figure 4: Using a distribution function in $\mathbb{R}^2$ to measure the box $I_1 \times I_2$.

Now consider Figure 4. We see that if $(a,b] = I_1 \times I_2 = (a_1, b_1] \times (a_2, b_2]$, then

$$\begin{aligned} \mu(a,b] &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \\ &\neq F(b_1, b_2) - F(a_1, a_2) \end{aligned} \tag{5.6.3}$$

(Note that we add back in the region that we had double subtracted.)

Now we generalize (5.6.3) to a formula for measuring r.s.c. intervals in n dimensions, rather than just 2 dimensions.

Theorem 5.6.1. *Let μ be a finite measure on $\mathcal{B}(\mathbb{R}^n)$. Define $F : \mathbb{R}^n \to \mathbb{R}$ via $F(x) = \mu(-\infty, x] = \mu(\omega \in \mathbb{R}^n : \omega_1 \leq x_1, ..., \omega_n \leq x_n)$. Then*

a) We have

$$\mu(a,b] = \Delta_{(a,b]} F := \Delta_{b_1 a_1} \cdots \Delta_{b_n a_n} F(x_1, ..., x_n) \tag{5.6.4}$$

where

$$\Delta_{b_i a_i} G(x_1, ..., x_n) := G(x_1, ..., x_{i-1}, b_i, x_{i+1}, ..., x_n) - G(x_1, ..., x_{i-1}, a_i, x_{i+1}, ..., x_n)$$

b) We have

$$\Delta_{(a,b]}F = \sum_{v\in V(a,b]} (-1)^{\#\, of\, a_i\text{'s in } v}\, F(v) \tag{5.6.5}$$

where $V(a,b]$ are the vertices of $(a,b]$ (see Definition 5.6.2).

Proof. We prove part (a) and leave (b) to the reader.

$$\begin{aligned}
\Delta_{b_n a_n}F(x_1,...,x_n) &= F(x_1,...,x_{n-1},b_n) - F(x_1,...,x_{n-1},a_n)\\
&= \mu(\{\omega_1 \le x_1,\ ...,\ \omega_{n-1} \le x_{n-1},\ \omega_n \le b_n\}) - \mu(\{\omega_1 \le x_1,\ ...,\ \omega_{n-1} \le x_{n-1},\ \omega_n \le a_n\})\\
&= \mu(\{\omega_1 \le x_1,\ ...,\ \omega_{n-1} \le x_{n-1},\ a_n < \omega_n \le b_n\})
\end{aligned}$$

where the last equality follows by the piece-and-difference decomposition of finite measures.

Similarly,

$$\begin{aligned}
&\Delta_{b_{n-1}a_{n-1}}\Delta_{b_n a_n}F(x_1,...,x_n)\\
&\qquad = \mu(\{\omega_1 \le x_1,\ ...,\ \omega_{n-2} \le x_{n-2},\ a_{n-1} < \omega_{n-1} \le b_{n-1},\ a_n < \omega_n \le b_n\})
\end{aligned}$$

Repeating this, we obtain

$$\Delta_{b_1 a_1}\cdots\Delta_{b_n a_n}F(x_1,...,x_n) = \mu(\{a_1 < \omega_1 \le b_1,\ ...\ a_n < \omega_n \le b_n\}) = \mu(a,b]$$

□

Remark 5.6.1. Note from the proof of Theorem 5.6.1 part (a) that the application of the nth difference operator restricts the set being measured to the bounds given in the nth dimension. See Figure 5.

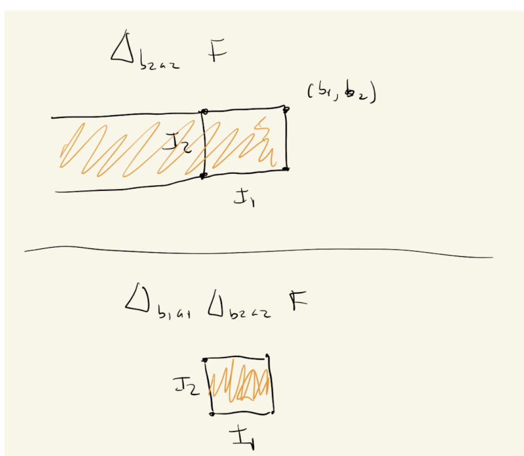


Figure 5: Repeated applications of the difference operator to a distribution function in $\mathbb{R}^2$.

△

Remark 5.6.2. Equation (5.6.5) tells us that we can measure any n-dimensional rectangle in $\mathbb{R}^n$ via 2^n evaluations of the distribution function. △

5.6.4 Defining distribution functions in $\mathbb{R}^n$

When defining distribution functions on $\mathbb{R}^n$, we must alter our notion of *increasing*. This is due to Theorem 5.6.1 part (a).

Definition 5.6.5. A *distribution function* on $\mathbb{R}^n$ is a map $F : \mathbb{R}^n \to \mathbb{R}$ that is:

a) *increasing*, i.e. its increments must be non-negative in the sense that

$$\Delta_{(a,b]}F \geq 0 \quad \text{for all r.s.c. intervals } (a, b] \tag{5.6.6}$$

b) *right continuous*, that is

$$\lim_{y \downarrow x} F(y) = F(x)$$

where $y \downarrow x$ means $y_i \downarrow x_i$ for each $i = 1, ..., n$.

△

Remark 5.6.3. Note that Definition 5.6.5 defines increasing in a different manner than what might be intuitive:

$$F(y) \geq F(x) \text{ if } y_i \geq x_i \quad \text{for all } i = 1, ..., n$$

However, such a condition would be insufficient to describe a distribution function in $\mathbb{R}^n$. For an example of a distribution function that is right continuous and increasing in this sense, but which can assign negative measure to an interval, see pp. 6-7 of [Durrett, 2010]. △

5.6.5 From distribution functions on $\mathbb{R}^n$ to Lebesgue-Stieltjes measures

Theorem 5.6.2. *Let F be a distribution function on $\mathbb{R}^n$, and let $\mu(a, b] = F(a, b], a, b \in \mathbb{R}^n, a \leq b$. Then there is a unique extension of μ to a Lebesgue-Stieltjes measure on $\mathbb{R}^n$.*

Proof. See text. □

5.6.6 Examples

Here we provide some examples of how Lebesgue-Stieltjes measures can be constructed on $\mathbb{R}^n$ via distribution functions.

1. Let $F_1, F_2, ..., F_n$ be distribution functions on $\mathbb{R}$, and define $F(x_1, ..., x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n)$. Then F is a distribution function on $\mathbb{R}^n$; it is clearly right-continuous, and it is increasing since

$$\Delta_{(a,b]}F = \prod_{i=1}^{n} [F(b_i) - F(a_i)] \geq 0$$

A special case is where each F_i is the distribution function corresponding to Lebesgue measure on $\mathcal{B}(\mathbb{R})$. Then each $F_i(x_i) = x_i$, and so we have

$$F(x_1, ..., x_n) = x_1 x_2 \cdots x_n$$

This μ is *Lebesgue measure* on $\mathcal{B}(\mathbb{R}^n)$. Note that

$$\mu(a, b] = \Delta_{(a,b]}F = \prod_{i=1}^{n}(b_i - a_i)$$

and more generally, the Lebesgue measure of any rectangular box is its volume (which can be seen by using a slight tweak to the arguments of parts (b)-(d) of the proof of Proposition 5.3.1).

2. Let f be any non-negative function from $\mathbb{R}^n$ to $\mathbb{R}$ such that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, ..., x_n)\; dx_1 \cdots dx_n < \infty$$

(For now, we assume the integration is in the Riemann sense.)

Define

$$F(x) = \int_{(-\infty, x]} f(t) dt$$

Then F is a distribution function. It is continuous by the fundamental theorem of calculus, and it is increasing since

$$\Delta_{(a,b]} F(x) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, ..., x_n)\; dx_1 \cdots dx_n < \infty$$

Remark 5.6.4. It may seem hard to verify (5.6.6), the condition that a distribution function on $\mathbb{R}^n$ must be increasing. Not to worry, the recipes above provide straightforward mechanisms for constructing distribution functions on $\mathbb{R}^n$ in which the condition will automatically be verified. $\triangle$

5.6.7 Summary

Let us summarize.[17] We have seen that if F is a distribution function on $\mathbb{R}^n$, then there is a unique Lebesgue-Stieltjes measure determined by $\mu(a, b] = \Delta_{(a,b]} F, a \leq b$. Also, if μ is a finite measure on $\mathcal{B}(\mathbb{R}^n)$ and $F(x) = \mu(-\infty, x], x \in \mathbb{R}^n$, then F is a distribution function on $\mathbb{R}^n$ and $\mu(a, b] = \Delta_{(a,b]} F, a \leq b$. It is possible to associate a distribution function with arbitrary Lebesgue-Stieltjes measure on $\mathbb{R}^n$, and thus to establish a one-to-one correspondence between Lebesgue-Stieltjes measures and distribution fucntions (provided distribution functions with the same increments $\Delta_{(a,b)} F, a, b \in \mathbb{R}^n, a \leq b$ are identified). However, the result will not be needed, and the details are quite tedious.

5.7 Properties of Borel sets under Lebesgue measure

Below we show some properties of Borel sets that hold under Lebesgue measure. How can we accomplish this? After all, the Borel sets are rather abstractly defined, and although we have been generating them via disjoint unions of r.s.c intervals, they also contain many types of members (e.g., proper open intervals, proper closed intervals, and singletons; bizarre sets like the Cantor set; inverse images of Borel measurable sets under Borel measurable functions; etc.).

To prove that such properties hold, we can take a standard tact: use the the Monotone Class Theorem, to show that the Borel sets have some property. Using this approach, we can show that a property holds for *all* Borel sets if we can just show that the property holds for some field generating the Borel sets (e.g., $\mathcal{F}_0 := \{\text{disjoint unions of } (a, b], a, b \in \mathbb{R}^n\}$) – a much more tangible object to work with.

Remark 5.7.1. (*Using the Monotone Class Theorem to prove that the Borel sets have some property*) Suppose you want to show that all Borel sets $\mathcal{B}(\mathbb{R}^n)$ have some property P. Define "good sets" as those that satisfy the property

$$\mathcal{G} := \{B \in \mathcal{B}(\mathbb{R}^n) : B \text{ has property } P\}$$

The strategy is then to simply

1. Show $\mathcal{G}$ contains $\mathcal{F}_0 := \{\text{disjoint unions of } (a, b], a, b \in \mathbb{R}^n\}$.

[17]This passage is basically a paragraph from [Ash et al., 2000] pp. 32 verbatim. However, we alter it slightly here to match our notation.

2. Show $\mathcal{G}$ is a monotone class.

$\triangle$

This is a particular version of the Good Sets Strategy (see Remark 4.1.2) in the special case where the σ-field of interest is the Borel sets (and where we take the field generating them to be $\mathcal{F}_0$). As pointed out in Remark 4.1.2, this strategy is very much like induction. Step #1 is the "base" step and Step #2 is the "induction" step.

For examples where this strategy is used, see the Approximation Theorem for Borel sets (Theorem 5.7.1) or the proof that Lebesgue measure is translation invariant (Proposition 5.7.1). We begin with the Approximation Theorem for Borel sets. This theorem shows that under appropriate conditions, a Borel set can be approximated from below by a compact set, and from above by an open set.

Theorem 5.7.1. ***(Approximation Theorem for Borel sets).*** *If μ is a σ-finite measure on $\mathcal{B}(\mathbb{R}^n)$, then for each $B \in \mathcal{B}(\mathbb{R}^n)$,*

a) $\mu(B) = \sup\{\mu(K) : K \subset B, K \text{ compact}\}$

b) If μ is in fact a Lebesgue-Stieltjes measure, then

$$\mu(B) = \inf\{\mu(V) : V \supset B, V \text{ open}\}$$

c) There is an example of a σ-finite measure on $\mathcal{B}(\mathbb{R}^n)$ that is not a Lebesgue-Stieljes measure for which (b) fails.

Proof.

a) We prove (a) for finite measures. For the extension to σ-finite measures, see the text.

We use the Monotone Class Theorem to show that all Borel sets have the desired property. Let $\mathcal{G}$ be the class of subsets that have the desired property.[18]

- *First, observe that $\mathcal{G}$ contains all compact sets.* If K is a compact set, then $\mu(K)$ is an upper bound on $\{\mu(K') : K' \subset K, K' \text{ compact}\}$ by monotonicity [$\mu(K) \geq \mu(K')$ for $K' \subset K, K'$ compact]. It is also the least upper bound since for each ϵ, there is a compact $K' \subset K$ satisfying $\mu(K') > \mu(K) - \epsilon$. [Just take $K' = K$].
- *Next, we show that $\mathcal{G}$ is a monotone class.* So we need to show that (i) if $B_n \in \mathcal{G}$ and $B_n \downarrow B$ then $B \in \mathcal{G}$ and (ii) if $B_n \in \mathcal{G}$ and $B_n \uparrow B$ then $B \in \mathcal{G}$.
  (i) Since each $B_n \in \mathcal{G}$, by definition of supremum (see Remark A.1.2), we can find $K_n \subset B_n$, K_n compact, such that
  $$\mu(B_n) \leq \mu(K_n) + \epsilon 2^{-n}$$
  Set $K = \cap_{n=1}^{\infty} K_n$. Then
  $$\begin{aligned} \mu(B) - \mu(K) &= \mu(B - K) && \text{piece-and-difference, } \mu \text{ finite} \\ &\overset{1}{\leq} \mu(\cup_{n=1}^{\infty}(B_n - K_n)) && \text{DeMorgan, monotonicity} \\ &\leq \sum_{n=1}^{\infty} \mu(B_n - K_n) && \text{countable subadditivity} \\ &= \sum_{n=1}^{\infty} \mu(B_n) - \mu(K_n) && \text{piece-and-difference, } \mu \text{ finite} \\ &= \epsilon \end{aligned}$$

[18] The reader may recognize that we are using the "good sets" strategy. See Section 3.1.3 and Remark **??**.

[For more detail, Equation (1) applies because $B - \cap_{n=1}^{\infty} K_n \overset{\text{DeMorgan}}{=} \cup_{n=1}^{\infty}(B - K_n) \overset{B \subset B_n}{\subset} \cap_{n=1}^{\infty}(B_n - K_n)$.]

So for all sets B formed by $B_n \downarrow B$ for $B_n \in \mathcal{G}$, we have that $\mu(B)$ satisfies the second property of the supremum (see Definition A.1.2). [It satisfies the first property immediately since $K_n \subset B_n \implies \cap_{n=1}^{\infty} K_n \subset \cap_{n=1}^{\infty} B_n$, so by monotonicity $\mu(K) \leq \mu(B)$, and so B is an upper bound.]

(ii) Up to reader or see text for proof.

- *Now we show that $\mathcal{G}$ contains $\mathcal{F}_0 := \{$disjoint unions of $(a, b], a, b \in \mathbb{R}^n\}$.* Consider that

$$(a, b] = \bigcup_{n=1}^{\infty} \underbrace{[a + \frac{1}{n}, b]}_{\text{compact}}$$

So $[a+1/n, b] \uparrow (a, b]$. And since $(a, b]$ is the limit of an increasing sequence of compact sets, $(a, b] \in \mathcal{G}$ by the first two bullet points. A similar argument holds for disjoint unions of sets which have the form $(a, b]$.

- *Now we use the Monotone Class Theorem to finish the proof.* By the previous bulletpoints, $\mathcal{G}$ contains $\mathcal{F}_0 := \{\text{disjoint unions of } (a, b], a, b \in \mathbb{R}^n\}$, and $\mathcal{G}$ is a monotone class. So by the Monotone Class Theorem (Theorem 4.1.2), $\mathcal{G}$ contains $\sigma(\mathcal{F}_0) = \mathcal{B}(\mathbb{R}^n)$.[19]

b) We prove part (b) for finite measures. For the extension to σ-finite measures, see the text.

We have

$$\begin{aligned}
\mu(B) &\overset{1}{\leq} \inf\{\mu(V) : V \supset B, V \text{ open}\} && \text{by monotonicity and Definition A.1.2} \\
&\overset{2}{\leq} \inf\{\mu(K^c) : K^c \supset B, K \text{ compact}\} && \text{by monotonicity and Proposition A.2.1} \\
&= \inf\{\mu(\mathbb{R}^n) - \mu(K) : K \subset B^c, K \text{ compact}\} && \text{by piece-and-difference, } \mu \text{ finite} \\
&\overset{3}{=} \mu(\mathbb{R}^n) - \sup\{\mu(K) : K \subset B^c, K \text{ compact}\} && \text{by Proposition A.2.3} \\
&= \mu(\mathbb{R}^n) - \mu(B^c) && \text{by part (a)} \\
&= \mu(B)
\end{aligned}$$

For more details, Equation (1) holds since, by monotonicity, the LHS is a lower bound on the RHS, so the statement must be true by definition of infimum. Equation (2) holds since the LHS is a smaller set than the RHS (because not every open set is the complement of a compact set)[20], and the infimum can only increase on subsets by Proposition A.2.1. Equation (3) holds by writing $\mu(K^c) = \mu(\mathbb{R}^n) - \mu(K)$. This has the form of a Minkowski set difference $A = \{c\} - B$, where c is a singleton. So we have $\inf A = \inf(\{c\} - B) \overset{Prop. A.2.3}{=} \inf\{c\} - \sup B = c - \sup B$.

c) See the text.

□

Proposition 5.7.1. ***(Translation Invariance of Lebesgue Measure).*** *Lebesgue measure is translation invariant. That is, if $B \in \overline{\mathcal{B}}(\mathbb{R}^n)$ and $c \in \mathbb{R}^n$, then $B + c \in \overline{\mathcal{B}}(\mathbb{R}^n)$ and $\mu(B + c) = \mu(B)$, where μ is Lebesgue measure.*

Proof. We prove the statement for $\mathcal{B}(\mathbb{R}^n)$ and leave the extension to $\overline{\mathcal{B}}(\mathbb{R}^n)$ to the reader. We shall use the Monotone Class Theorem as our vehicle for executing the Good Sets Strategy (see Remark

[19] In other words, *all* Borel sets are are “good" - they have the property stated in part (a).

[20] Recall that in $\mathbb{R}^n$, a compact set is both closed *and* bounded.

4.1.2). That is, we will let $\mathcal{G}$ be the class of "good sets" that have the desired property. Then we must show: (a) that $\mathcal{G}$ is a monotone class and (b) that $\mathcal{G}$ contains $\mathcal{F}_0 = \{\text{disjoint union of sets of the form } (a,b], a, b \in \mathbb{R}^n\}$. We will use this strategy twice, to show: (1) that $B \in \mathcal{B}(\mathbb{R}^n)$ and $c \in \mathbb{R}^n$ implies $B + c \in \mathcal{B}(\mathbb{R}^n)$ (2) that $\mu(B+c) = \mu(B)$ for all $B \in \mathcal{B}(\mathbb{R}^n)$.

1. We want to show that $B \in \mathcal{B}(\mathbb{R}^n) \implies B + c \in \mathcal{B}(R^n)$. Let $\mathcal{G}$ be the sets where the property holds.

   a) Consider a sequence $B_n \uparrow B$ such that $B_n \in \mathcal{G}$. That is, by hypothesis, we have $B_n \in \mathcal{B}(\mathbb{R}^n) \implies B_n + c \in \mathcal{B}(R^n)$. Then

   $$B + c = (\bigcup_{n=1}^{\infty} B_n) + c = \bigcup_{n=1}^{\infty} \underbrace{(B_n + c)}_{\text{in } \mathcal{B}(\mathbb{R}^n) \text{ by hypothesis}} \underbrace{\in \mathcal{B}(\mathbb{R}^n)}_{\text{by } \sigma\text{-field}}$$

   So $B \in \mathcal{G}$.

   b) This property holds on $\mathcal{F}_0$; that is $\mathcal{G} \supset \mathcal{F}_0$. Given $(a,b] \in \mathcal{F}_0$, $(a,b]+c = (a+c, b+c] \in \mathcal{F}_0$. A similar statement holds for disjoint unions of r.s.c. intervals.

2. We want to show that $\mu(B+c) = \mu(B)$ for all $B \in \mathcal{B}(\mathbb{R}^n)$.

   a) Let $\mathcal{G}$ be the sets where the property holds. We show $\mathcal{G}$ is a monotone class.

   First, we handle increasing sequences. So we want to show $B_n \in \mathcal{G}, B_n \uparrow B \implies B \in G$. Now by hypothesis, $\mu(B_n + c) = \mu(B_n)$. So

   $$\begin{aligned}
   \mu(B+c) &= \mu\big(\cup_{n=1}^{\infty} (B_n) + c\big) && \text{def. of } B \\
   &= \mu\big(\cup_{n=1}^{\infty} (B_n + c)\big) && \text{def. of union and +; still an increasing sequence} \\
   &= \lim_{n\to\infty} \mu(B_n + c) && \text{continuity from below} \\
   &= \lim_{n\to\infty} \mu(B_n) && \text{hypothesis} \\
   &= \mu(B) && \text{continuity from below}
   \end{aligned}$$

   Now we handle decreasing sequences. So we want to show $B_n \in \mathcal{G}, B_n \downarrow B \implies B \in G$. We could use the same argument as above with continuity from above instead of continuity from below, but continuity from above only applies for sets with finite measure. However, Lebesgue measure is σ-finite, so we can handle this problem in the standard way.[21]

   b) Now we show that $\mathcal{G}$ contains $\mathcal{F}_0$. The property certainly holds for r.s.c intervals; that is, $\mu(a+c, b+c] = \mu(a,b]$. [23] For Lebesgue measure on $\mathbb{R}^n$, the distribution function is

---

[21] In particular, we write $\Omega = \cup_{n=1}^{\infty} \Omega_n$ where $\mu(\Omega_n) < \infty$.[22] Then we define a finite measure μ_n via $\mu_n(A) := \mu(A \cap \Omega_n)$. We have

$$\mu(A) = \sum_{n=1}^{\infty} \mu_n(A), \quad (*)$$

since $\mu(A) = \mu(\cup A \cap \Omega_n) = \sum_n \mu(A \cap \Omega_n) = \sum_n \mu_n(A)$.

Now, using the continuity from above argument for finite measures, we have

$$\mu_n(B+c) = \mu_n(B), \quad (+)$$

And so

$$\mu(B+c) \overset{(*)}{=} \sum_n \mu_n(B+c) \overset{(+)}{=} \sum_n \mu_n(B) \overset{(*)}{=} \mu(B)$$

[23] For example, in $\mathbb{R}$, we have

$$\mu\big((a,b]+c\big) = \mu\big((a+c, b+c]\big) = (b+c) - (a+c) = b - a = \mu(a,b]$$

$F(x_1, ..., x_n) = x_1 \cdots x_n$, and so $\mu(a, b] = \Delta_{b_1 a_1} \cdots \Delta_{b_n a_n}[x_1 \cdots x_n]$. Now note that

$$\begin{aligned}\Delta_{b_i+c_i, a_i+c_i}[x_1 \cdots x_n] &= x_1 \cdots x_{i-1}\Big((b_i + c_i) - (a_i + c_i)\Big)x_{i+1} \cdots x_n \\ &= x_1 \cdots x_{i-1}\Big(b_i - a_i\Big)x_{i+1} \cdots x_n \\ &= \Delta_{b_i a_i}[x_1 \cdots x_n]\end{aligned}$$

So $\Delta_{b_i+c_i, a_i+c_i} = \Delta_{b_i, a_i}$ for each $i = 1, ..n$. Thus,

$\mu(a+c, b+c] = \Delta_{b_1+c_1, a_1+c_1} \cdots \Delta_{b_n+c_n, a_n+c_n}[x_1 \cdots x_n] = \Delta_{b_1 a_1} \cdots \Delta_{b_n a_n}[x_1 \cdots x_n] = \mu(a, b]$

A similar proof holds for disjoint unions of r.s.c intervals.

□

5.8 A set that is not Lebesgue measurable

Proposition 5.8.1. *Call $x, y \in \mathbb{R}$ equivalent iff $x - y \in \mathbb{Q}$.*[24] *Define $A \subset [0, 1]$ as a set containing one member from each class. (This set exists by axiom of choice.) Then A is not Lebesgue measurable.*

Proof. a) First, we show that we can partition the reals as $\mathbb{R} = \dot\bigcup_{q\in\mathbb{Q}}(q + A)$.

- *(disjointedness)* Suppose $x \in q + A$ and $x \in r + A$ for some $r, q \in \mathbb{Q}, r \neq q$. Then $\exists\, a_1, a_2 \in A : q + a_1 = r + a_2 \implies a_1 - a_2 = r - q$. Now since $r \neq q$, we know that a_1 and a_2 are not the same. But they are in the same equivalence class, since $r - q \in \mathbb{Q}$. This contradicts how A was constructed.
- *(containment)* We show $\mathbb{R} \subset \dot\bigcup_{q\in\mathbb{Q}}(q + A)$ (as the other direction is obvious). If $x \in \mathbb{R}$ and a_x is its representative in A, then $x - a_x = q \in \mathbb{Q}$, and so $x \in q + A$.

b) Now we note that since $A \subset [0, 1]$, we have

$$\dot\bigcup_{q\in\mathbb{Q}, 0\leq q\leq 1} (q + A) \subset [0, 2]$$

So

$$2 = [0, 2] \overset{\text{subadditivity}}{\geq} \mu\Bigg(\dot\bigcup_{q\in\mathbb{Q}, 0\leq q\leq 1} q+A\Bigg) = \sum_{q\in\mathbb{Q}, 0\leq q\leq 1} \mu(q+A) \overset{\text{translation invariance}}{=} \sum_{q\in\mathbb{Q}, 0\leq q\leq 1} \mu(A)$$

This implies $\mu(A) = 0$, since the RHS of this equation is a countable sum of a constant, and so can only take on values 0 or ∞.

But then

$$\infty = \mu(\mathbb{R}) \overset{\text{part (a)}}{=} \mu\Bigg(\dot\bigcup_{q\in\mathbb{Q}}(q + A)\Bigg) = \sum_{q\in\mathbb{Q}} \mu(q + A) \overset{\text{translation invariance}}{=} \sum_{q\in\mathbb{Q}} \mu(A) \overset{\text{see above}}{=} 0. \Rightarrow\Leftarrow$$

□

[24] So, for example, some equivalence classes are:

- $e \sim e + \frac{1}{10} \sim e - \frac{1}{10} \sim e + \frac{50}{3} \sim ...$
- $\pi \sim \pi + \frac{1}{10} \sim \pi - \frac{1}{10} \sim \pi + \frac{50}{3} \sim ...$
- $1 \sim 1 + \frac{1}{10} \sim 1 - \frac{1}{10} \sim 1 + \frac{50}{3} \sim ...$

Remark 5.8.1. The proof of Proposition 5.8.1 only used the following two properties of Lebesgue measure:

- translation invariance
- finiteness on bounded intervals

Therefore, our argument shows that there cannot be a translation invariant measure λ (except $\lambda \equiv 0$) on the class of all subsets of $\mathbb{R}$ such that $\lambda(I) < \infty$ for all bounded intervals I. $\triangle$

6 § 1.5 Measurable functions and Integration

In this section, we will introduce the general theory of integration of a function with respect to a general measure, as introduced by Lebesgue. We will refer to this as *integration* or *Lebesgue integration*. We use use *(Lebesgue) integration against Lebesgue measure* to refer to the special case of integrating a function defined on a sub-domain of the real line with respect to the Lebesgue measure.

6.1 Intuition[25]

Folland summarizes the difference between the Riemann and Lebesgue approaches thus: "to compute the Riemann integral of f, one partitions the domain [...] into subintervals", while in the Lebesgue integral, "one is in effect partitioning the range of f " [Folland, 1999].

Figure 6 compares how the Riemann and Lebesgue approaches would approximate the area under the curve of a function $f : \mathbb{R} \to \mathbb{R}$.

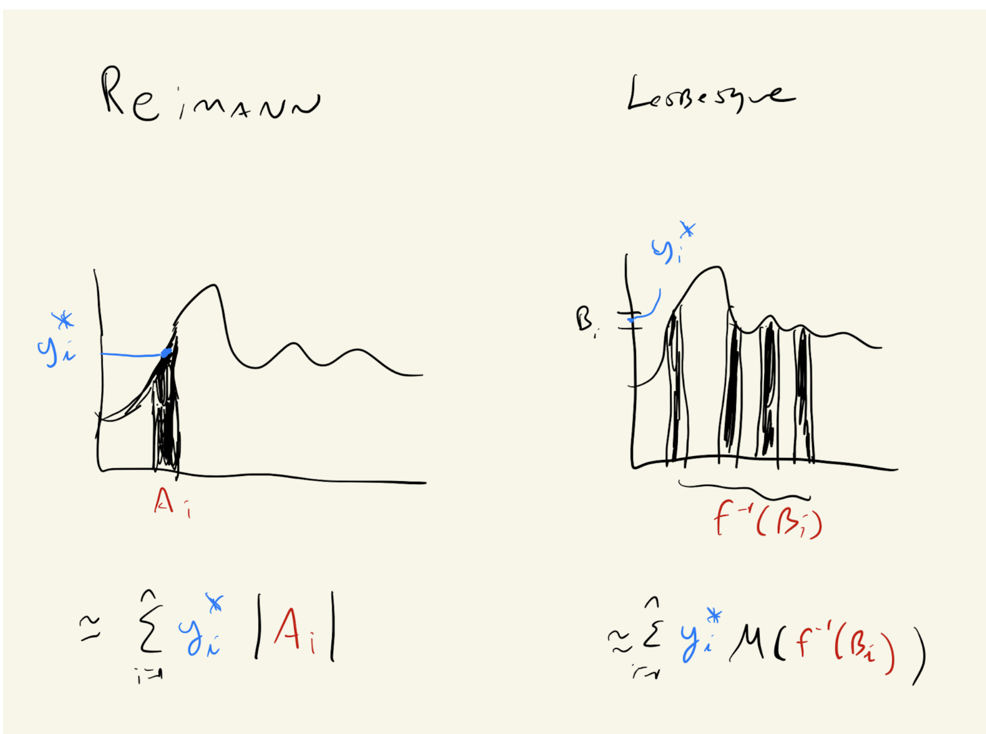


Figure 6: Illustrating the fundamental differences between Riemann and Lebesgue integration

[25] Here we borrow freely from some sections of Wikipedia.

In particular, notice:

1. Lebesgue integration partitions the range of f, whereas Riemann integration partitions the domain of f. As a result, the Lesbegue approach provides *adaptive grouping* when computing the area under the curve as the sum over n contributions. Whereas a function can vary a lot in Riemann subintervals of the form $A_i = (a_i, b_i)$, in the Lebesgue approach, the function will have controlled amount of variation for each of the n contributions. While the approaches give equivalent answers for sufficiently nice functions (like continuous functions), the Lebesgue definition makes it possible to calculate integrals for a broader class of functions. For example, as we will see, the *Dirichlet function*, which is 0 where its argument is irrational and 1 otherwise, has a Lebesgue integral, but does not have a Riemann integral.

   Lebesgue summarized his approach to integration in a letter to Paul Montel:

   > I have to pay a certain sum, which I have collected in my pocket. I take the bills and coins out of my pocket and give them to the creditor in the order I find them until I have reached the total sum. This is the Riemann integral. But I can proceed differently. After I have taken all the money out of my pocket I order the bills and coins according to identical values and then I pay the several heaps one after the other to the creditor. This is my integral.

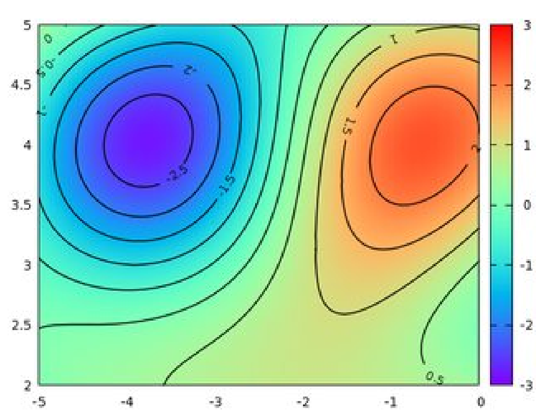
   The insight is that one should be able to rearrange the values of a function freely, while preserving the value of the integral. This process of rearrangement can convert a very pathological function into one that is “nice" from the point of view of integration, and thus let such pathological functions be integrated.

2. The Riemann approach implicitly assumes that sets in the domain have sizes that are given by Lebesgue measure ($\mu(A) = |A|$), whereas the Lebesgue approach allows sets in the domain to have sizes given by any arbitrary measure μ.

For another example with domain in $\mathbb{R}^2$, suppose we want to find a mountain’s volume (above sea level).

- **The Riemann approach**: Divide the base of the mountain into a grid of 1 meter squares. Measure the altitude of the mountain at the center of each square. The volume on a single grid square is approximately 1 $m^2\times$ (that square’s altitude), so the total volume is 1 m^2 times the sum of the altitudes.

- **The Lebesgue approach**: Draw a contour map of the mountain, where adjacent contours are 1 meter of altitude apart. The volume of earth a single contour contains is approximately 1 m $\times$ (that contour’s area), so the total volume is the sum of these areas times 1 m.



While the Riemann integral considers the area under a curve as made out of vertical rectangles, the Lebesgue definition considers slabs that are not necessarily just rectangles, and so it is more flexible.

6.2 § 1.5.1 Measurable functions

6.2.1 Definitions

Definition 6.2.1. If $\mathcal{F}$ is a σ-field of subsets of Ω, then $(\Omega, \mathcal{F})$ is called a *measurable space* and sets in $\mathcal{F}$ are called *measurable sets*. $\triangle$

We can now define measurable functions as those which preserve measurability under inverse images.

Definition 6.2.2. If $h : \Omega_1 \to \Omega_2$, h is a *measurable function* relative to the σ-fields $\mathcal{F}_j$ of subsets of Ω_j, $j = 1, 2$, iff $h^{-1}(A) \in F_1$ for all $A \in \mathcal{F}_2$. We sometimes denote measurable functions as an explicit mapping between measurable spaces: $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$. $\triangle$

Borel measurable functions are a special case of particular interest.

Definition 6.2.3. A *Borel measurable function* is a measurable function $h : (\Omega_1, \mathcal{F}_1) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ or $h : (\Omega_1, \mathcal{F}_1) \to (\overline{\mathbb{R}}^n, \mathcal{B}(\overline{\mathbb{R}}^n))$. $\triangle$

Note that the *Borel* in Borel measurability refers to the measurable sets in the *range*, not the domain. A more precise term would be $(\mathcal{F}_1, \textit{Borel})$*-measurable*, since the condition to be a measurable function depends on both sigma-fields. However, people do not say that. Unless stated otherwise, we assume $\mathcal{F}_1 = \mathcal{B}$ whenever Ω_1 is a Borel subset of $\mathbb{R}^k$ or $\overline{\mathbb{R}}^k$.

6.2.2 “Computational" definitions

In practice, to show that a function is measurable, it suffices to apply what we might call the “computational definition of measurable functions."

Claim 6.2.1. (*Computational definition of measurable functions*) For $h : \Omega_1 \to \Omega_2$ to be measurable relative to the σ-fields $\mathcal{F}_j$ of subsets of Ω_j, $j = 1, 2$, it suffices to show that $h^{-1}(B) \in \mathcal{F}_1$ for all $B \in \mathcal{C} : \sigma(\mathcal{C}) = \mathcal{F}_2$. $\triangle$

Proof. We apply the “Good Sets" strategy (see Section 3.1.3). The “base" condition is satisfied by hypothesis. For the “induction step", we need to show that the good sets form a σ-field.

Let us define the good sets as $\mathcal{G} := \{B \in \mathcal{F}_2 : h^{-1}(B) \in \mathcal{F}_1\}$. We check the three conditions:

- $\Omega_2 \in \mathcal{G}$? ✓ . True because $h^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$ by the fact that $\mathcal{F}_1$ is a σ-field.
- $B \in \mathcal{G} \implies B^c \in \mathcal{G}$? ✓ . Since complements and inverse images commute (see Section 6.2.4), we have $h^{-1}(B^c) = h^{-1}(B)^c \in \mathcal{F}_1$ by assumption and the fact that $\mathcal{F}_1$ is a σ-field, and hence closed under complements.
- $B_1, B_2, ... \in \mathcal{G} \implies \cup_{i=1}^{\infty} B_i \in \mathcal{G}$? ✓ . Since unions and inverse images commute (see Section 6.2.4), we have $h^{-1}(\cup_{i=1}^{\infty} B_i) = \cup_{i=1}^{\infty} h^{-1}(B_i) \in \mathcal{F}_1$ by assumption and the fact that $\mathcal{F}_1$ is a σ-field, and hence closed under countable unions.

$\square$

Remark 6.2.1. (*The computational definition of Borel measurability.*) Thanks to Claim 6.2.1, to show that a function h is Borel measurable, we simply need to show that $h^{-1}(B) \in \mathcal{F}_1$ for $B \in \mathcal{C}$ for any collection $\mathcal{C}$ that generates the Borel sets. For instance, when h is real-valued, it suffices to show

- $h^{-1}\big((a, \infty)\big) \in \mathcal{F}_1$ for each $a \in \mathbb{R}$.

- $h^{-1}\big([a,\infty)\big) \in \mathcal{F}_1$ for each $a \in \mathbb{R}$.
- $h^{-1}\big((a,b)\big) \in \mathcal{F}_1$ for each $a, b \in \mathbb{R}$.
- $h^{-1}\big([a,b]\big) \in \mathcal{F}_1$ for each $a, b \in \mathbb{R}$.
- $h^{-1}(U) \in \mathcal{F}_1$ for each open set $U \subset \mathbb{R}$.
- $h^{-1}(V) \in \mathcal{F}_1$ for each closed set $V \subset \mathbb{R}$.
- etc.

For a larger list, recall Section 3.1.4. △

6.2.3 Examples

Example 6.2.1. (*Constant functions are Borel measurable*) Consider a constant function, i.e. $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ such that $h(\omega) = c$ for all $\omega \in \Omega_1$. Then h is Borel measurable, since

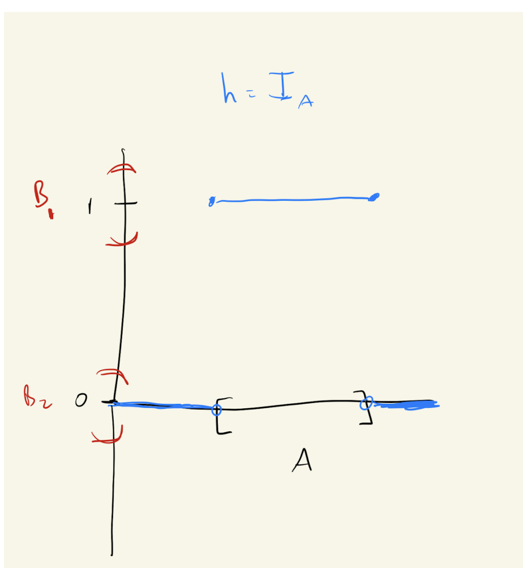
$$h^{-1}(B) = \begin{cases} \Omega, & \text{if } c \in B \\ \emptyset, & \text{if } c \notin B \end{cases}$$

△

Example 6.2.2. (*Any function is measurable with respect to the trivial σ-field.*) Consider any function $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$, where $\mathcal{F}_2$ is the trivial σ-field: $\mathcal{F}_2 = \{\emptyset, \Omega_2\}$. Then h is measurable since $h^{-1}(\emptyset) = \emptyset \in \mathcal{F}_1$ and $h^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$ △

Example 6.2.3. (*Indicators of Borel sets are Borel measurable.*) Let A be a Borel subset of $\mathbb{R}$,[26] and let $I_A : \mathbb{R} \to \mathbb{R}$ be the indicator of A; that is $I_A(\omega) = 1$ for $\omega \in A$ and 0 for $\omega \notin A$. Then I_A is Borel measurable, since for all $B \in \mathcal{B}(\mathbb{R})$, we have

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



[26] Recall Section 3.1.4. For instance, we might take A to be an open interval, or a disjoint union of open intervals, or the Cantor set.

$\triangle$

Example 6.2.4. (*Indicators of non-Borel sets are not Borel measurable - but they may still be measurable.*) Let A be a subset of $\mathbb{R}$ that is not Borel (e.g., A could be the non-Borel set described in Section 5.8), and let $I_A : \mathbb{R} \to \mathbb{R}$ be the indicator of A. Then I_A is *not* Borel measurable.

However, I_A *is* measurable with respect to the trivial sigma-field; that is, if we take the mapping to be $I_A : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{F}_2)$, where $F_2 := \{\emptyset, \mathbb{R}\}$. See Example 6.2.2. $\triangle$

Example 6.2.5. (*Continuous functions are Borel measurable*). Let $h : \mathbb{R}^k \to \mathbb{R}^n$ be continuous. Since h is continuous, the inverse image of any open set is open. Hence h is Borel measurable by the computational definition of Borel measurability – see Remark 6.2.1. $\triangle$

6.2.4 Why define measurability this way?

Measurable functions do *not* preserve measurability in *both* directions. That is, if $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ is a measurable function, it is not necessarily true that $h(A) \in \mathcal{F}_2$ for all $A \in \mathcal{F}_1$. For a counterexample, we take $\mathcal{F}_2 = \{\emptyset, \Omega_2\}$, recalling Example 6.2.2. Then any h is measurable. But if there is $A \in F_1$ such that $h(A)$ is a nonempty proper subset of Ω_2, then it is not a measurable set ($h(A) \notin \mathcal{F}_2$).

So why is measurability defined by preserving measurability over *inverse* images, rather than in terms of direct images? In measure theory, inverse images are much nicer objects than direct images. This is because basic set operations are preserved by inverse images, but in general not by images.

In particular, for any function f

a) Inverse images and complements commute: $f^{-1}(B^c) = \big(f^{-1}(B)\big)^c$

   Proof. $\big(f^{-1}(B)\big)^c := \{x : x \notin f^{-1}(B)\} = \{x : f(x) \notin B\} = \{x : f(x) \in B^c\} = \{x : x \in f^{-1}(B^c)\} := f^{-1}(B^c)$.

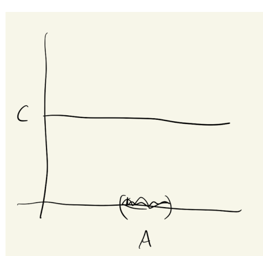
b) Inverse images and unions commute: $f^{-1}(\cup_i B_i) = \cup_i \big(f^{-1}(B_i)\big)$

c) Inverse images and intersections commute: $f^{-1}(\cap_i B_i) = \cap_i \big(f^{-1}(B_i)\big)$

However,

d) Direct images and complements do not in general commute: $f(A^c) \neq \big(f(A)\big)^c$

   Proof. Let $f : \Omega \to \mathbb{R}$ be the constant function, i.e. $f(\omega) = c$ for some $c \in \mathbb{R}$ for all $\omega \in \Omega$. Let A be a non-empty proper subset of Ω. Then $h(A) = c$ and $h(A^c) = c$. So $\mathbb{R} \setminus c = \big(h(A)\big)^c \neq h(A^c) = c$.



e) Direct images and intersections do not in general commute: $f(\cap_i A_i) \neq \cap_i \big(f(A_i)\big)$

Recall in Section 6.1 that Lebesgue and Riemann integration are distinguished in terms of whether they partition the range or domain of the function. My speculation is that the nice interplay of basic set operations and inverse images (but not direct images) at least partially explains why Lebesgue's approach has been more successful than Riemann's approach (in the sense of better limit theorems, better handling of non-Euclidean spaces, etc.).

6.2.5 Closure properties

Proposition 6.2.1. *If $h_1, h_2 : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ are Borel measurable, then so are $h_1 + h_2$ and $h_1 h_2$.*

Proof. See [Folland, 1999] Proposition 2.6. □

Proposition 6.2.2. *If $\{h_n\}$ is a sequence of $\overline{\mathbb{R}}$-valued Borel measurable functions on $(\Omega, \mathcal{F})$, then the functions*

$$\begin{aligned} &\sup_n h_n(\omega), \qquad &&\limsup_{n\to\infty} h_n(\omega) \\ &\inf_n h_n(\omega), \qquad &&\liminf_{n\to\infty} h_n(\omega) \end{aligned}$$

are all measurable. Thus, if $h(\omega) = \lim_{n\to\infty} h_n(\omega)$ exists for all $\omega \in \Omega$, then h is measurable.

Proof. See [Folland, 1999] Proposition 2.7. □

6.3 § 1.5.2-1.5.3 Integrating Borel measurable functions

In this section, we define integral of a Borel measurable function $h : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ against arbitrary measure μ. The integral can be written as:

$$\int_\Omega h \, d\mu \;, \qquad \int_\Omega h(\omega) \, d\mu(\omega) \;, \quad \text{or} \quad \int_\Omega h(\omega)\mu(d\omega)$$

We proceed in three steps: first, we consider where h is simple, then we consider h non-negative, then we consider h arbitrary.

6.3.1 Integrals of simple functions

Definition of simple functions

Definition 6.3.1. Let $(\Omega, \mathcal{F})$ be a measurable space, fixed throughout the discussion. If $h : \Omega \to \overline{\mathbb{R}}$, h is said to be *simple* iff h is measurable and takes on only finitely many distinct values. That is, h is simple iff it can be written $h = \sum_{i=1}^r y_i I_{A_i}$ where the A_i are disjoint sets in $\mathcal{F}$ and I_{A_i} is the indicator of A_i; the y_i need not be distinct. △

Remark 6.3.1. (*Simple functions generalize step functions*) A special case of simple functions are the step functions used in Riemann integration.

For $a, b \in \overline{\mathbb{R}}$ with $a < b$, $f : [a, b] \to \mathbb{R}$ is a *step function* if there exists a partition $a = x_0 < x_1 < ... < x_n = b$ and constants $y_1, ..., y_n \in \mathbb{R}$ such that $f(x) = y_i$ for all $y \in (x_{i-1}, x_i)$ and each $i = 1, ..., n$. Then f is equal to the following simple function:

$$y_i I_{(x_{i-1}, x_i)} + f(x_i) I_{\{x_i\}}$$

Note that in this case, the sets A_i take on a specific form, as open intervals or one of finitely many singletons. In general, simple functions allow more general A_i.

△

Remark 6.3.2. Note that the indicator function for a non-Borel set (see Example 6.2.4) is *not* a simple function, even though it only takes on values 0 and 1. △

Definition of the integral of simple functions

Definition 6.3.2. Let h be simple, say $h = \sum_{i=1}^{r} y_i I_{A_i}$ where the A_i are disjoint sets in $\mathcal{F}$. Then

$$\int_{\Omega} h \, d\mu \; := \sum_{i=1}^{r} y_i \; \mu(A_i). \tag{6.3.1}$$

△

The integral of a simple function can also be expressed as

$$\int_{\Omega} h \, d\mu \; := \sum_{i=1}^{r} y_i \; \mu\Big(h^{-1}(y_i)\Big)$$

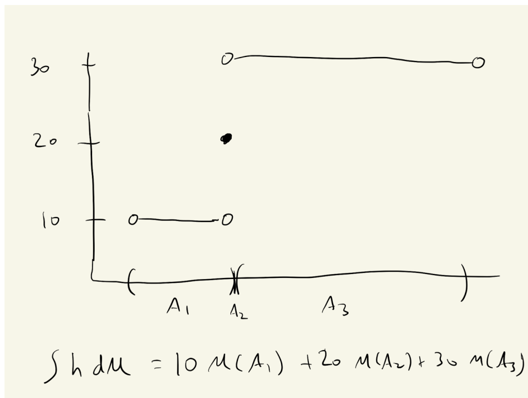
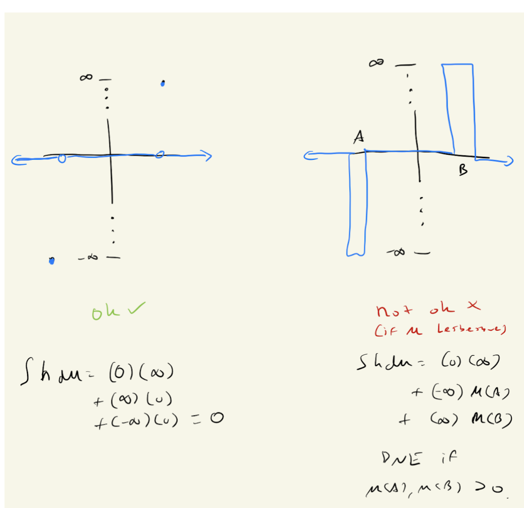


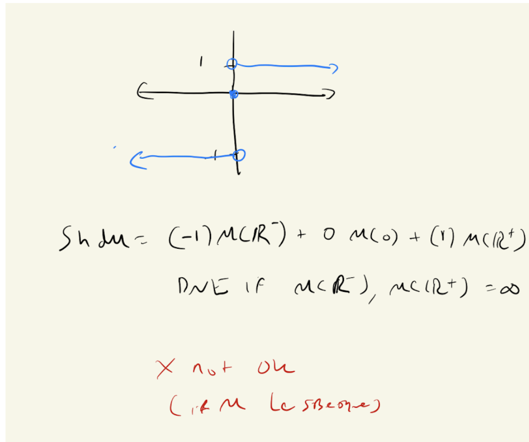
Figure 7: The Lebesgue integral of a simple function. (In this case, the simple function is also a step function.)

Remark 6.3.3. (*When does the integral of a simple function exist*) The integral of a simple function exists whenever ∞ and $-\infty$ do not both appear in the sum. So in particular, the integral for h does not exist when

- The finite values it takes on $\{y_i\}_{i=1}^{r}$ include ∞ and $-\infty$ on sets that are not of measure zero.



- It takes on values of opposite signs on two sets of infinite measure.



△

Remark 6.3.4. (*Integrating simple functions over arbitrary measurable subsets*) For any $A \in \mathcal{F}$, we define

$$\int_A h \, d\mu \; := \int_\Omega h \, I_A \, d\mu$$

This definition is possible because whenever h is a simple function, then so is hI_A for any measurable set A. Indeed, if we express $h = \sum_{i=1}^r y_i I_{B_i}$, then $hI_A = \sum_{i=1}^r y_i I_A I_{B_i} = \sum_{i=1}^r y_i I_{A \cap B_i}$, and each $A \cap B_i \in \mathcal{F}$. △

Comparison to Riemann integration

Example 6.3.1. (*Integrating a step function*) Let h be a step function, as defined in Remark 6.3.1. So $a = x_0 < x_1 < \ldots < x_n = b$ is a partition of $\Omega := \text{dom}(h)$, and h takes on values y_i on (x_{i-1}, x_i) for $i = 1, \ldots, n$. Then

$$\begin{aligned}
\int_\Omega h \, d\mu \; &= \sum_{i=1}^n y_i \; \mu(x_{i-1}, x_i) + \sum_{i=1}^n f(x_i)\mu\{x_i\} \\
&\overset{\text{(if } \mu \text{ is Lebesgue measure)}}{=} \sum_{i=1}^n y_i \; (x_i - x_{i-1}) + \overset{0}{\cancel{\sum_{i=1}^n f(x_i)\mu\{x_i\}}}
\end{aligned}$$

So the Lebesgue integral of the step function agrees with the Riemann integral when μ is Lebesgue measure (although not for general measure). △

Now we integrate a simple function that is not a step function – in fact, a simple function for which there is not a Riemann integral.

Example 6.3.2. (*Integrating the Dirichlet function*) Let h be the Dirichlet function; that is $h =$

$I_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ is the indicator of the rationals. Let us integrate h against Lebesgue measure μ.

$$\begin{aligned}
\int_\Omega h \, d\mu &= 1 \; \mu(\mathbb{Q}) + 0 \; \mu(\mathbb{R} - \mathbb{Q}) && \text{def. integral of simple function} \\
&= 1 \; \mu(\mathbb{Q}) && \text{arithmetic of } \overline{\mathbb{R}}\text{: } 0 \cdot x = 0 \text{ for } x \in \overline{\mathbb{R}} \\
&= 1 \; \mu(\dot{\bigcup_{q \in \mathbb{Q}}} \{q\}) && \text{rewrite } \mathbb{Q} \\
&= 1 \sum_{q \in \mathbb{Q}} \mu(\{q\}) && \text{countable additivity, } \mathbb{Q} \text{ is countable} \\
&\stackrel{1}{=} 1 \sum_{q \in \mathbb{Q}} 0 && \text{Proposition 5.3.1} \\
&= 0.
\end{aligned}$$

Note Equation 1 holds for any Lebesgue-Stieljes measure with a continuous distribution function (see Remark 5.3.1). However, other measures may yield other results. $\triangle$

Now we show that the Dirichlet function does not have a Riemann integral.

Remark 6.3.5. (*The Dirichlet function does not have a Riemann integral.*) Let h be the Dirichlet function; that is $h = I_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ is the indicator of the rationals. Fix $[a, b] \subset \mathbb{R}$, and let $f = h|_{[a,b]}$; that is $f : [a, b] \to \mathbb{R}$ is the Dirichlet function restricted to $[a, b]$.

We consider an arbitrary *partition* P of $[a, b]$ into a collection of n subintervals via a finite sequence $P = \{x_i\}_{i=0}^n$ such that $a = x_0 < x_1 < ... < x_n = b$. Now by Proposition B.0.1, a function on $[a, b]$ is Riemann integrable iff $\text{Osc}(f, P) \to 0$ as the maximum interval length of a partition P goes to 0.[27] But for any partition P, any subinterval (x_{k-1}, x_k) will contain at least one rational number and at least one irrational number, and so

$$S^+(f, P) = \sum_{k=1}^n 1 \cdot (x_k - x_{k-1}) = b - a$$
$$S^-(f, P) = \sum_{k=1}^n 0 \cdot (x_k - x_{k-1}) = 0$$

Thus $\text{Osc}(f, P) = b - a$ for all partitions P. In particular $\text{Osc}(f, P) \not\to 0$ as the maximum interval length of a partition P goes to 0.[28] Thus f is not Riemann integrable.

The Riemann integral of $h : \mathbb{R} \to \mathbb{R}$ would be an improper integral defined as the limiting value of the Riemann integrals $h|_{[a,b]}$ as $a \to -\infty, b \to \infty$. But since the proper Riemann integrals for $h|_{[a,b]}$ don't exist, neither does the improper Riemann integral for h. $\triangle$

Properties of integrals of simple functions

In this section, for now, we only provide properties that come up in our exposition. For other useful properties, see Proposition 2.13 of [Folland, 1999].

Proposition 6.3.1. *Let g and h be simple functions. Then*

a) (monotonicity) If $g \leq h$ then $\int g \, d\mu \leq \int h \, d\mu$.

[27] For a refresher on how these terms are defined, see Section B.

[28] To be more explicit, the condition of Proposition B.0.1 is that $\forall \epsilon > 0, \quad \exists \delta > 0 : \forall P$ where the maximum interval length of $P < \delta, \;\; \text{Osc}(f, P) < \epsilon$. But since $\text{Osc}(f, P) = b - a$ for all partitions P, the condition is contradicted: $\exists \epsilon = \frac{b-a}{2} > 0 : \forall P, \text{Osc}(f, P) > \epsilon$.

b) (addivity) $\displaystyle\int (g+h)\, d\mu \; = \int g\, d\mu \; + \int h\, d\mu \;.$

c) (scalar multiple property) $\displaystyle\int c\, g\, d\mu \; = c \int g\, d\mu \,.$

Proof. a) By definition of simple functions, we write

$$g = \sum_{i=1}^{r} x_i I_{A_i}, \qquad h = \sum_{j=1}^{s} y_j I_{B_j}$$

But if we use a common (finer) partition of Ω into sets $\{A_i \cap B_j\}$, we can write

$$g = \sum_{i=1}^{r}\sum_{j=1}^{s} x_i I_{A_i \cap B_j}, \qquad h = \sum_{i=1}^{r}\sum_{j=1}^{s} y_j I_{A_i \cap B_j} \tag{6.3.2}$$

where by assumption, $x_i \leq y_j$ on each $A_i \cap B_j$.
So

$$\int g\, d\mu \; = \sum_{i=1}^{r}\sum_{j=1}^{s} x_i\, \mu(A_i \cap B_j) \leq \sum_{i=1}^{r}\sum_{j=1}^{s} y_i\, \mu(A_i \cap B_j) = \int h\, d\mu \;.$$

b) We sum the two forms in (6.3.2) and apply the definition of integral of simple functions to obtain

$$\int (g+h)\, d\mu \; = \sum_{i=1}^{r}\sum_{j=1}^{s} (x_i + y_j)\, \mu(A_i \cap B_j) \tag{6.3.3}$$

But since we have

$$A_i = \dot{\bigcup_j} (A_i \cap B_j), \quad B_j = \dot{\bigcup_i} (A_i \cap B_j)$$

Then by finite additivity

$$\mu(A_i) = \sum_{j=1}^{s} \mu(A_i \cap B_j), \quad \mu(B_j) = \sum_{i=1}^{r} \mu(A_i \cap B_j)$$

And so

$$\int g\, d\mu \; = \sum_{i=1}^{r} x_i \mu(A_i) = \sum_{i=1}^{r} x_i \left(\sum_{j=1}^{s} \mu(A_i \cap B_j) \right)$$
$$\int h\, d\mu \; = \sum_{j=1}^{s} y_i \mu(B_i) = \sum_{j=1}^{s} y_i \left(\sum_{i=1}^{r} \mu(A_i \cap B_j) \right)$$

and summing these together yields (6.3.3).

c) This follows immediately from the distributive property.

$$c \int g\, d\mu \; = c \sum_{i=1}^{r} x_i \mu(A_i) = \sum_{i=1}^{r} c\, x_i \mu(A_i) = \int cg\, d\mu$$

□

Proposition 6.3.2. *Let s be a non-negative simple function. Then*

$$A \mapsto \int_A s \, d\mu$$

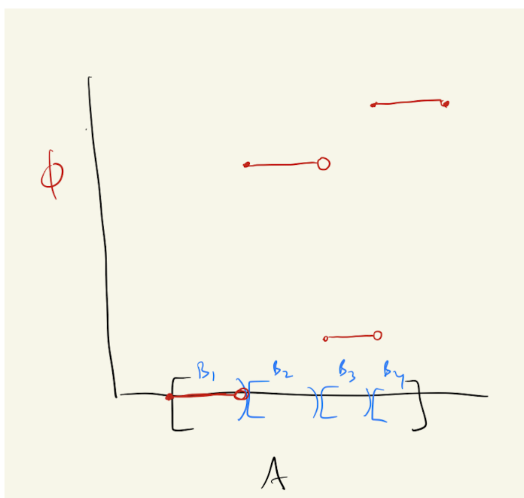
is a measure on $\mathcal{F}$.

Proof. We show countable additivity.[29] Let $A = \dot{\bigcup}_{i=1}^{\infty} A_i$. Then

$$\begin{aligned}
\int_A s \, d\mu &= \sum_{j=1}^{n} y_j \, \mu(B_j \cap A) && \text{Remark 6.3.4} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{\infty} y_i \, \mu(B_j \cap A_i) && \text{countable additivity} \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{n} y_i \, \mu(B_j \cap A_i) && \text{sum of limit is limit of sum} \\
&= \sum_{i=1}^{\infty} \int_{A_i} s \, d\mu
\end{aligned}$$

□

Remark 6.3.6. Proposition 6.3.2 says that we can measure a set by integrating a simple function over it. Given some initial measure for A, we can get a new measure by breaking A into finitely many (measurable) pieces and giving different weights to the measures of the different pieces.



△

Remark 6.3.7. (*Continuity of measure applied to measures given by integrals of simple functions over sets*) Recall (Theorem 3.4.1) that measure satisfies continuity from below: if A_n are measurable and $A_n \uparrow A$, then $\lim_{n\to\infty} \mu(A_n) = \mu(A)$. Combining this with Proposition 6.3.2, we have that if A_n are measurable and $A_n \uparrow A$,

$$\lim_{n\to\infty} \int_{A_n} s \, d\mu = \int_A s \, d\mu$$

for any simple function s. △

[29] Non-negativity is clear from the definition. See Remark 6.3.4.

6.3.2 Integrals of non-negative Borel measurable functions

Definition of integral of non-negative Borel measurable functions

Definition 6.3.3. If h is non-negative Borel measurable, we define

$$\int_\Omega h\, d\mu \;= \sup\left\{ \int_\Omega s\, d\mu \;: s \quad \text{simple}, \quad 0 \le s \le h \right\} \tag{6.3.4}$$

△

When h is simple, this definition agrees with the definition of the integral for simple functions (Definition 6.3.2). This follows from Proposition 6.3.1 (a) and the fact that the family of functions over which the supremum is taken includes h itself.[30]

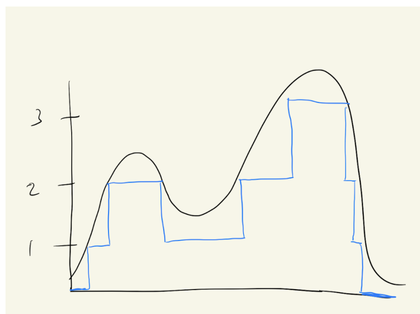


Figure 8: A simple function approximating a non-negative function in terms of its integral

Remark 6.3.8. (*When does the integral of a non-negative Borel measurable function exist?*) The integral of a non-negative Borel measurable function *always* exists (although it may take on the value $+\infty$). Note that neither case discussed in Remark 6.3.3 apply, since the supremum is simply taken over the set of simple functions that never take on negative values. △

Properties of integral of non-negative Borel measurable functions (Part I)

We will use these basic properties below:

Proposition 6.3.3. *Let f, g be non-negative Borel measurable functions. Then*

a) (monotonicity) $\displaystyle\int f \le \int g$ *whenever* $f \le g$

b) (non-negative constant multiples) $\displaystyle\int cf = c\int f$ *for all* $c \ge 0$.

[30]Let us show in more detail that when h is simple, the definition of the integral for non-negative Borel measurable functions (Definition 6.3.3) agrees with the definition of the integral for simple functions (Definition 6.3.2). Let h be simple and take $\int_\Omega h\, d\mu$ to be given by Definition 6.3.2. Then by Proposition 6.3.1 (a), $\int_\Omega h\, d\mu$ must be an upper bound for A, which we define as the set of integrals on the RHS of Definition 6.3.3. Moreover, it is the least upper bound by Remark A.1.4, since for every $M' < M := \int_\Omega h\, d\mu$ there is an $a \in A$ such that $a > M'$, namely $a = \int_\Omega h\, d\mu$.

Proof. a)

$$\begin{aligned}
f \leq g &\implies \{s : s \text{ simple }, 0 \leq s \leq f\} \subset \{s : s \text{ simple }, 0 \leq s \leq g\} \\
&\implies \{\int s \, d\mu \ : s \text{ simple }, 0 \leq s \leq f\} \subset \{\int s \, d\mu \ : s \text{ simple }, 0 \leq s \leq g\} \\
&\overset{Prop.A.2.1}{\implies} \sup\{\int s \, d\mu \ : s \text{ simple }, 0 \leq s \leq f\} \leq \sup\{\int s \, d\mu \ : s \text{ simple }, 0 \leq s \leq g\} \\
&\implies \int f \, d\mu \ \leq \int g \, d\mu
\end{aligned}$$

b)

$$\begin{aligned}
\int cf \, d\mu \ &= \sup\{\int s' \, d\mu \ : s' \text{ simple }, 0 \leq s' \leq cf\} \\
&= \sup\{\int cs \, d\mu \ : s \text{ simple }, 0 \leq s \leq f\} && \text{Let } s' = cs \\
&= \sup\{c \int s \, d\mu \ : s \text{ simple }, 0 \leq s \leq f\} && \text{Linearity of integral of simple functions} \\
&= c\{\int s \, d\mu \ : s \text{ simple }, 0 \leq s \leq f\} && \text{Prop. A.2.2} \\
&= c \int f \, d\mu
\end{aligned}$$

□

Computing the integral of non-negative Borel measurable functions

The following proposition may provide additional confidence in Definition 6.3.3, since simple functions approximate non-negative Borel measurable functions.

Proposition 6.3.4. *Let h be a non-negative Borel measurable function. Then there is a sequence $\{s_n\}$ of simple functions such that $0 \leq s_1 \leq s_2 \leq ... \leq h$, $s_n \to h$ pointwise, and $s_n \to h$ uniformly on any set on which h is bounded.*

Proof. Define

$$s_n(\omega) = \begin{cases} \dfrac{k-1}{2^n}, & \text{if} \quad \dfrac{k-1}{2^n} \leq h(\omega) \leq \dfrac{k}{2^n}, \quad k = 1, 2, ..., n2^n \\ n, & \text{if} \quad h(\omega) \geq n \end{cases}$$

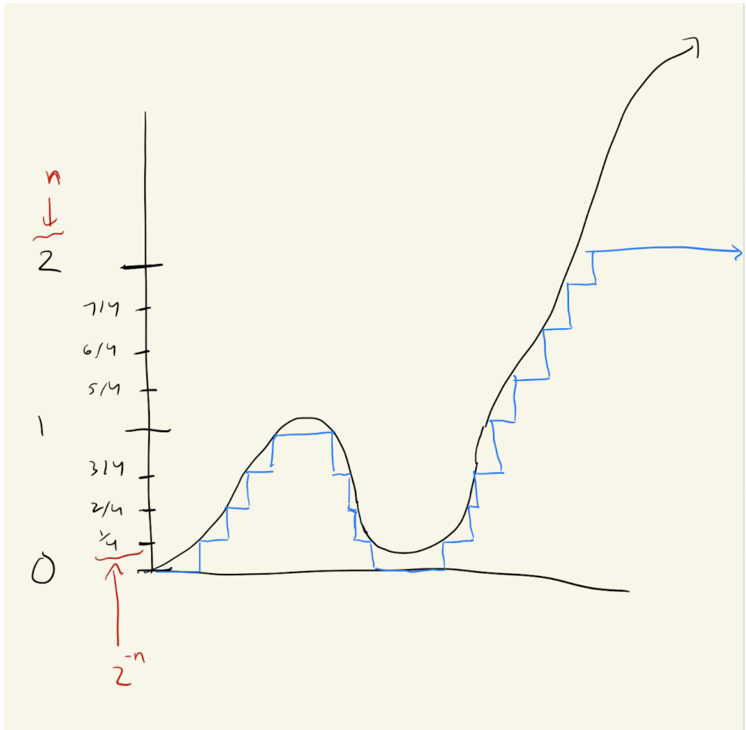


Figure 9: An element of an increasing sequence of simple functions approximating an arbitrary non-negative Borel measurable function. The nth function in the sequence has a maximum value of n and divides the range into bins of size 2^{-n}.

Then

- $s_n \le s_{n+1}$ for all n
- $0 \le h - s_n \le 2^{-n}$ on the set where $h \le n$.

□

Question 6.3.1. When working with $\overline{\mathbb{R}}$, how does one check convergence of a function at a point ω such that $h(\omega) = \infty$? Does the proof of Proposition 6.3.4 still work at such points? △

Now we establish one of the fundamental convergence theorems.

Theorem 6.3.1. ***The Monotone convergence theorem.*** *If $\{h_n\}$ is a sequence of non-negative Borel measurable functions such that $h_n \le h_{n+1}$ for all n and $h = \lim_{n\to\infty} h_n (= \sup_n h_n)$, then*

$$\int h_n \, d\mu \uparrow \int h \, d\mu \tag{6.3.5}$$

Proof. We break the proof into four parts. Consider

$$\int h \, d\mu = \lim_{n\to\infty} \int h_n \, d\mu \tag{6.3.6}$$

- We first show that both quantities in (6.3.6) exist. First, recall from Remark 6.3.8 that the integral of a non-negative Borel measurable function always exists. So the LHS exists because

h is Borel measurable (by Proposition 6.2.2) and non-negative. The RHS exists since each $\int h_n$ exists and is increasing (by monotonicity; see Prop. 6.3.3), and therefore has a limit (possibly ∞).

- Next, we show $\geq$ for (6.3.6); that is we show $\lim_{n\to\infty} \int h_n \leq \int h$.

$$h_n \leq h \overset{\text{monotonicity}}{\Longrightarrow} \int h_n \leq \int h \overset{\text{limits preserve (non-strict) inequalities}}{\Longrightarrow} \lim_{n\to\infty} \int h_n \leq \int h$$

- Now, we show $\leq$ for (6.3.6); that is we show $\lim_{n\to\infty} \int h_n \geq \int h$. Let $\alpha \in (0,1)$ and s be a simple function such that $0 \leq s \leq h$. Now define

$$A_n := \{\omega : h_n(\omega) \geq \alpha s(\omega)\}$$

And note that $A_n \uparrow \Omega$.

So

$$\int h_n \overset{\text{monotonicity}}{\geq} \int_{A_n} h_n \overset{\text{def. } A_n\text{, monotonicity}}{\geq} \int_{A_n} \alpha s \overset{\text{linearity}}{=} \alpha \int_{A_n} s \tag{6.3.7}$$

Now we recognize the right hand side as a measure on A_n, and since $A_n \uparrow \Omega$, we can apply continuity from below (see Remark 6.3.7), so taking the limit as $n \to \infty$, Equation (6.3.7) becomes

$$\lim_{n\to\infty} \int h_n \geq \alpha \int s$$

Now since the equality holds for all $\alpha < 1$, it holds for $\alpha = 1$, and so we have

$$\lim_{n\to\infty} \int h_n \geq \int s$$

Since the LHS is an upper bound on the set in the RHS, it must be greater than the least upper bound, so

$$\lim_{n\to\infty} \int h_n \geq \int h$$

- Now we know that (6.3.6) holds. It just remains to show that (6.3.6) $\implies$ (6.3.5). By monotonicity, $\int h_{n+1}\, d\mu \geq \int h_n\, d\mu$ for all n, and since the limit of an increasing sequence is its supremum (Prop. A.2.4), $\int h\, d\mu \geq \int h_n\, d\mu$ for all n.

□

Remark 6.3.9. (*The monotone convergence theorem aids in computation.*) The monotone convergence theorem can actually be used to make it easier to do computations with integrals of non-negative Borel measurable functions! Let us quote [Folland, 1999] (except with changes of notation and references)

> The definition of $\int h$ involves the supremum over a huge (usually uncountable) family of simple functions, so it may be difficult to evaluate $\int h$ directly from the definition (see Definition 6.3.3). The monotone convergence theorem, however, assures us that to compute $\int h$, it is enough to compute $\lim \int s_n$, where $\{s_n\}$ is any sequence of simple functions that increase to h, and Proposition 6.3.4 guarantees that such sequences exist.

△

Remark 6.3.10. (*Monotone convergence theorem fails for the Riemann integral.*) Using an enumeration of the rational numbers between 0 and 1, we define the function f_n (for all nonnegative integer n) as the indicator function of the set of the first n terms of this sequence of rational numbers. The increasing sequence of functions f_n (which are nonnegative, Riemann-integrable with a vanishing integral) pointwise converges to the Dirichlet function which is not Riemann-integrable (see Remark 6.3.5). △

Properties of integral of non-negative Borel measurable functions (Part II)

With the monotone convergence theorem in hand, we can now provide some additional properties of the integrals of non-negative Borel measurable functions.

Proposition 6.3.5. *Let f, g be non-negative Borel measurable functions. Then*

$$\int f + \int g = \int f + g$$

Proof. By Proposition 6.3.4, there are sequences of simple functions $\{s_n\}, \{t_n\}$ such that $s_n \uparrow f$ and $t_n \uparrow g$. Thus, by limit properties, $(s_n + t_n) \uparrow (f + g)$. So we have

$$\begin{aligned}
\int f + \int g &= \lim_{n\to\infty} \int s_n + \lim_{n\to\infty} \int t_n && \text{Monotone convergence theorem} \\
&= \lim_{n\to\infty} \Big(\int s_n + \int t_n\Big) \\
&= \lim_{n\to\infty} \int (s_n + t_n) && \text{Linearity of integral for simple functions} \\
&= \int (f + g) && \text{Monotone convergence theorem}
\end{aligned}$$

□

Below we show that Proposition 6.3.5 actually extends to linearity with countably infinite sums.

Corollary 6.3.1. *(Linearity with countably infinite sums of non-negative measurable functions) If $\{f_n\}$ is a sequence of non-negative measurable functions, and*

$$f(\omega) = \sum_{n=1}^{\infty} f_n(\omega), \qquad \textit{for all } \omega \in \Omega$$

Then

$$\int_{\Omega} f \, d\mu = \sum_{n=1}^{\infty} \int_{\Omega} f_n \, d\mu \tag{6.3.8}$$

Thus, any series of non-negative Borel measurable functions may be integrated term by term.

Proof. By induction, the linearity of Proposition 6.3.5 extends to a finite collection $\{f_n\}_{n=1}^{N}$. Now $\sum_{n=1}^{N} f_n \uparrow \sum_{n=1}^{\infty} f_n$, so we apply Monotone Convergence Theorem (MCT)

$$\underbrace{\lim_{N\to\infty} \int \sum_{n=1}^{N} f_n \, d\mu}_{\overset{\text{(additivity)}}{=} \sum_{n=1}^{\infty} \int f_n \, d\mu} \overset{\text{(MCT)}}{=} \int \sum_{n=1}^{\infty} f_n \, d\mu$$

□

Remark 6.3.11. If we let μ be the counting measure on a countable set, the statement of Corollary 6.3.1 becomes a statement about double series of non-negative real numbers (which can be proved by more elementary means) [Rudin, 1987]:

That is, if $a_{ij} \geq 0$ for i and $j = 1, 2, 3, ...$, then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}$$

More explicitly, on $\Omega = \{1, 2, 3, \ldots\}$, we define sequence $\{f_i\}$ such that $f_i(j) = a_{ij}$. Then we define $f = \sum_{i=1}^{\infty} f_i$, so $f(j) = \sum_{i=1}^{\infty} f_i(j) = \sum_{i=1}^{\infty} a_{ij}$.

Then, since μ is the counting measure, the LHS of (6.3.8) becomes

$$\int_{\Omega} f(j)\, d\mu\,(j) = \sum_{j=1}^{\infty} f(j) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}$$

and the RHS of (6.3.8) is

$$\sum_{i=1}^{\infty} \int_{\Omega} f_i\, d\mu = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij}.$$

△

6.3.3 Integrals of arbitrary Borel measurable functions

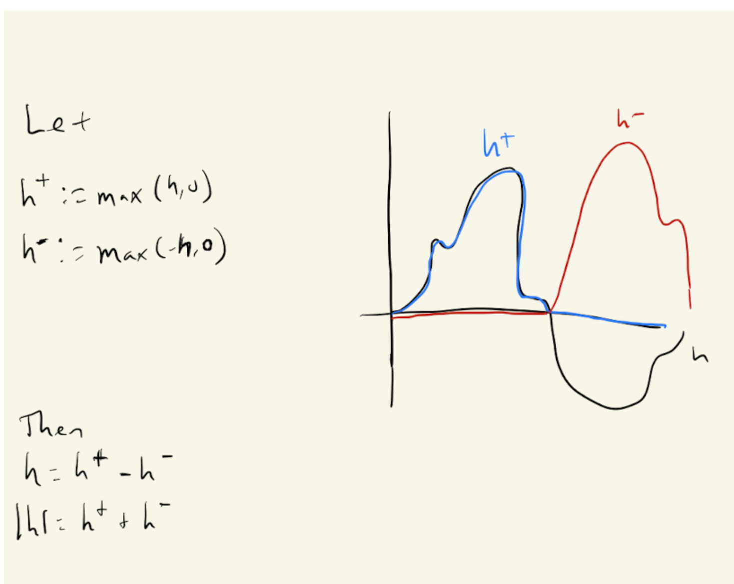
Let h be an arbitrary Borel measurable function. We will express an arbitrary Borel measurable function as as the difference of two non-negative Borel measurable functions.

Define:

$$h^+ := \max(h, 0)$$
$$h^- := \max(-h, 0)$$

Then

$$h = h^+ - h^-$$
$$|h| = h^+ + h^-$$



Now both h^+ and h^- are Borel measurable as well. This holds because if h_1, h_2 are Borel measurable, then so are $\max(h_1, h_2)$ and $\min(h_1, h_2)$:

$$\left\{\omega : \max\Big(h_1(\omega), h_2(\omega)\Big) < c\right\} = \{\omega : h_1(\omega) < c\} \cap \{\omega : h_2(\omega) < c\}$$
$$\left\{\omega : \min\Big(h_1(\omega), h_2(\omega)\Big) < c\right\} = \{\omega : h_1(\omega) < c\} \cup \{\omega : h_2(\omega) < c\}$$

which is sufficient to show measurability by the "computational definition of measurability" (see Remark 6.2.1).

So we have expressed an arbitrary Borel measurable function as as the difference of two non-negative Borel measurable functions. Therefore, we can define its integral as follows.

Definition 6.3.4. (*Integral of an arbitrary Borel measurable function*)

$$\int_\Omega h \, d\mu = \int_\Omega h^+ \, d\mu - \int_\Omega h^- \, d\mu \tag{6.3.9}$$

△

Remark 6.3.12. (*When does the integral of an arbitary Borel measurable function exist?*) Recall from Remark 6.3.8 that the integral of a non-negative Borel measurable function *always* exists (although it may take on the value $+\infty$). Thus, the integral of an arbitrary non-negative Borel function exists so long as it does not take the form $+\infty - \infty$. △

Definition 6.3.5. (*Integrable and extended integrable functions.*) We say that a function h is μ-integrable (or just integrable if μ is understood) if $\displaystyle\int_\Omega h \, d\mu$ is finite, that is, iff $\displaystyle\int_\Omega h^+ \, d\mu$ and $\displaystyle\int_\Omega h^- \, d\mu$ are both finite. Following [Folland, 1999, pp. 86], we say that a function h is extended μ-integrable iff at least one of $\displaystyle\int_\Omega h^+ \, d\mu$ and $\displaystyle\int_\Omega h^- \, d\mu$ are finite (which means that the integral $\displaystyle\int_\Omega h \, d\mu$ exists). △

Remark 6.3.13. (*Integrals on subsets*) For $A \in \mathcal{F}$, we define

$$\int_A h \, d\mu = \int_\Omega h I_A \, d\mu$$

This definition works because whenever h is measurable, then so is hI_A:

$$\{\omega : hI_A(\omega) < c\} = \underbrace{\{\omega : h(\omega) < c\}}_{\in \mathcal{F} \text{ since } h \text{ measurable}} \cap \underbrace{\{\omega : \omega \in A\}}_{\in \mathcal{F} \text{ by assumption}}$$

△

Properties of the integral of arbitrary Borel measurable functions

Proposition 6.3.6. *Let f, g, h arbitrary Borel measurable functions. Then*

a) (scalar multiple) If $\int f \, d\mu$ exists and $c \in \mathbb{R}$, then $\int cf$ exists and $\int cf \, d\mu = c \int f \, d\mu$.

b) *(monotonicity) If* $g \geq h$ *and both integrals exist, then* $\int g \, d\mu \geq \int h \, d\mu$ *. Moreover, if* $g \geq h$*,* $\int h \, d\mu$ *exists and* $\int h \, d\mu > -\infty$*, then* $\int g \, d\mu$ *exists. And if* $g \geq h$*,* $\int g \, d\mu$ *exists and* $\int g \, d\mu < \infty$*, then* $\int h \, d\mu$ *exists.*[31]

c) *(delayed truncation of simple functions) If* $h \geq 0$ *and* $B \in \mathcal{F}$*, then*[32]

$$\int_B h \, d\mu = \sup \left\{ \int_B s \, d\mu \ : \ 0 \leq s \leq h, \ s \text{ simple} \right\}$$

d) *(existence of integral transfers to subsets)*

$$\int_\Omega h \, d\mu \text{ exists} \implies \int_A h \, d\mu \text{ exists} \quad \forall A \in \mathcal{F} \tag{6.3.10}$$

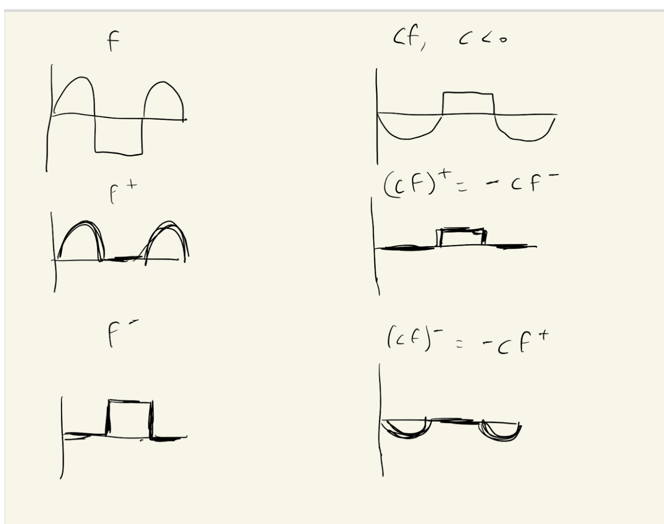
$$\int_\Omega h \, d\mu \text{ finite} \implies \int_A h \, d\mu \text{ finite} \quad \forall A \in \mathcal{F} \tag{6.3.11}$$

Proof. a) Since cf is a Borel measurable function[33], we apply Definition 6.3.4.

We have

$$\text{if } c \geq 0, \qquad (cf)^+ = cf^+ \qquad (cf)^- = cf^- \tag{6.3.12a}$$

$$\text{if } c < 0, \qquad (cf)^+ = -cf^- \qquad (cf)^- = -cf^+. \tag{6.3.12b}$$



Now we will use the fact that if f is non-negative Borel measurable and $c \geq 0$, then we already know the identity holds (see Prop. 6.3.3 (b)).

[31] We might consider this as a "dominance criterion for existence." For more on why the monotonicity statement is concerned with existence, see Remark 6.3.14.

[32] For why this needs to be proven, see Remark 6.3.15.

[33] The function cf is a Borel measurable fucntion by Proposition 6.2.1 and Example 6.2.1. Alternatively, we could verify this directly. If $c \geq 0$, then $\{\omega : cf(\omega) \leq k\} = \{\omega : f(\omega) \leq k/c\} \in \mathcal{F}$ by the Borel measurability of f. Similarly if $c < 0$, then $\{\omega : cf(\omega) \leq k\} = \{\omega : f(\omega) \geq k/c\} \in \mathcal{F}$ by the Borel measurability of f.

So if $c \geq 0$

$$\begin{aligned}
\int cf\, d\mu &= \int (cf)^+\, d\mu - \int (cf)^-\, d\mu && \text{Def. 6.3.4} \\
&= \int cf^+\, d\mu - \int cf^-\, d\mu && \text{(6.3.12a)} \\
&\overset{*}{=} c\int f^+\, d\mu - c\int f^-\, d\mu && \text{Prop. 6.3.3 (b)} \\
&= c\int f\, d\mu && \text{Def. 6.3.4}
\end{aligned}$$

Likewise if $c < 0$

$$\begin{aligned}
\int cf\, d\mu &= \int (cf)^+\, d\mu - \int (cf)^-\, d\mu && \text{Def. 6.3.4} \\
&= \int -cf^-\, d\mu - \int -cf^+\, d\mu && \text{(6.3.12b)} \\
&\overset{**}{=} -c\int f^-\, d\mu + c\int f^+\, d\mu && \text{Prop. 6.3.3 (b)} \\
&= c\int f\, d\mu && \text{Def. 6.3.4}
\end{aligned}$$

Equations (*) and (**) reveal that $\int cf\, d\mu$ exists whenever $\int f\, d\mu$ exists.

b) First we show that $g \geq h \implies \int g\, d\mu \geq \int h\, d\mu$ when both integrals exist. We decompose each function into its positive and negative parts

$$g = g^+ - g^-, \quad h = h^+ - h^-.$$

By hypothesis,

$$g^+ \geq h^+, \quad g^- \leq h^-.$$

So by monotonicity for non-negative functions (Prop. 6.3.3 (a)), we have

$$\int g^+\, d\mu \geq \int h^+\, d\mu\,, \quad \int g^-\, d\mu \leq \int h^-\, d\mu\,. \tag{6.3.13}$$

So

$$\begin{aligned}
\int g\, d\mu &= \int g^+\, d\mu - \int g^-\, d\mu && \text{(def. integral; existence assumed)} \\
&\geq \int h^+\, d\mu - \int h^-\, d\mu && \text{(6.3.13)} \\
&= \int g\, d\mu && \text{(def. integral; existence assumed)}
\end{aligned}$$

Now we consider the "dominance criterion for existence". We prove the second sentence of (b), as the third is proved similarly.

If $\int h\, d\mu$ exists and $\int h\, d\mu > -\infty$, then by definition of the integral, $\int h^-\, d\mu < \infty$. Since $g \geq h$, then $g^- \leq h^-$, so

$$\int g^-\, d\mu \leq \int h^-\, d\mu < \infty$$

Thus, $\int g \, d\mu$ exists.[34]

c) We want to prove that if $h \geq 0$ and $B \in \mathcal{F}$, then

$$\int_B h \, d\mu = \sup\left\{ \int_B s \, d\mu \ : \ 0 \leq s \leq h, \ s \text{ simple} \right\}.$$

We prove $\boxed{\geq}$, $\boxed{\leq}$ separately, using the strategy of Remark A.1.5.

- $\boxed{\geq}$. For $0 \leq s \leq h$, s simple,

$$\int_B h \, d\mu \geq \int_B s \, d\mu \qquad \text{monotonicity}$$

Since the LHS is an upper bound on the set of the integrals on the RHS, $\boxed{\geq}$ holds.

- $\boxed{\leq}$

$$\{t : t \text{ simple}, 0 \leq t \leq h1_B\} \subseteq \{s1_B : s \text{ simple}, 0 \leq s \leq h\}$$

$$\implies \underbrace{\sup\left\{ \int t \, d\mu \ : \ t \text{ simple}, \ 0 \leq t \leq h1_B \right\}}_{:= \int_B h \, d\mu} \leq \sup\left\{ \int s1_B \, d\mu \ : \ s \text{ simple}, \ 0 \leq s \leq h \right\}$$

d)

$$(h1_A)^+ = h^+ 1_A \leq h^+, \quad (h1_A)^- = h^- 1_A \leq h^-$$

So by monotonicity,

$$\underbrace{\int (h1_A)^+ \, d\mu}_{\text{A1}} \leq \underbrace{\int h^+ \, d\mu}_{\text{B1}}$$

$$\underbrace{\int (h1_A)^- \, d\mu}_{\text{A2}} \leq \underbrace{\int h^- \, d\mu}_{\text{B1}}$$

So $B_i < \infty \implies A_i < \infty$.

By assuming the conditional holds for at least one $i \in \{1, 2\}$, we prove transfer of existence. By assuming the conditional holds for both i, we prove transfer of finiteness.

□

Remark 6.3.14. (*Why the monotonicity property is concerned with existence*) Why is Proposition 6.3.6 b) concerned with monotonicity? Answer: even if $\int g \, d\mu$ exists and $g \geq h$, we can still have $\int h \, d\mu$ not exist, because of (6.3.13). For example, we can have

$$\int g^+ \, d\mu = \int h^+ \, d\mu = \infty \tag{6.3.14}$$

$$\int g^- \, d\mu < \int h^- \, d\mu = \infty \tag{6.3.15}$$

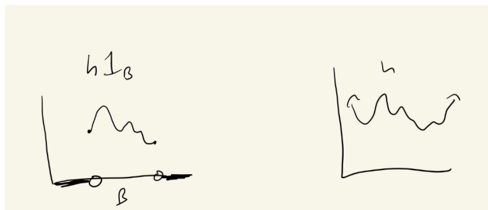
and so $\int h \, d\mu$ DNE. △

[34] Recall that for $\int f \, d\mu$ to exist, at least one of $\int f^- \, d\mu$, $\int f^+ \, d\mu$ must be finite.

Remark 6.3.15. (*Why delayed truncation of simple functions is something that needs to be proven*) Proposition 6.3.6 c) needs to be proven because it is *not* what is given by the definition of the integral for an arbitrary Borel measurable function (after observing that $h1_B$ is still measurable). Note that

$$\int_\Omega h\, d\mu \ = \sup\left\{\int s\, d\mu \ :\ 0 \le s \le h1_B,\ s \text{ simple}\right\} \qquad \text{def. integral} \qquad (6.3.16a)$$

$$\int_\Omega h\, d\mu \ = \sup\left\{\int s1_B\, d\mu \ :\ 0 \le s \le h,\ s \text{ simple}\right\} \qquad \text{Prop. 6.3.6 c).} \qquad (6.3.16b)$$



What does each say?

- (6.3.16a) : truncate first, then “simplify"
- (6.3.16b) “simplify" first, then truncate

. △

7 § 1.6 Basic Integration Theorems

Here we prove some basic integration theorems, and further properties of integration that can be derived thereof.

Definition 7.0.1. We say that λ is an *indefinite integral*[35] with respect to μ if for any $A \in \mathcal{F}$ we have

$$\lambda(A) = \int_A f\, d\mu \qquad (7.0.1)$$

where f is a Borel measurable function and where $\int_\Omega f\, d\mu$ exists. △

Theorem 7.0.1. ***Indefinite integrals are countably additive set functions.*** *Let f be a Borel measurable function such that $\int_\Omega f\, d\mu$ exists. Define $\lambda(B) = \int_B f\, d\mu$, $B \in \mathcal{F}$. Then λ is countably additive on $\mathcal{F}$; thus if $f \ge 0$, λ is a measure.*

Proof. Recall that Prop. 6.3.2 proved this for non-negative simple functions.

So let f be any non-negative Borel measurable function. We want to show that if $\lambda(B) = \int_B f\, d\mu$, $B = \dot\bigcup_{n=1}^\infty B_n$, then $\lambda(B) = \displaystyle\sum_{n=1}^\infty \lambda(B_n)$.

[35]This interpretation of “indefinite integral" is used in [Ash et al., 2000, pp. 61]

- $\boxed{\leq}$ Let s be simple, $0 \leq s \leq f$. Then

$$\begin{aligned} \int_B s \, d\mu &= \sum_{n=1}^{\infty} \int_{B_n} s \, d\mu && \text{Prop. 6.3.2} \\ &\leq \sum_{n=1}^{\infty} \int_{B_n} f \, d\mu && \text{monotonicity} \\ &:= \sum_{n=1}^{\infty} \lambda(B_n) \end{aligned}$$

Since the RHS is an upper bound, the supremum (over s) cannot exceed it. Thus, applying Prop. 6.3.6 c), we have

$$\begin{aligned} \int_B f \, d\mu &\leq \sum_{n=1}^{\infty} \lambda(B_n) \\ \implies \lambda(B) &\leq \sum_{n=1}^{\infty} \lambda(B_n) && \text{def. } \lambda \end{aligned}$$

- $\boxed{\geq}$ By monotonicity of the integral,

$$B \supset B_n \implies 1_B \geq 1_{B_n} \implies f 1_B \geq f 1_{B_n} \implies \lambda(B) \geq \lambda(B_n) \qquad (1)$$

If $\lambda(B_n) = \infty$ for one n, we are done. (Why? Since $f \geq 0$, by monotonicity, $\int_A f \, d\mu \geq 0$ for any $A \in \mathcal{F}$. So each $\lambda(B_n) \geq 0$. So if one $\lambda(B_n) = \infty$, then $\sum_{n=1}^{\infty} \lambda(B_n) = \infty$, and (1) and $\lambda(B) \geq \sum_{n=1}^{\infty} \lambda(B_n)$ are saying the same thing, that $\lambda(B) = \infty$.)

So let each $\lambda(B_n) < \infty$.

Fix N. Consider $\dot{\bigcup}_{n=1}^{N} B_n$. By Prop. 6.3.6 c) and properties of the supremum (if we subtract ϵ from it, there $\exists$ a member of the set exceeding that), for all $\epsilon > 0$, we have simple functions $s_n : 0 \leq s_n \leq f$ for each n so that

$$\int_{B_n} s_n \, d\mu \geq \int_{B_n} f \, d\mu - \frac{\epsilon}{N} \quad \text{for all } n \qquad (2)$$

Let s^* be the pointwise maximum of $\{s_n\}_{n=1}^{N}$. This is still a simple function, and $s^* 1_{B_n} \geq s_n 1_{B_n}$ for each n, so

$$\int_{B_n} s^* \, d\mu \geq \int_{B_n} s_n \quad \text{for all } n \qquad (3)$$

So (2) and (3) gives that

$$\int_{B_n} s^* \, d\mu \geq \int_{B_n} f \, d\mu - \frac{\epsilon}{N} \quad \text{for all } n \qquad (4)$$

Thus, for any $N, \epsilon > 0$, we have

$$
\begin{aligned}
\lambda(B) &\geq \lambda\Big(\dot{\bigcup_{n=1}^{N}} B_n\Big) && \text{By monotonicity; need argument like } 1\text{; we don't know it's a measure yet!} \\
&:= \int_{\dot\bigcup_{n=1}^{N} B_n} f \, d\mu \\
&\geq \int_{\dot\bigcup_{n=1}^{N} B_n} s^* \, d\mu && \text{monotonicity} \\
&= \sum_{n=1}^{N} \int_{B_n} s^* \, d\mu && \text{what's we're trying to prove holds for simple functions (Prop. 6.3.2)} \\
&\geq \sum_{n=1}^{N} \int_{B_n} f_n \, d\mu \; - \epsilon && \text{see } 4 \\
&:= \sum_{n=1}^{N} \lambda(B_n) - \epsilon \\
\implies \lambda(B) &\geq \sum_{n=1}^{\infty} \lambda(B_n) && \text{justified below}
\end{aligned}
$$

What justifies the last line above? **Claim.** Let $\{a_n\}_{n=1}^{\infty} : 0 \leq a_n < \infty$. Then $M \overset{*}{\geq} \sum_{n=1}^{N} a_n - \epsilon$ for any $N, \epsilon > 0 \implies M \geq \sum_{n=1}^{\infty} a_n$. **Proof.** If $\sum_{n=1}^{\infty} a_n = \infty$ then the claim obviously holds. If $\sum_{n=1}^{\infty} a_n < \infty$ then for all $\epsilon > 0$, the tail of the series is less than ϵ for some N^*. So write $\sum_{n=1}^{N^*} a_n < \sum_{n=1}^{\infty} a_n - \epsilon$, and equation (*) becomes $M \geq \sum_{n=1}^{\infty} a_n - 2\epsilon = \sum_{n=1}^{\infty} a_n - \tilde{\epsilon}$ for all $\tilde{\epsilon} > 0$. Take the limit as $\tilde{\epsilon} \to 0$, and the non-strict inequality is preserved in the limit.

So now assume f is an arbitrary Borel-measurable function. Since we have assumed $\int f \, d\mu$ exists, we have $\int f^+ \, d\mu \, , \int f^- \, d\mu \; < \infty$. So by what we have shown for non-negative functions, there exists measures λ^+, λ^- corresponding to each of these integrals. So if $B = \dot\bigcup_{n=1}^{\infty} B_n$,

$$
\begin{aligned}
&\int_B f \, d\mu \; = \int_B f^+ \, d\mu \; - \int_B f^- \, d\mu && \text{def. integral} \\
\implies\; &\lambda(B) = \lambda^+(B) - \lambda^-(B) && \text{by def. of } \lambda, \lambda^+, \lambda^- \\
\implies\; &\lambda(B) = \sum_{n=1}^{\infty} \lambda^+(B_n) - \sum_{n=1}^{\infty} \lambda^-(B_n) && \text{by result with non-negative functions}
\end{aligned}
$$

and this expression is *NOT* of the form $\infty - \infty$, since the first line isn't, by the existence of $\int_\Omega f \, d\mu$ (and the fact that, by monotonicity, $\int_\Omega f \, d\mu \; < \infty \implies \int_B f \, d\mu \; < \infty$).

□

Corollary 7.0.1. ***Indefinite integrals as measures.*** *Let $f \geq 0$ be a Borel measurable function such that $\int_\Omega f \, d\mu$ exists. Define $\lambda(B) = \int_B f \, d\mu$, $B \in \mathcal{F}$. Then λ is a measure.*

Proof. We apply Definition 3.2.1 to show that λ is a measure. λ is countably additive by Theorem 7.0.1. The non-negativity of λ is immediate from Definition 6.3.3. □

Remark 7.0.1. Indefinite integrals are signed measures. As will become clear in Section 9, Theorem 7.0.1 more generally tells us that *indefinite integrals are signed measures*. If we remove the constraint that $f \geq 0$, then λ is a countably additive set function, but it may be negative.

The Radon-Nikodym theorem (to be covered later in the document) provides an important converse: instead of obtaining a signed measure λ from a measure μ and function f, we will be given signed measure λ and measure μ, and will obtain the *Radon-Nikodym derivative* f. $\triangle$

Remark 7.0.2. Change of measure and differential notation.

By Cor. 7.0.1, given Borel measurable $f \geq 0$, the indefinite integral

$$\lambda(A) = \int_A f \, d\mu \tag{7.0.2}$$

can be interpreted as a change in measure specifically, as a change from measure μ to measure λ.

To express this relationship, we sometimes use the following notation [Folland, 1999, pp. 89]:

$$d\lambda = f \, d\mu \tag{7.0.3}$$

And sometimes, by a slight abuse of language, we refer to "the measure $f\,d\mu$".

The notation may make more sense if we interpret it, as does [Rudin, 1987, pp. 24]:

$$\int_\Omega g \, d\lambda \; = \int_\Omega gf \, d\mu \tag{7.0.4}$$

for every measurable function g on Ω.[36] (See Proposition 7.0.1 for a proof.)

As pointed out by [Rudin, 1987, pp. 24], we assign no independent meaning to the symbols $d\lambda$ and $d\mu$; (7.0.3) simply means that (7.0.2) (and therefore (7.0.4)) holds for every measurable $f \geq 0$.

$\triangle$

Theorem 7.0.2. ***Additivity theorem.*** *Let f and g be Borel measurable, and assume that $f+g$ is well-defined. If $\int_\Omega f \, d\mu$ and $\int_\Omega g \, d\mu$ exist and $\int_\Omega f \, d\mu + \int_\Omega g \, d\mu$ is well-defined (not of the form $+\infty - \infty$ or $-\infty + \infty$), then*

$$\int_\Omega f + g \, d\mu \; = \int_\Omega f \, d\mu \; + \int_\Omega f \, d\mu$$

In particular, if f and g are integrable, so is $f+g$.

Proof. See [Ash et al., 2000], Theorem 1.6.3. $\square$

Remark 7.0.3. (*Additivity holds automatically for integrable functions.*) If f and g are integrable, the conditions of Theorem 7.0.2 are always met.

Moreover, in this situation, the proof of Theorem 7.0.2 is straightforward. Suppose f and g are integrable. Let $h = f + g$. Then

$$h^+ - h^- = f^+ - f^- + g^+ - g^-$$

Rearranging, we have

$$h^+ + f^- + g^- = f^+ + g^+ + h^-$$

Applying additivity for non-negative functions (see Prop. 6.3.5) twice, we get

$$\int h^+ + \int f^- + \int g^- = \int f^+ + \int g^+ + \int h^-$$

[36] Note that if also $g \geq 0$, this constructs yet another measure by $\xi(A) = \int_A g \, d\lambda$ for all $A \in \mathcal{F}$.

Rearranging (possible by integrability), we get

$$\int h^+ - \int h^- = \int f^+ - \int f^- + \int g^+ - \int g^-$$
$$\overset{\text{def. integral, def. h}}{\Longrightarrow} \int f + g = \int f + \int g$$

$\triangle$

Remark 7.0.4. The conditions of the additivity theorem imply the conditions of the scalar multiple property (Prop 6.3.6 (a)). Thus, linearity holds whenever additivity holds. $\triangle$

Proposition 7.0.1. *(Change of differential.)*[37] *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $f \geq 0$ a non-negative Borel measurable function on Ω. Recalling Cor. 7.0.1, define a measure λ on $\mathcal{F}$ by*

$$\lambda(A) = \int_A f \, d\mu$$

Then for any Borel measurable function g on Ω, we have

$$\int_\Omega g \, d\lambda \; = \int_\Omega gf \, d\mu$$

in the sense that if one of the integrals exists, so does the other, and the two integrals are equal.

Proof. We proceed through the steps in constructing the integral.

a) *Simple functions.* First let g be a simple function, which we write as $s = \sum_{i=1}^r x_i 1_{E_i}$. Then

$$\int s \, d\lambda \overset{\int \text{ for simple f'n}}{=} \sum_{i=1}^r x_i \lambda(E_i) \overset{\text{hypothesis}}{=} \sum_{i=1}^r x_i \int_{E_i} f \, d\mu \overset{\text{linearity}}{=} \int \sum_{i=1}^r x_i 1_{E_i} f \, d\mu \overset{\text{def. } s}{=} \int sf \, d\mu$$

b) *Non-negative Borel measurable functions.* Now let g be a non-negative Borel measurable function. By Prop 6.3.4, there exists a sequence of simple functions $\{s_n\}$ such that $s_n \uparrow g$. Since $s_n \uparrow g$, then also $s_n f \uparrow fg$. So applying Monotone Convergence Theorem to both, we obtain $\int s_n d\lambda \uparrow \int g d\lambda$ and $\int s_n f d\mu \uparrow \int fg d\mu$. But since $\int s_n d\lambda = \int s_n f d\mu$ for all n by the previous bullet point, the sequences must have the same limit (by uniqueness of limits), so $\int g d\lambda = \int gf d\mu$.

c) *Arbitrary Borel measurable functions.* Now let g be an arbitrary Borel measurable function.

$$\int g \, d\lambda \overset{\int \text{ for general f'n}}{=} \int g^+ \, d\lambda \; - \int g^- \, d\lambda \overset{\text{part (b)}}{=} \int g^+ f \, d\mu \; - \int g^- f \, d\mu$$
$$\overset{\text{Additivity Thm}}{=} \int (g^+ - g^-) f \, d\mu \overset{\text{def. } g^+, g^-}{=} \int gf \, d\mu$$

which holds if $\int g \, d\lambda$ exists. In that case, $\int g^+ \, d\lambda \; - \int g^- \, d\lambda$ is well-defined, and so the Additivity Theorem (Thm. 7.0.2) can be applied.

$\square$

Corollary 7.0.2. *(Additivity corollaries)*

a) If h is Borel measurable, h is integrable iff $|h|$ is integrable.

b) If g and h are Borel measurable with $|g| \leq h$, h integrable, then g is integrable.

[37] This is Exercise 4 from [Ash et al., 2000, pp. 71].

Proof. a) If h is integrable, then by assumption we have

$$\left|\int h\, d\mu\right| \overset{\text{def. integral}}{=} \left|\int h^+\, d\mu - \int h^-\, d\mu\right| < \infty$$

which is true iff BOTH of $\left\{\int h^+\, d\mu\, , \int h^-\, d\mu\right\} < \infty$

If $|h|$ is integrable, then by assumption we have

$$\left|\int |h|\, d\mu\right| \overset{\text{additivity (Theorem 7.0.2)}}{=} \left|\int h^+\, d\mu + \int h^-\, d\mu\right| < \infty$$

which is also true iff BOTH of $\left\{\int h^+\, d\mu\, , \int h^-\, d\mu\right\} < \infty$

b)

$$\begin{aligned} &\int h\, d\mu < \infty && \text{by hypothesis} \\ \implies &\int |g|\, d\mu < \infty && \text{by monotonicity} \\ &\implies g \text{ integrable} && \text{by item b) above} \end{aligned}$$

□

Definition 7.0.2. **Almost everywhere** A condition is said to hold *almost everywhere* with respect to the measure μ (written a.e $[\mu]$ or simply a.e. if μ is understood) if there exists a set $B \in \mathcal{F}$ of μ-measure 0 such that the condition holds outside B. △

From the point of view of integration theory, functions that differ only on a set of measure zero may be identified, as is established by the following result.

Theorem 7.0.3. ***Almost everywhere***. *Let f, g, h be Borel measurable functions.*

a) If $f = 0$ a.e. $[\mu]$, then $\int_\Omega f\, d\mu = 0$.

b) If $g = h$ a.e. $[\mu]$, and $\int_\Omega g\, d\mu$ exists, then so does $\int_\Omega h\, d\mu$, and $\int_\Omega g\, d\mu = \int_\Omega h\, d\mu$.

Proof. a) i) $\underline{f \text{ simple}}$. If f is simple, we can write $f = \sum_{i=1}^n x_i 1_{A_i}$. By hypothesis, $\forall$ i, $x_i = 0$ or $\mu(A_i) = 0$. Thus, $\int f\, d\mu = \sum_{i=1}^n x_i 1_{A_i} = 0$.

ii) $\underline{f \text{ non-negative}}$. Since $f = 0$ a.e. $[\mu]$, then $\forall\, s \in \{s \text{ simple } : 0 \le s \le f\}, s = 0$ a.e. $[\mu]$. So by item i), $\int s\, d\mu = 0 \forall s$. So by definition of the integral for non-negative functions

$$\int f\, d\mu = \sup\{\int s\, d\mu : s \text{ simple }, 0 \le s \le f\} = \sup\{0\} = 0$$

iii) $\underline{f \text{ arbitrary}}$. $f = 0$ a.e. $[\mu] \implies f^+ = 0, f^- = 0$ a.e. $[\mu]$. So by item ii), $\int f^+\, d\mu = 0, \int f^-\, d\mu = 0$.. So by definition of the integral $\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu = 0$.

b) We prove i) $\int h\, d\mu$ exists and then that ii) $\int h\, d\mu = \int g\, d\mu$

i) $\int g\, d\mu$ exists means that $\int g^+\, d\mu$, $\int g^-\, d\mu$ are not BOTH ∞. WLOG, suppose that $\int g^+\, d\mu < \infty$ 1.

Now $h = g$ a.e. $\implies$ $h^+ = g^+, h^- = g^-$ a.e. $\implies$ $h^+ - g^+ = 0$ a.e. 2. So

$$0 \overset{\text{by part (a) and } 2}{=} \int (h^+ - g^+)\, d\mu \overset{\text{linearity}}{=} \int h^+\, d\mu - \int g^+\, d\mu \qquad 3$$

where we can apply linearity (see Remark 7.0.4) because

- Integrals of non-negative functions always exist, and multiplication by a scalar doesn't change existence (see Prop. 6.3.6 a)).
- The difference can't be of the form $\infty - \infty$ by 1.

So again by 1, we can add to sides of 3 to get

$$\int h^+ = \int g^+ < \infty$$

so $\int h\, d\mu$ exists.

ii) Let $A := \{\omega : h(\omega) = g(\omega)\}$. By hypothesis, $\mu(A^c) = 0$. Now we decompose each function by partitioning their domains

$$h = h1_A + h1_{A^c} \overset{\text{def. } A}{=} g1_A + h1_{A^c} \tag{7.0.5}$$

$$g = g1_A + g1_{A^c} \tag{7.0.6}$$

Now since $g1_{A^c}, h1_{A^c}$ equal 0 except on a set of measure 0, by part (a),

$$\int_{A^c} g\, d\mu = 0, \quad \int_{A^c} h\, d\mu = 0 \tag{7.0.7}$$

And so we can apply additivity to (7.0.5) and (7.0.6), since:

- $\int g\, d\mu$, $\int h\, d\mu$ exist, so since existence transfers to subsets (see Prop. 5.3.1d)), $\int_A g\, d\mu$, $\int_{A^c} g\, d\mu$, $\int_A h\, d\mu$, $\int_{A^c} h\, d\mu$ exist.
- By (7.0.7),

$$\int_A g\, d\mu + \int_{A^c} g\, d\mu \neq \infty - \infty$$
$$\int_A h\, d\mu + \int_{A^c} h\, d\mu \neq \infty - \infty$$

So applying linearity to (7.0.5) and (7.0.6), we get

$$\int h\, d\mu = \int_A g\, d\mu + \cancelto{0}{\int_{A^c} h\, d\mu} \qquad \text{cancelation by (7.0.7).}$$

$$\int g\, d\mu = \int_A g\, d\mu + \cancelto{0}{\int_{A^c} g\, d\mu} \qquad \text{cancelation by (7.0.7).}$$

And so $\int h\, d\mu = \int g\, d\mu$

□

Remark 7.0.5. Thanks to Theorem 7.0.3, in any integration theorem, we may freely use the phrase "almost everywhere" in the hypotheses, and the conclusions will still follow. For example

- If g, h are Borel measurable and $g \geq h$ a.e., then $\int g \, d\mu \geq \int h \, d\mu$. (This is the monotonicity property from Prop. 6.3.6 b), but with the condition weakened to a.e.).
- If $\{h_n\}$ is a sequence of non-negative Borel measurable functions such that $h_n \to h$ a.e., then $\int_\Omega h_n \, d\mu \to \int_\Omega h \, d\mu$. (This is the Monotone Convergence Theorem but with the condition weakened to a.e. In more detail: we can simply define h_n^* such that it equals h_n almost everywhere and h on the set of measure 0. Then $h_n^* \to h$. So by MCT, $\int_\Omega h_n^* \, d\mu \to \int_\Omega h \, d\mu$. But by Theorem 7.0.3 b), $\int_\Omega h_n^* \, d\mu = \int_\Omega h_n \, d\mu$ for all n, and so the conclusion holds.)

$\triangle$

Theorem 7.0.4. *Let h be Borel measurable*

a) If h is integrable, then h is finite a.e.

b) If $h \geq 0$ and $\int_\Omega h \, d\mu = 0$ then $h = 0$ a.e.

Proof. a) By contraposition. If h is not finite a.e., then $\exists B \in \mathcal{F} : \mu(B) > 0$ and $|h 1_B| = \infty$. Then

$$\int_\Omega |h| \, d\mu \overset{\text{monotonicity}}{\geq} \int_B |h| \, d\mu \overset{\text{simple function}}{=} \infty \, \mu(B) = \infty.$$

So $|h|$ is not integrable. So by Corollary 7.0.2 b), h is not integrable.

b) Let $B_n := \{\omega : h(\omega) \geq \frac{1}{n}\}$. Then $B_n \uparrow B := \{\omega : h(\omega) > 0\}$.[38] (1)

Now

$$0 \overset{\text{first hypothesis}}{\leq} h 1_{B_n} \overset{B_n \subset B}{\leq} h 1_B \overset{\text{def. } B}{=} h$$

$$\implies \int_{B_n} h \, d\mu \overset{\text{monotonicity}}{\leq} \int_\Omega h \, d\mu \overset{\text{second hypothesis}}{=} 0$$

$$\overset{\text{first hypothesis, monotonicity}}{\implies} \int_{B_n} h \, d\mu = 0 \qquad (2)$$

Now (1) and monotonicity again give

$$\int_{B_n} h \, d\mu \overset{\text{def. } B_n\text{, monotonicity}}{\geq} \int \frac{1}{n} 1_{B_n} \, d\mu \overset{\text{simple function}}{=} \frac{1}{n} \mu(B_n) \qquad (3)$$

Now (2) and (3) together give $\mu(B_n) = 0 \; \forall n$. So by continuity of measure

$$\mu(B) = \lim_{n \to \infty} \mu(B_n) = 0$$

$\square$

The monotone convergence theorem as stated earlier only applies to non-negative functions and only to increasing sequences. We relax those assumptions below.

Theorem 7.0.5. ***Extended Monotone Convergence Theorem.*** *Let $f_1, f_2, ..., f, g$ be Borel measurable*

a) If $f_n \uparrow f$ and $f_n \geq g$ for all n, where $\int_\Omega g \, d\mu > -\infty$, then

$$\int_\Omega f_n \, d\mu \uparrow \int_\Omega f \, d\mu$$

[38] Recall that $\cup_{n=1}^\infty [\frac{1}{n}, \infty) = (0, \infty)$.

b) If $f_n \downarrow f$ and $f_n \leq g$ for all n, where $\int_\Omega g \, d\mu < \infty$, then

$$\int_\Omega f_n \, d\mu \downarrow \int_\Omega f \, d\mu$$

Proof. a) If $\int g \, d\mu = \infty$, then by monotonicity (and the fact that the limit of an increasing sequence equals its supremum), $\int f \, d\mu \geq \int f_n \, d\mu \geq \int g \, d\mu$, and the conclusion holds. So assume $\int g \, d\mu < \infty$. Along with the hypothesis, we have that $\int g \, d\mu$ is finite.

Now

$$\begin{aligned}
& f_n - g \geq 0 \quad \text{and} \quad f_n - g \uparrow f - g && \text{Hypothesis (and Prop A.2.4)} \\
\implies & \int (f_n - g) \, d\mu \uparrow \int (f - g) \, d\mu && \text{Monotone Convergence Theorem} \\
\overset{1}{\implies} & \int f_n \, d\mu - \int g \, d\mu \uparrow \int f \, d\mu - \int g \, d\mu && \text{Linearity} \\
\implies & \int f_n \, d\mu \uparrow \int f \, d\mu && \text{Since } \int g \text{ is finite}
\end{aligned}$$

To check that linearity holds in 1, note that $\int f \, d\mu$ and $\int f_n \, d\mu$ exist by monotonicity (Prop 6.3.6 a), and the sum cannot be of form $\infty - \infty$ or $-\infty + \infty$ since $\int g \, d\mu$ is finite.

b) We have

$$\begin{aligned}
& -f_n \geq -g \quad \text{and} \quad -f_n \uparrow -g && \text{Hypothesis} \\
\implies & -\int f_n \, d\mu \uparrow -\int f \, d\mu && \text{Part (a) (and constant multiple property; Prop 6.3.6 a)} \\
\implies & \int f_n \, d\mu \downarrow \int f \, d\mu
\end{aligned}$$

□

Fatou's Lemma

Theorem 7.0.6. ***Extended Fatou's Lemma***[39] *Let $f_1, f_2, ..., g$ be Borel measurable for each positive integer n.*

a) If $f_n \geq g$ for all n where $\int_\Omega g \, d\mu > -\infty$ then

$$\int_\Omega \left(\liminf_{n\to\infty} f_n \right) d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \, d\mu \tag{7.0.8}$$

b) If $f_n \leq g$ for all n where $\int_\Omega g \, d\mu < \infty$ then

$$\int_\Omega \left(\limsup_{n\to\infty} f_n \right) d\mu \geq \limsup_{n\to\infty} \int_\Omega f_n \, d\mu \tag{7.0.9}$$

[39] We refer to Theorem 7.0.6 as *extended* Fatou's lemma in parallel with the extended monotone convergence theorem (Theorem 7.0.5). Some presentations, e.g. [Folland, 1999], present a (non-extended) version of Fatou's lemma that only gives part (a) and which only applies to non-negative measurable functions. We prefer the extended formulation due to its greater generality and supporting of intuition from the "big picture view". Note that in the case of non-negative functions, the hypotheses reduce to simply $f_n \uparrow f$, as there is automatically a measurable g satisfying the remaining conditions, namely $g \equiv 0$.

Proof. a) By definition of the limit inferior,

$$\underbrace{\liminf_{n\to\infty} f_n}_{:= h} = \lim_{n\to\infty} \underbrace{\inf_{m\geq n} f_n}_{:= h_n}$$

Now $h_n \uparrow h$ (due to taking the infimum over successively smaller sets; see Prop. A.2.1), $h_n \geq g$ (since g is a lower bound by hypothesis and the infimum is the *greater* lower bound) where $\int_\Omega g\, d\mu > -\infty$ (by hypothesis).

Hence, by the extended Monotone Convergence Theorem (Thm. 7.0.5)

$$\underbrace{\lim_{n\to\infty} \int h_n\, d\mu}_{= \liminf_{n\to\infty} \int h_n\, d\mu} \overset{\text{(MCT)}}{=} \int \lim_{n\to\infty} h_n\, d\mu \qquad (1)$$

So

$$\liminf_{n\to\infty} \int f_n\, d\mu \overset{(\text{monotonicity, } f_n \geq h_n)}{\geq} \liminf_{n\to\infty} \int h_n\, d\mu \overset{(1)}{=} \int \lim_{n\to\infty} h_n\, d\mu \overset{(\text{def. } h_n)}{=} \int \liminf_{n\to\infty} f_n\, d\mu$$

b)

$$\begin{aligned}
\int_\Omega \limsup_{n\to\infty} f_n\, d\mu &\overset{*}{=} -\int_\Omega \liminf_{n\to\infty} (-f_n)\, d\mu \\
&\geq -\liminf_{n\to\infty} \int_\Omega (-f_n)\, d\mu && \text{part (a)} \\
&\overset{*}{=} -\limsup_{n\to\infty} \int_\Omega (f_n)\, d\mu
\end{aligned}$$

Equality (*) holds by the constant multiple property of the infimum and supremum (Prop A.2.2), which gives that $\limsup_{n\to\infty} f_n = -\liminf_{n\to\infty}(-f_n)$. (Part (a) applies because $f_n \leq g$ where $\int g\, d\mu < \infty$ implies that $-f_n \geq -g$, where $-\int g\, d\mu > -\infty$. Note also that multiplying by a negative reverses the order of the inequality.)

□

Remark 7.0.6. (*Big picture view of Fatou's Lemma*) We can interpret Fatou's lemma as *integrals of asymptotics give more extreme values than asymptotics of integrals*.

If $|f_n| \leq g$ where $\int_\Omega g\, d\mu$ is finite, we have

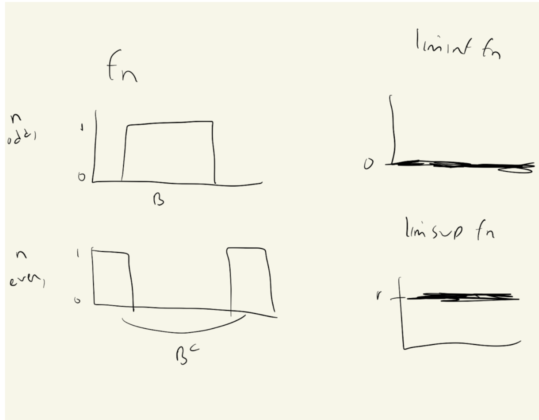
$$\int_\Omega \left(\liminf_{n\to\infty} f_n \right) d\mu \overset{(7.0.8)}{\leq} \liminf_{n\to\infty} \int_\Omega f_n\, d\mu \overset{(A.4.1)}{\leq} \limsup_{n\to\infty} \int_\Omega f_n\, d\mu \overset{(7.0.9)}{\leq} \int_\Omega \left(\limsup_{n\to\infty} f_n \right) d\mu \tag{7.0.10}$$

△

Example 7.0.1. (*Strict inequalities can occur in Fatou's lemma*) We show that that strict inequalities can occur in the expanded view of Fatou's lemma, (7.0.10).

Consider a measure space $(\Omega, \mathcal{F}, \mu)$ and set $B \in \mathcal{F}$ such that $0 < \mu(B^c) < \mu(B) < \mu(\Omega)$. Define a sequence of functions $\{f_n\}$ such that

$$f_n = \begin{cases} 1_B, & \text{n odd} \\ 1_{B^c}, & \text{n even} \end{cases}$$



	Integrate first	**Asymptotics first**
Observation	$\int f_n \, d\mu \; = \begin{cases} \mu(B), & \text{n odd} \\ \mu(B^c), & \text{n even} \end{cases}$	$\liminf_{n\to\infty} f_n = 0$ $\limsup_{n\to\infty} f_n = 1$
Implication	$\liminf_{n\to\infty} \int f_n \, d\mu \; = \mu(B^c)$ $\limsup_{n\to\infty} \int f_n \, d\mu \; = \mu(B)$	$\int \liminf_{n\to\infty} f_n \, d\mu \; = 0$ $\int \limsup_{n\to\infty} f_n \, d\mu \; = \mu(\Omega)$

And so we see that strict inequalities occur in the expanded view of Fatou's lemma, (7.0.10).

$$\underbrace{\int_\Omega \left(\liminf_{n\to\infty} f_n \right) d\mu}_{0} < \underbrace{\liminf_{n\to\infty} \int_\Omega f_n \, d\mu}_{\mu(B^c)} < \underbrace{\limsup_{n\to\infty} \int_\Omega f_n \, d\mu}_{\mu(B)} < \underbrace{\int_\Omega \left(\limsup_{n\to\infty} f_n \right) d\mu}_{\mu(\Omega)}$$

The example captures the fact that *integrals of asymptotics give more extreme values than asymptotics of integrals*.

△

Dominated Convergence Theorem

Theorem 7.0.7. ***Dominated Convergence Theorem.**. If $f_1, f_2, ..., f, g$ are Borel measurable, $|f_n| \le g$ for all n, where g is μ-integrable, and $f_n \to f$ a.e. $[\mu]$, then f is μ-integrable, and $\int_\Omega f_n \, d\mu \to \int_\Omega f \, d\mu$.*

Proof. We have $|f| \le g$ a.e. (In detail, $|f_n| \le g$ can be unpacked into $f_n \le g$ and $f_n \ge -g$. Since limits preserve non-strict inequalities, we have ($f \le g$ and $f \ge -g$) a.e. So repacking the absolute value operator, the conclusion follows.), hence f is integrable by Cor. 7.0.2 (b).

By hypothesis, both sides of the expanded Fatou's lemma apply, and so we have (7.0.10):

$$\underbrace{\int_\Omega \left(\liminf_{n\to\infty} f_n \right) d\mu}_{1} \le \underbrace{\liminf_{n\to\infty} \int_\Omega f_n \, d\mu}_{2} \le \underbrace{\limsup_{n\to\infty} \int_\Omega f_n \, d\mu}_{3} \le \underbrace{\int_\Omega \left(\limsup_{n\to\infty} f_n \right) d\mu}_{4}$$

But since $f_n \to f$ a.e, $\liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n = \lim_{n\to\infty} f_n$ a.e., and so by the a.e. theorem (Thm. 7.0.3), they have the same integrals: $\int_\Omega \liminf_{n\to\infty} f_n \, d\mu = \int_\Omega \limsup_{n\to\infty} f_n \, d\mu = \int_\Omega \lim_{n\to\infty} f_n \, d\mu$. In other words, 1=4, and so by sandwiching 2=3. Since the limit inferior and limit superior of the integrals are equal, the limit of the integrals exists as well, and all together we have

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega \lim_{n\to\infty} f_n \, d\mu .$$

□

Remark 7.0.7. (*Example where limits and integrals cannot be exchanged.*) Let $f_n = \frac{1}{n} 1_{[0,n]}$. Then $\int_\mathbb{R} f_n = 1$ for all n, and so $\lim_{n\to\infty} \int_\mathbb{R} f_n = 1$. But the pointwise limit $f := \lim_{n\to\infty} f_n = 0$ and so $\int_\mathbb{R} f = \int_\mathbb{R} 0 = 0$. △

Continuity and differentiability of functions defined with an integral

We now consider *dependence on a parameter*. Specifically, we consider integrals where the integrand depends on a real parameter.[40] The following theorem describes continuity and the computation of derivative for such functions.

Theorem 7.0.8. ***Continuity and differentiability of functions defined with an integral.*** *Let*

$$\begin{aligned}
&f : \mathcal{X} \times [a,b] \to \mathbb{R}, && \textit{where } (-\infty < a < b < \infty) \\
&f(\cdot, t) : \mathcal{X} \to \mathbb{R} \textit{ be integrable} && \forall t \in [a,b] \\
&F(t) := \int_\mathcal{X} f(x,t) \, d\mu\,(x)
\end{aligned}$$

a) Suppose

$$|f(x,t)| \le g(x) \qquad \forall x, t$$

for some integrable g. If $f(x,\cdot)$ is continuous for each x, then F is continuous.

b) Suppose $\partial f/\partial t$ exists and

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

for some integrable g. Then F is differentiable and

$$F'(t) = \int_\mathcal{X} \frac{\partial f}{\partial t}(x,t) \, d\mu\,(x)$$

Proof. a) We need to show that

$$\lim_{t\to t_0} f(x,t) = f(x,t_0) \implies \lim_{t\to t_0} F(t) = F(t_0)$$

So

$$\begin{aligned}
\lim_{t\to t_0} F(t) &= \lim_{t\to t_0} \int_\mathcal{X} f(x,t) \, d\mu\,(x) && \text{definition of } F \\
&= \int_\mathcal{X} \lim_{t\to t_0} f(x,t) \, d\mu\,(x) && \text{Dominated Convergence Thm} \\
&= \int_\mathcal{X} f(x,t_0) \, d\mu\,(x) && \text{hypothesis} \\
&= F(t_0) && \text{definition of } F
\end{aligned}$$

[40] Here, a “parameter" refers to a variable in the domain of the function that is not integrated over.

b) First we note that $\frac{\partial f}{\partial t}$ is measurable. This is true by the closure properties of measurable functions (Sec. 6.2.5), since

$$\frac{\partial f}{\partial t}(x,t_0) = \lim_{n\to\infty} \underbrace{\frac{f(x,t_n)-f(x,t_0)}{t_n-t_0}}_{:=h_n(x)}$$

for any sequence $\{t_n\}$ converging to t_0.

Next we note that $h_n(x)$ is bounded uniformly in n. For all n, there is an s_n between t_0 and t_n such that

$$|h_n| \stackrel{\text{(def)}}{=} \left| \frac{f(x,t_n)-f(x,t_0)}{t_n-t_0} \right| \stackrel{\text{(MVT)}}{=} \left| \frac{\partial f}{\partial t}(x,s_n) \right| \stackrel{\text{(hypothesis)}}{\leq} g(x)$$

where MVT stands for the Mean Value Theorem.

Thus, we can apply the Dominated Convergence Theorem to h_n, i.e. $\lim_{n\to\infty} \int h_n(x)\, d\mu\,(x) = \int \lim_{n\to\infty} h_n(x)\, d\mu\,(x)$. Using this, we obtain

$$\begin{aligned}
F'(t_0) &= \lim_{n\to\infty} \frac{F(t_n)-F(t_0)}{t_n-t_0} && \text{def. derivative} \\
&= \lim_{n\to\infty} \frac{\int f(x,t_n)\, d\mu\,(x) - \int f(x,t_0)\, d\mu\,(x)}{t_n-t_0} && \text{def. } F \\
&= \lim_{n\to\infty} \int \frac{f(x,t_n)-f(x,t_0)}{t_n-t_0}\, d\mu\,(x) && \text{linearity, applies since } f \text{ integrable} \\
&= \int \frac{\partial}{\partial t} f(x,t)\, d\mu\,(x) && \text{Dominated Convergence Theorem, def. derivative}
\end{aligned}$$

□

Remark 7.0.8. (*Extensions to real-valued parameters with unbounded support*) Theorem 7.0.8 may seem overly restrictive, since the real-valued parameter has bounded support. However, as noted by [Folland, 1999] pp. 56, continuity and differentiability are *local* in nature. Thus, if the hypotheses of (a) or (b) hold for all $[a,b] \subset I$ of an open interval I (which is perhaps $\mathbb{R}$ itself), perhaps with the dominating function g depending on a and b, one obtains the continuity and differentiability of the integrated function F on all of I! △

8 § 1.7 Comparison of Lebesgue and Riemann integrals

In this section, we show that integration with respect to Lebesgue measure is more general than Riemann integration, and we give a precise criterion for Riemann integration.

Review of Riemann integration. Let $[a,b]$ be a bounded closed subset of the reals, and f be a bounded real valued function on $[a,b]$. We assume f is fixed throughout the discussion (i.e., we suppress dependence on f in the notation). Let $P : a = x_0 < x_1 < ... < x_n = b$ be a partition of $[a,b]$. We construct the upper and lower sums as follows. Let

$$\begin{aligned}
M_i &:= \sup\{f(y) : x_{i-1} < y \leq x_i\}, && i = 1,...,n \\
m_i &:= \inf\{f(y) : x_{i-1} < y \leq x_i\}, && i = 1,...,n
\end{aligned}$$

And define step functions α and β, called the *upper* and *lower* functions for f via

$$\begin{aligned}
\alpha(x) &= M_i && \text{if } x_{i-1} < x \leq x_i && i = 1,...,n \\
\beta(x) &= m_i && \text{if } x_{i-1} < x \leq x_i && i = 1,...,n
\end{aligned}$$

[$\alpha(a)$ and $\beta(a)$ may be chosen arbitrarily]. The upper and lower sums are defined as

$$U(P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}) \tag{8.0.1a}$$

$$L(P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \tag{8.0.1b}$$

Now let $P_1, P_2, ...$ be a sequence of partitions such that P_{k+1} is a refinement of P_k for each K, and such that $|P_k|$ (the length of the largest subinterval of P_k) approaches 0 as $k \to \infty$.

If

$$\lim_{k\to\infty} L(P_k) = \lim_{k\to\infty} U(P_k) = r, \tag{8.0.2}$$

independent of the particular sequence of partitions, then f is said to be *Riemann integrable* on $[a, b]$, and r is the value of the *Riemann integral*.[41]

The criterion for Riemann integrability criterion in terms of Lebesgue integration. Now consider the measure space $(\Omega, \mathcal{F}, \mu) = ([a, b], \overline{\mathcal{B}}([a, b]), \text{Lebesgue measure})$, where $\overline{\mathcal{B}}([a, b])$ are the Lebesgue measurable sets (see Section 5.5). Let P_k be a sequence of partitions described earlier, with α_k and β_k the corresponding upper and lower functions. Now since α_k and β_k are simple functions, we can express the upper and lower sums (8.0.1) as integrals with respect to Lebesgue measure μ:

$$U(P_k) = \int_{[a,b]} \alpha_k \, d\,\mu$$

$$L(P_k) = \int_{[a,b]} \beta_k \, d\,\mu$$

Now we can bound the upper and lower functions by an integrable function. (In detail, since we assumed f is bounded, we can write $|f| \le M$, and therefore $|\alpha_k|, |\beta_k| \le M$. Moreover M is integrable, since $\int_{[a,b]} M \; d\,\mu \; = M(b - a) < \infty$.) Moreover, (by Proposition A.2.1), we have

$$\alpha_1 \ge \alpha_2 \ge ... \ge f \ge ... \ge \beta_2 \ge \beta_1$$

so α_k and β_k approach limit functions α and β. Thus, we can apply Lebesgue dominated convergence theorem to obtain

$$\lim_{k\to\infty} U(P_k) = \lim_{k\to\infty} \int_{[a,b]} \alpha_k \, d\,\mu \overset{\text{(LDCT)}}{=} \int_{[a,b]} \alpha \, d\,\mu$$

$$\lim_{k\to\infty} L(P_k) = \lim_{k\to\infty} \int_{[a,b]} \beta_k \, d\,\mu \overset{\text{(LDCT)}}{=} \int_{[a,b]} \beta \, d\,\mu$$

Thus, we can write the criterion for Riemann integrability (8.0.2) in terms of the Lebesgue integral. In particular, f is Riemann integrable over $[a, b]$ with value r iff

$$\int_{[a,b]} \alpha \, d\,\mu \; = \int_{[a,b]} \beta \, d\,\mu \; = r \tag{8.0.3}$$

independent of the sequence of partitions $\{P_k\}$.

[41] TODO: integrate this with my review of the Riemann integral in the appendix.

Continuity at a point as equality of the upper and lower functions. Here we provide a key observation which will help us to relate the Riemann integral and the Lebesgue integral.

Lemma 8.0.1. *If x is not an endpoint of any subintervals of P_k, then*

$$f \text{ is continuous at } x \quad \text{iff} \quad \alpha(x) = f(x) = \beta(x)$$

Proof.

- $\boxed{\implies}$. f is continuous at x means that $\forall \epsilon > 0, \exists \delta > 0:$

$$\begin{aligned} |y - x| < \delta &\implies |f(y) - f(x)| < \epsilon \\ &\implies f(y) \leq f(x) + \epsilon \qquad\qquad \forall y \in B_x(\delta) \\ &\implies \sup_{y \in B_x(\delta)} f(y) \leq f(x) + \epsilon \qquad\qquad 1. \end{aligned}$$

  where $B_x(\delta)$ refers to a ball centered at x with radius δ.

  Now let $I_k(x)$ be the subinterval in the kth partition to which x belongs. Since $|P_k| \to 0$, $|I_k(x)| \to 0$, and so

$$\forall \delta > 0, \exists K : \forall k \geq K, \quad |y - x| < \delta \quad \forall y \in I_k(x) \qquad 2$$

  Combining 1 and 2, we obtain

$$\forall \delta > 0, \exists K : \forall k \geq K, \quad \sup_{y \in I_k(x)} f(y) \leq f(x) + \epsilon \qquad 3$$

  And of course, since the supremum is an upper bound and $x \in I_k(x)$,

$$\sup_{y \in I_k(x)} f(y) \geq f(x). \qquad 4$$

  So combining 3 and 4, and recalling that $\alpha_k(x) := \sup_{y \in I_k(x)} f(y)$, we have

$$\forall \delta > 0, \exists K : \forall k \geq K, \quad |\alpha_k(x) - f(x)| \leq \epsilon$$

  which is the definition of the limit. That is,

$$\lim_{k \to \infty} \alpha_k(x) = f(x)$$

  A similar argument holds for β_k.

- $\boxed{\impliedby}$. By hypothesis,

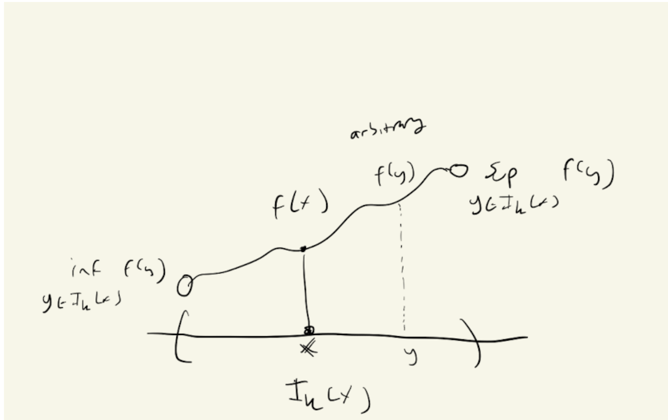
  $\forall \epsilon > 0, \exists K : \forall k \geq K,$

$$\begin{aligned} &|\sup_{y \in I_k(x)} f(y) - f(x)| \leq \epsilon \quad \text{and} \quad |\inf_{y \in I_k(x)} f(y) - f(x)| \leq \epsilon \\ \overset{1}{\implies}\ & f(y) - f(x) < \epsilon \qquad \text{and} \qquad f(y) - f(x) > -\epsilon \quad \forall y \in I_k(x) \end{aligned}$$

  where (1) holds since the supremum is an upper bound and the infimum is a lower bound. So

$$\forall \epsilon > 0, \exists K : \forall k \geq K, \quad |f(y) - f(x)| < \epsilon \quad \forall y \in I_k(x)$$

  and taking δ to be the radius of $I_k(x)$, the definition of the continuity of f at x holds.



$\square$

The theorem.

Theorem 8.0.1. *Let f be a bounded real-valued function on $[a, b]$.*

a) The function f is Riemann integrable on $[a, b]$ iff f is continuous almost everywhere on $[a, b]$ (with respect to Lebesgue measure).

b) If f is Riemann integrable on $[a, b]$, then f is integrable with respect to Lebesgue measure on $[a, b]$, and the two integrals are equal.

Proof. a) • $\boxed{\Longrightarrow}$. By (8.0.3), if f is Riemann integrable, then

$$\int_{[a,b]} \alpha \, d\mu \; = \int_{[a,b]} \beta \, d\mu \; = r$$

By linearity (which holds immediately since α, β are integrable by the above equation; see Remark 7.0.3),

$$\int_{[a,b]} (\alpha - \beta) \, d\mu \; = 0$$

Since $\beta \le f \le \alpha$ (since each β_k and α_k are lower and upper bounds by construction, and limits preserve non-strict inequalities), we have $\alpha - \beta \ge 0$, so by Theorem 7.0.4 (b), $\alpha - \beta = 0$ a.e. So (by sandwiching) $\alpha = \beta = f$ a.e. So by Lemma 8.0.1, f is continuous a.e.[42]

• $\boxed{\Longleftarrow}$. By hypothesis and Lemma 8.0.1, $\alpha = f = \beta$ a.e. As a result, f is measurable. (α, β are limits of simple functions, and hence measurable by closure properties of simple functions. Thus, f differs from a measurable function only on a subset of a set of measure 0. Since the Lebesgue measurable sets $\overline{\mathcal{B}}([a, b])$ are complete, f must be measurable by Problem 4.2.1.)[43] Since $\alpha = f = \beta$ a.e., by Theorem 7.0.3 (b), we have

$$\int_{[a,b]} \alpha \, d\mu \; = \int_{[a,b]} f \, d\mu \; = \int_{[a,b]} \beta \, d\mu \tag{8.0.4}$$

Thus f is Riemann integrable by (8.0.3).[44]

[42] Implicit in this last statement, I think, is that the endpoints are are set of measure 0, even in the limit. If E_k denotes the set of endpoints of the subintervals of P_k, and E_{ik} denotes the ith such endpoint for $i = 1, ..., N_k$, then $\mu(\lim_{k\to\infty} E_k) \overset{1}{=} \lim_{k\to\infty} \mu(E_k) = \lim_k \mu(\cup_{i=1}^{N_k} E_{ik}) \overset{2}{=} 0$, where (1) holds by continuity of measure (from below; recall each successive partition is a refinement of the previous) and (2) holds since by additivity of measure, since each E_{ik} is a singleton and the Lebesgue measure of any singleton is 0.

[43] After this line [Ash et al., 2000] also adds that f is integrable. (Since f is bounded, say $|f| \le L$, we have $\int_{[a,b]} f \, d\mu \le \int_{[a,b]} L \, d\mu = L(b-a) < \infty$.) But on my reading, we can just use the a.e. theorem directly after arguing that f is measurable.

[44] Presumably the “independent of the sequence of partitions" condition of (8.0.3) is met here, since the argument – in particular, Lemma 8.0.1 – does not seem to depend on the sequence of partitions.

b) If f is Riemann integrable, then f is continuous a.e. by part (a). So (8.0.4) holds. Then, by (8.0.3), $\int_{[a,b]} f \, d\mu = r$, the value of the Riemann integral.

□

9 §2.1 Signed measures

This section plays the same role as Section 2.1 of [Ash et al., 2000], although the primary reference here is Section 3.1 of [Folland, 1999].

9.1 Overview of signed measures

Definition 9.1.1. Let $(\Omega, \mathcal{F})$ be a measurable space.[45] A **signed measure** on $(\Omega, \mathcal{F})$ is a function $\nu : \mathcal{F} \to [-\infty, \infty]$ such that

- $\nu(\emptyset) = 0$;
- ν assumes at most one of the values $\pm\infty$;
- (*countable additivity*) If $\{A_j\}$ is a sequence of disjoint sets in $\mathcal{F}$, then $\nu(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \nu(A_j)$.

△

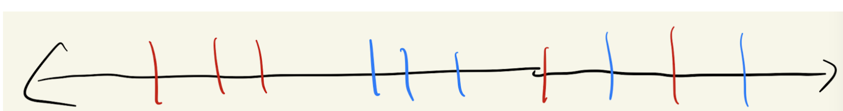
Remark 9.1.1. Every measure is a signed measure. △

More generally, two mechanisms for forming signed measures come to mind:

1. $\nu = \mu_1 - \mu_2$, where μ_1 and μ_2 are measures on $\mathcal{F}$, and at least one measure is finite.
2. The set function defined by $\nu(A) = \int_A h \, d\mu$, where $h : \Omega \to [-\infty, \infty]$ is a measurable function such that at least one of $\int h^+ \, d\mu$ and $\int h^- \, d\mu$ is finite. (The function is defined $\forall A \in \mathcal{F}$ on measure space $(\Omega, \mathcal{F}, \mu)$.).

Remark 9.1.2. [Folland, 1999] points out that these are the *only* two examples; every signed measure can be represented in either of these two forms. △

Example 9.1.1. We can define a signed measure on all subsets of the reals by the number of blue ticks minus the number of red ticks.[46]



△

Example 9.1.2. We can define a signed measure on the reals by

$$\nu(A) := \int_A |x| \, d\mu \, (x)$$

where μ is the Lebesgue measure. To illustrate, compare the behavior of these set functions on two different sets:

[45] Recall Definition 6.2.1.

[46] Note that even though *Lebesgue measure* could not be defined on *all* subsets of the reals, clearly other signed measures can be. Recall from Remark that the problem was translation invariance: we showed that there cannot be a translation invariant measure that assigns a finite value to all intervals. The (signed) measure in this example, however, is clearly not translation invariant.

	Set	
Output	$[-5, -4]$	$[-5, 5]$
μ	1	10
ν	-1	0

△

Remark 9.1.3. Signed measures are continuous by Theorem 3.4.1 (continuity of countably additive set functions). △

9.2 Hanh and Jordan Decompositions

Definition 9.2.1. Given a signed measure ν on a measurable space $(\Omega, \mathcal{F})$, a set $A \in \mathcal{F}$ is called

- **positive** if $\nu(B) \geq 0$ for all $B \subset A$.
- **negative** if $\nu(B) \leq 0$ for all $B \subset A$.
- **null** if $\nu(B) = 0$ for all $B \subset A$.

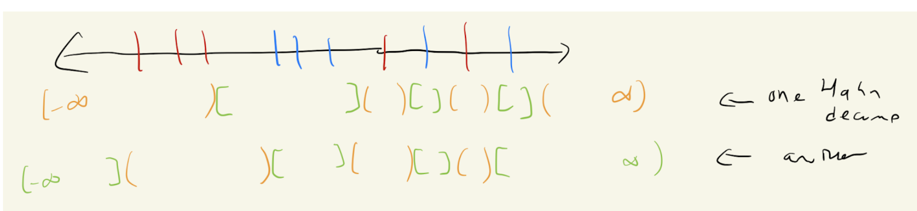
△

Remark 9.2.1. To motivate Definition 9.2.1, consider that unlike measures, *signed* measures do not have the monotonicity property. So, for example, the condition $\nu(A) = 0$ in isolation provides no information about the value of ν on subsets. For an illustration, recall Example 9.1.1. △

Theorem 9.2.1. ***The Hanh Decomposition Theorem*** *If ν is a signed measure on $(\Omega, \mathcal{F})$, there exist a positive set P and a negative set N such that $\Omega = P \cup N$. Moreover, if P', N' is another such pair, then $P \triangle P' (= N \triangle N')$ is null for ν.*

Proof. See [Folland, 1999] pp. 86. □

Example 9.2.1. Let us consider two possible Hanh decompositions $(P, N), (P', N')$ for the signed measure of Example 9.1.1. The union of green intervals is a positive set P, and the union of orange intervals is a negative set N. While there are a number of possible Hanh decompositions, these choices are "not that different" from each other in the sense that for any two decompositions $(P, N), (P', N')$, the set $P \triangle P'$ (i.e. the set of points in exactly one version of green coloring), which equals $N \triangle N'$ (i.e. the set of points in exactly one version of orange coloring), is null for ν. [In other words, the set of "disagreement" is such that all subsets have signed measure 0.]



△

Definition 9.2.2. Let ν, λ be signed measures on a measurable space $(\Omega, \mathcal{F})$. Then ν and λ are said to be **mutually singular**, written $\nu \perp \lambda$, if there exist $A, B \in \mathcal{F}$ such that $\Omega = A \cup B$ where ν is null for A and λ is null for B. △

So two signed measures are mutually singular if there is a partition of the universe into two cells such that each cell is null for a different signed measure.

Example 9.2.2. The normal distributions truncated to the positive reals ($\mathcal{N}_+$) and negative reals ($\mathcal{N}_-$) provide an example of mutually singular measures. Note that in Definition 9.2.2, we can take either $(A, B) = (\mathbb{R}_0^+, \mathbb{R}^-)$ or $(A', B') = (\mathbb{R}^+, \mathbb{R}_0^-)$, illustrating the non-uniqueness of the Hanh Decomposition. △

Now we show that any signed measure can be expressed as the difference of two (latent) measures, which are moreover mutually singular.

Theorem 9.2.2. ***The Jordan decomposition theorem.*** *If ν is a signed measure, there exist unique measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.*

Proof.

- *Existence*. Let $\Omega = P \cup N$ be a Hanh decomposition for ν. (So $\nu(A) \geq 0$ for all $A \subset P$ and $\nu(B) \leq 0$ for all $B \subset N$).

  Define set functions ν^+ and ν^- by:

  $$\nu^+(E) := \nu(E \cap P)$$
  $$\nu^-(E) := -\nu(E \cap N)$$

  which are both measures (as they are non-negative and countably additive).

  Then:

  - $\nu = \nu^+ - \nu^-$.

    This holds because E = $(E \cap P) \cup (E \cap N)$, so by countable additivity, $\nu(E) = \nu(E \cap P) + \nu(E \cap N) := \nu^+(E) - \nu^-(E)$.
  - $\nu^+ \perp \nu^-$.

    This holds because

    $$\nu^+(N) = \nu(N \cap P) = \nu(\emptyset) = 0$$
    $$\nu^-(P) = -\nu(P \cap N) = \nu(\emptyset) = 0,$$

    so $\Omega = P \cup N$ is the partition required for mutual singularity. (By monotonicity of measure, the equalities hold for subsets as well).
- *Uniqueness*. TBD, or see [Folland, 1999] pp. 87.

□

Remark 9.2.2. Don't let the notation confuse you. Both ν^+ and ν^- are (positive) measures. The superscripts are meant to designate that ν^+ is the minuend and ν^- is the subtrahend in $\nu = \nu^+ - \nu^-$. △

9.3 Total Variation

Definition 9.3.1. The **total variation** of a signed measure ν, denoted $|\nu|$, is the measure given by

$$|\nu| := \nu^+ + \nu^-$$

△

Remark 9.3.1. Definition 9.3.1 is well-defined because the sum of two measures is a measure:

- Non-negativity ✓
- Countable additivity ✓

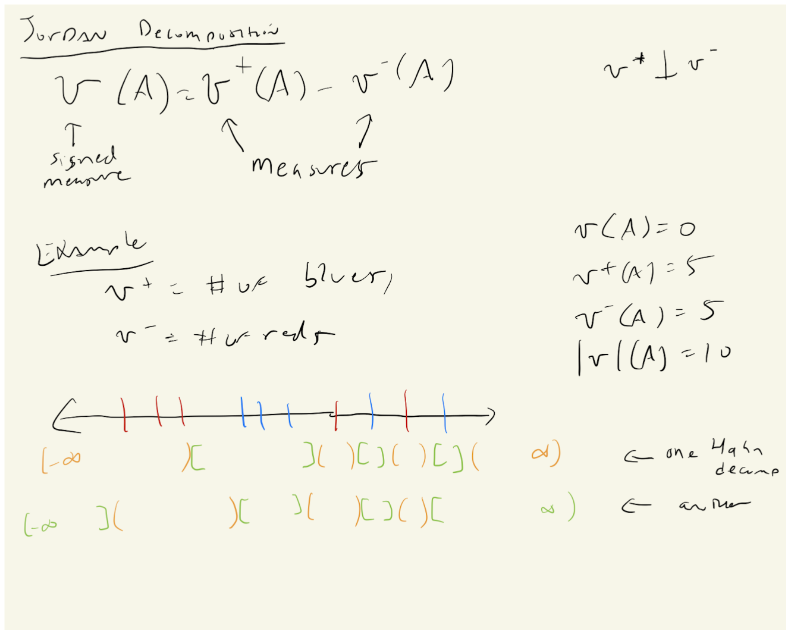
  Countable additivity holds because[47]

  $$\begin{aligned} |\nu|\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \nu^+\Big(\bigcup_{i=1}^{\infty} A_i\Big) + \nu^-\Big(\bigcup_{i=1}^{\infty} A_i\Big) \\ &= \sum_{i=1}^{\infty} \nu^+(A_i) + \sum_{i=1}^{\infty} \nu^-(A_i) && \nu^+, \nu^- \text{ are measures} \\ &= \sum_{i=1}^{\infty} \nu^+(A_i) + \nu^-(A_i) && \text{Limits and sums commute; see [Strichartz, 2000] Theorem 2.3.2 and apply it to partial sums} \\ &= \sum_{i=1}^{\infty} |\nu|(A_i) && \text{def. of } |\nu| \end{aligned}$$

[47] TODO: Argue instead by appealing to a more general theorem.

$\triangle$

Remark 9.3.2. (*An illustrative example of Hanh-Jordan Decomposition.*) The picture below illustrates the structure of a signed measure ν, as provided by a Jordan decomposition. As we will see in Problem 9.3.1, we can think about the "set of disagreement" ($P \triangle P' (= N \triangle N')$) between any two Hanh Decompositions ($(P, N), (P', N')$) as having a total variation $|\nu|$ of 0, so both ν^+ and ν^- give zero measure to the sets of disagreement.



$\triangle$

Problem 9.3.1. (*Null sets are those with zero total variation.*) Given a signed measure ν on a measure space $(\Omega, \mathcal{F})$ and a set $E \in \mathcal{F}$, show that

$$E \text{ is null for } \nu \quad \text{iff} \quad |\nu|(E) = 0$$

Solution. First recall the definitions

- E is null for ν: $\nu(F) = 0 \quad \forall F \subset E$.
- $|\nu|(E) = 0$: $\nu^+(E) + \nu^-(E) = 0$.

Now the proof

- $\boxed{\Longrightarrow}$. Let $\Omega = P \uplus N$ be a Hanh decomposition for ν. Then

$$\nu^+(E) = \nu(E \cap P) = 0$$
$$\nu^-(E) = -\nu(E \cap N) = 0$$

where the first equalities are the Jordan decomposition, and the second equalities follow since E is null for ν (and $E \cap P$, $E \cap N \subset E$).

So

$$|\nu|(E) = \nu^+(E) + \nu^-(E) = 0.$$

• $\boxed{\Longleftarrow}$.

$$\begin{aligned}
|\nu|(E) = 0 &\iff \nu^+(E) + \nu^-(E) = 0 && \text{def. } |\nu| \\
&\implies \nu^+(E) = 0, \nu^-(E) = 0 && \text{non-negativity of measure} \\
&\implies \nu^+(F) = 0, \nu^-(F) = 0 \quad \forall F \subset E && \text{monotonicity of measure} \\
&\implies \nu^+(F) + \nu^-(F) = 0 \quad \forall F \subset E \\
&\implies \nu(F) = 0 \quad \forall F \subset E
\end{aligned}$$

□

10 §2.2 Lebesgue-Radon-Nikodym Theorem

This section plays the same role as Section 2.2 of [Ash et al., 2000], although the primary reference here is Section 3.2 of [Folland, 1999].

10.1 Absolute continuity

In this section, we define absolute continuity and give some useful properties.

Definition 10.1.1. Let ν be a signed measure and μ a measure on a measurable space $(\Omega, \mathcal{F})$. We say that ν is **absolutely continuous** with respect to μ, and write

$$\nu \ll \mu$$

if

$$\mu(A) = 0 \implies \nu(A) = 0 \quad \forall A \in \mathcal{F}.$$

△

Example 10.1.1. Below we give examples. For these examples, let $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

1. Let μ be a univariate Gaussian and ν be a Gaussian that is truncated (e.g. to the set of positive reals). Then $\nu \ll \mu$ but $\mu \not\ll \nu$.
2. Let μ be Lebesgue measure and ν give the number of integers in a set. Then $\nu \not\ll \mu$ and $\mu \not\ll \nu$.
3. Let μ be Lebesgue measure and ν be twice Lebesgue measure. Then $\nu \ll \mu$ and $\mu \ll \nu$.

△

Remark 10.1.1. (*Indefinite integrals are absolutely continuous.*)

Let ν be an "indefinite integral" with respect to μ.[48]

$$\text{i.e. } \nu(A) = \int_A h \, d\mu \,, \qquad \forall A \in \mathcal{F}$$

Then $\nu \ll \mu$

$$\text{i.e. } \mu(A) = 0 \implies \nu(A) = 0 \qquad \forall A \in \mathcal{F}$$

△

[48] This interpretation of "indefinite integral" is used in [Ash et al., 2000, pp. 61].

Proof. [49] If $\mu(A) = 0$, $h1_A = 0$ a.e. So by Theorem 7.0.3 (a),

$$\int_A h \, d\mu \; = 0$$

$\square$

Remark 10.1.2. (*Radon-Nikodym as a converse.*) The Radon-Nikodym theorem is an assertion in the opposite direction of Remark 10.1.1: if $\nu \ll \mu$ (and μ is σ-finite on $\mathcal{F}$), then ν is an indefinite integral with respect to μ. $\triangle$

Remark 10.1.3. (*A signed measure is absolutely continuous iff its total variation is absolutely continuous.*)[50] Let ν be a signed measure and μ a measure on a measurable space $(\Omega, \mathcal{F})$. Then

$$\nu \ll \mu \quad \text{iff} \quad |\nu| \ll \mu$$

$\triangle$

Proof. Using the Jordan decomposition and the definitions (of absolute continuity and total variation), we write the following for reference:

$$\begin{aligned}
\nu \ll \mu \;\; & \text{means} \\
& \mu(A) = 0 \implies \nu(A) = 0 \\
& \qquad \text{i.e. } \nu^+(A) - \nu^-(A) = 0 \\
|\nu| \ll \mu \;\; & \text{means} \\
& \mu(A) = 0 \implies |\nu|(A) = 0 \\
& \qquad \text{i.e. } \nu^+(A) + \nu^-(A) = 0
\end{aligned}$$

for all $A \in \mathcal{F}$.

And now we proceed:

- $\boxed{\Longleftarrow}$. The hypothesis plus non-negativity implies

$$\nu^+(A), \; \nu^-(A) = 0$$

- $\boxed{\Longrightarrow}$.

$$\begin{aligned}
\mu(A) = 0 \quad & \overset{\text{monotonicity}}{\Longrightarrow} \quad A \text{ is } \mu\text{-null} \\
& \qquad\qquad\qquad \text{i.e. } \mu(B) = 0 \quad \forall\, B \subset A \\
& \overset{\text{hypothesis}}{\Longrightarrow} \quad \nu(B) = 0 \quad \forall\, B \subset A \\
& \overset{\text{null sets have zero total variation (Problem 9.3.1)}}{\Longrightarrow} \quad |\nu|(A) = 0
\end{aligned}$$

$\square$

Remark 10.1.4. (*Absolute continuity is the antithesis of mutual singularity.*)

[49] My first attempt at a proof worked through the definition for Lebesgue integral, working up from simple functions to non-negative to arbitrary functions. However, the proof we provide here, which uses the strategy in Section 2.2 of [Ash et al., 2000], is much nicer. Note that my original strategy of proving via the definition just basically recapitulates the proof of Theorem 7.0.3 (a), so we might as well rely on the Theorem to do that dirty work.

[50] This is [Folland, 1999] Problem 3.2.8a

This statement (from [Folland, 1999]) captures the fact that

$$\begin{array}{rl} \text{if} & \nu \text{ is a signed measure} \\ & \mu \text{ is a measure} \\ & \nu \ll \mu \\ & \nu \perp \mu \\ \text{then} & \nu \equiv 0 \end{array}$$

△

Proof. First recall

$$\nu \ll \mu \overset{\text{Remark 10.1.3}}{\iff} |\nu| \ll \mu$$

Now by mutual singularity

$$\begin{array}{lllll} \exists A, B \in \mathcal{F}: & \Omega = A \,\dot\cup\, B & & & \\ & \nu \text{ is null for } A & \overset{\text{Problem 9.3.1}}{\iff} & |\nu|(A) = 0 & \\ & \mu \text{ is null for } B & \overset{\text{special case}}{\implies} & \mu(B) = 0 \overset{|\nu| \ll \mu}{\implies} & |\nu|(B) = 0 \end{array}$$

So

$$\begin{aligned} |\nu|(\Omega) &\overset{\text{countable additivity}}{=} |\nu|(A) + |\nu|(B) \overset{\text{see above}}{=} 0 \\ &\overset{\text{monotonicity, since } |\nu| \text{ is a measure}}{\implies} |\nu| \equiv 0 \\ &\overset{*}{\implies} \nu \equiv 0. \end{aligned}$$

Implication (*) holds since

$$\begin{aligned} |\nu| \equiv 0 &\overset{\text{def}}{\iff} \nu^+ + \nu^- \equiv 0 \\ &\overset{\text{non-negativity}}{\implies} \nu^+ \equiv 0, \nu^- \equiv 0 \\ &\implies \nu^+ - \nu^- \equiv 0 \\ &\overset{\text{def}}{\iff} \nu \equiv 0. \end{aligned}$$

□

We now provide some motivation for the name “absolute continuity".

Theorem 10.1.1. (Name-justifying characterization of absolute continuity.) *Let ν be a finite signed measure and μ a positive measure on $(\Omega, \mathcal{F})$. Then*

$$\begin{array}{lcl} \nu \ll \mu & \textit{iff} & \forall \epsilon > 0,\ \exists \delta > 0 : \forall A \in \mathcal{F}, \\ (\textit{i.e. } \mu(A) = 0 \implies \nu(A) = 0) & & \mu(A) < \delta \implies |\nu(A)| < \epsilon \\ \text{A} & & \text{B} \end{array}$$

Proof.

- $\boxed{\Longleftarrow}$.

$$\mu(A) = 0 \implies \mu(A) < \delta \quad \forall \delta \overset{\text{hypothesis}}{\implies} |\nu(A)| < \epsilon \quad \forall \epsilon \implies \nu(A) = 0$$

- $\boxed{\implies}$. We proceed by contraposition. Since not B,

$$\begin{array}{llll} \exists \epsilon > 0 : & \forall n \in \mathbb{N}, & \exists A_n \in \mathcal{F} : & \\ \mu(A_n) \le 2^{-n} & \text{but} & |\nu|(A_n) \overset{*}{\ge} |\nu(A_n)| \ge \epsilon & + \end{array}$$

where in Equation (*), we move work with $|\nu|$ since it is a measure.
The inequality holds since

$$|\nu(A)| \leq |\nu|(A)$$
$$|\nu^+(A) - \nu^-(A)| \leq |\nu^+(A) + \nu^-(A)|$$
$$|x - y| \leq |x + y|$$

where we have applied the Jordan decomposition and properties of real numbers.

To show A, recall from Remark 10.1.3 that $\nu \ll \mu$ iff $|\nu| \ll \mu$, so

$$\text{N.T.S.} \quad \exists B \in \mathcal{F}: \ \mu(B) = 0 \quad \text{but} \quad |\nu|(B) \neq 0$$

Let $B_k = \cup_{n=k}^{\infty} A_n, \ B = \cap_{k=1}^{\infty} B_k$. So

$$B = \limsup A_n = \{x : x \in A_n \text{ i.o. }\}$$

Then by the Borel-Cantelli Lemma (Lemma 3.5.1), $\mu(B) = 0$. But

$$|\nu|(B) \overset{\text{cty from above}}{=} \lim_{k \to \infty} |\nu|(B_k) \overset{\text{monotonicity, } +}{\geq} \lim_{k \to \infty} \epsilon = \epsilon.$$

where continuity from above holds because B_k are decreasing, and $|\nu|(B_1)$ is finite because ν finite $\implies$ $|\nu|$ finite.
In fact ν finite $\iff$ $|\nu|$ finite. This can be seen by appealing to the Jordan decomposition of ν and the definition of $|\nu|$. For any $E \in \mathcal{F}$,

$$\nu(E) = \nu^+(E) - \nu^-(E) \notin \{\infty, -\infty\}$$
$$\iff$$
$$|\nu|(E) = \nu^+(E) + \nu^-(E) \notin \{\infty, -\infty\}.$$

Both statements say that the two numbers $\nu^+(E)$, $\nu^-(E)$ are both finite.

□

As a corollary, we have that an integral of any (integrable) function over a set can be made arbitrarily small if the measure of the set is sufficiently small.

Corollary 10.1.1. *Let f be an integrable function with respect to μ. Then $\forall \epsilon > 0, \ \exists \delta > 0$ such that*

$$\mu(A) < \delta \implies \left| \int_A f \, d\mu \right| < \epsilon$$

Proof. By Theorem 7.0.1, the set function $\nu : \mathcal{F} \to \overline{\mathbb{R}}$ defined by $\nu(A) = \int_A f \, d\mu$ is a signed measure. Moreover, $\nu \ll \mu$, since indefinite integrals are absolutely continuous (see Remark 10.1.1). Thus, the implication follows from the name-justifying characterization of absolute continuity (Theorem 10.1.1). □

10.2 The theorem

Lemma 10.2.1. *Let ν, μ be a finite measures on $(\Omega, \mathcal{F})$. Then either $\nu \perp \mu$ or there exist $\epsilon > 0$ and $A \in \mathcal{F}$ such that $\mu(A) > 0$ and $\nu \geq \epsilon\mu$ on A.*

Proof. See [Folland, 1999, pp. 89]. □

Theorem 10.2.1. ***The Lebesgue-Radon-Nikodym Theorem.*** *Let ν be a σ-finite signed measure and μ a σ-finite (positive) measure on $(\Omega, \mathcal{F})$. There exist unique σ-finite signed measures λ, ρ on $(\Omega, \mathcal{F})$ such that*

$$\nu = \lambda + \rho, \quad \textit{where} \quad \lambda \perp \mu \quad \textit{and} \quad \rho \ll \mu$$

Moreover, there is an extended μ-integrable function $f : \Omega \to \mathbb{R}$ such that $d\rho = f \, d\mu$, and any two such functions are equal μ-a.e.

Proof. We prove the theorem in the special case that ν and μ are finite (positive) measures. For a full proof, see [Folland, 1999, pp. 90].

Let

$$\mathcal{S} := \left\{ f : \Omega \to [0, \infty] : \int_E f \, d\mu \leq \nu(E) \quad \forall E \in \mathcal{F} \right\}$$

$\mathcal{S}$ is nonempty since $0 \in \mathcal{S}$. Now if $f, g \in \mathcal{S}$, then $h := \max(f, g) \in \mathcal{S}$, because if $A := \{x : f(x) > g(x)\}$, then for any $E \in \mathcal{F}$, we have

$$\begin{aligned} \int_E h \, d\mu &= \int_\Omega 1_E h \, d\mu = \int_\Omega \Big(1_{E \cap A} + 1_{E \cap A^c} \Big) h \, d\mu \\ &= \int_{E \cap A} h \, d\mu + \int_{E \cap A^c} h \, d\mu \\ &= \int_{E \cap A} f \, d\mu + \int_{E \cap A^c} g \, d\mu \\ &\overset{\text{(since } f, g \in \mathcal{S})}{\leq} \nu(E \cap A) + \nu(E \cap A^c) \\ &\overset{\text{countable additivity}}{=} \nu(E). \end{aligned}$$

Now let $a := \sup\{\int_\Omega f \, d\mu : f \in \mathcal{S}\}$, and note that $a \overset{\text{sup is LUB}}{\leq} \nu(\Omega) \overset{\text{assumption}}{<} \infty$. By a property of the supremum (see Remark A.1.3), we can find a sequence $\{f_n\} \in \mathcal{S}$ such that $\int_\Omega f_n \, d\mu \to a$. Now if we let $g_n := \max(f_1, \ldots, f_n)$, then $g_n \in \mathcal{S}$ by applying induction, since we have shown that $\mathcal{S}$ is closed under the maximum operator. It follows that $\int_\Omega g_n \, d\mu \to a$.

Let us show that $\lim_{n \to \infty} \int_\Omega g_n \, d\mu = a$.

- $\boxed{\geq}$.

$$\begin{aligned} &\int_\Omega g_n \, d\mu \geq \int_\Omega f_n \, d\mu && \text{by monotonicity of integral} \\ \implies &\lim_{n \to \infty} \int_\Omega g_n \, d\mu \geq \lim_{n \to \infty} \int_\Omega f_n \, d\mu = a && \text{limits preserve non-strict inequalities} \end{aligned}$$

- $\boxed{\leq}$. Since $g_n \in \mathcal{S}$,

$$\begin{aligned} &\int_\Omega g_n \, d\mu \leq a && \text{supremum is upper bound} \\ \implies &\lim_{n \to \infty} \int_\Omega g_n \, d\mu \leq a && \text{limits preserve non-strict inequalities} \end{aligned}$$

Since g_n is an increasing sequence, there exists a function $f : \Omega \to [0, \infty]$ such that $g_n \uparrow f$ (see Prop. A.4.2),[51] and by monotone convergence theorem

$$\int_\Omega f \, d\mu = \lim_{n \to \infty} \int_\Omega g_n \, d\mu = a < \infty$$

So $f < \infty$ a.e. (by Theorem 7.0.4 a), and so we may take f to be real-valued everywhere.

Now define ρ by $\rho(A) = \int_A f \, d\mu$, and define $\lambda := \nu - \rho$. (In differential notation, we can express both conditions simultaneously via $d\lambda = d\nu - f d\mu$.) Then the existence conditions of the theorem hold:

- $\nu = \lambda + \rho$? ✓ Addition.
- $\rho \ll \mu$? ✓ This holds because indefinite integrals are absolutely continuous (see Remark 10.1.1).

[51] TODO: Show specifically that $f = \sup_n f_n$. The proposition gives us that $g = \sup_n g_n = \sup_n \max_{m \leq n} f_n$.

- $\lambda \perp \mu$? ✓ First note $\lambda \geq 0$ since $f \in \mathcal{S}$. So we proceed via Lemma 10.2.1. BWOC, suppose not $\lambda \perp \mu$. Then there is $\epsilon > 0, A \in \mathcal{F} : \mu(A) > 0$ and $\lambda \geq \epsilon\mu$ on A. Now

$$\begin{aligned} \lambda &\geq 1_A \epsilon\mu && \text{by hypothesis, and } \lambda \geq 0 \\ \implies d\nu - f\, d\mu &\geq \epsilon 1_A\, d\mu && \text{by def. } \lambda\text{, differential notation} \\ \implies d\nu &\geq (f + \epsilon 1_A)\, d\mu && \text{addition} \\ \implies (f + \epsilon 1_A) &\in \mathcal{S} && \text{by def. } \mathcal{S} \end{aligned}$$

But by linearity (which applies immediately since $(f + \epsilon 1_A)$ is integrable), we have

$$\int_\Omega (f + \epsilon 1_A)\, d\mu = a + \epsilon\mu(A) > a,$$

which contradicts the definition of the supremum.

Now we show uniqueness.

$$\begin{aligned} \text{We have} \quad & d\nu = d\lambda + f d\mu \\ \text{Suppose also} \quad & d\nu = d\lambda' + f' d\mu \end{aligned}$$

Then[52]

$$\begin{array}{llll} & d\lambda - d\lambda' & = (f - f')d\mu & \\ \text{But} & \lambda - \lambda' & \perp \mu & \text{Justified below} \\ \text{And} & \lambda - \lambda' & \ll \mu & \text{Since } d(\lambda - \lambda') = (f - f')\, d\mu \text{; indefinite integrals are absolutely continuous (Remark 10.1.1)} \\ \text{So} & \lambda - \lambda' & = 0 & \ll \text{ and } \perp \text{ are antitheses; see Remark 10.1.4} \\ \text{So} & \lambda & = \lambda' & \\ \text{So} & f & = f' \quad [\mu] - a.e. & \text{See [Folland, 1999] Prop. 2.23} \end{array}$$

Let us show that if $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 - \lambda_2 \perp \mu$.

$$\begin{array}{lll} \lambda_1 \perp \mu & \text{means} & \exists E_1, F_1 : E_1 \uplus F_1 = \Omega \\ & & E_1 \text{ null for } \lambda_1 \\ & & F_1 \text{ null for } \mu \\ \lambda_2 \perp \mu & \text{means} & \exists E_2, F_2 : E_2 \uplus F_2 = \Omega \\ & & E_2 \text{ null for } \lambda_2 \\ & & F_2 \text{ null for } \mu \end{array}$$

Now let $\lambda = \lambda_1 - \lambda_2$.

$$\begin{array}{ll} \text{Let} & E = E_1 \cap E_2 \\ & \text{Then } E \text{ null for } \lambda_1 \text{ and } \lambda_2\text{, so } E \text{ null for } \lambda = \lambda_1 - \lambda_2 \\ \text{Let} & F := E^c = (E_1 \cap E_2)^c = E_1^c \cup E_2^c = F_1 \cup F_2 \\ & \text{So } F \text{ null for } \mu. \end{array}$$

□

11 Implications

TODO: Mention RN as a special case

Proposition 11.0.1. *Suppose that ν is a σ-finite signed measure and μ, λ are σ-finite measures on $(\Omega, \mathcal{F})$ such that $\nu \ll \mu$ and $\mu \ll \lambda$. Then*

a) If g is integrable, then $g(d\nu/d\mu)$ is integrable, and

$$\int g\, d\nu = \int g \frac{d\nu}{d\mu}\, d\mu$$

b) We have $\nu \ll \lambda$, and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \quad \lambda - a.e.$$

Proof. a) See [Folland, 1999, pp.91]. The proof is similar to that of Prop. 7.0.1.

[52] Include [Folland, 1999] Prop. 2.23 in these notes. It is strange that [Ash et al., 2000] omits this. It can probably be obtained fairly immediately from Theorem 7.0.4.

b) For all $A \in \mathcal{F}$, we have

$$\begin{aligned}
\nu(A) &= \int_A d\nu && \text{Def. } \int \text{ simple functions, since } \int_A d\nu := \int 1_A \, d\nu \\
&= \int_A \frac{d\nu}{d\mu} \, d\mu && \nu \ll \mu, \text{ Radon-Nikodym} \\
&= \int_A \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \, d\lambda && \mu \ll \lambda \text{ and part (a) (replacing } \nu, \mu \text{ by } \mu, \lambda\text{) where } g = \frac{d\nu}{d\mu} 1_A.
\end{aligned} \tag{11.0.1}$$

On the other hand, we have $\nu \ll \lambda$. ($\ll$ can be seen to be transitive immediately from the definition). Thus, by Radon-Nikodym,

$$\nu(A) = \int_A \frac{d\nu}{d\lambda} \, d\lambda \tag{11.0.2}$$

Since (11.0.1) and (11.0.2) both hold for all $A \in \mathcal{F}$, then by [Folland, 1999] Prop. 2.23,

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \quad \lambda - \text{a.e.}$$

□

References

[Apostol, 1974] Apostol, T. (1974). *Mathematical Analysis (2nd edition*.

[Ash et al., 2000] Ash, R. B., Robert, B., Doleans-Dade, C. A., and Catherine, A. (2000). *Probability and measure theory*. Academic Press.

[Durrett, 2010] Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.

[Folland, 1999] Folland, G. B. (1999). *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons.

[Rosenthal, 2006] Rosenthal, J. S. (2006). *First Look At Rigorous Probability Theory, A*. World Scientific Publishing Company.

[Rudin, 1987] Rudin, W. (1987). Real and complex analysis.

[Strichartz, 2000] Strichartz, R. S. (2000). *The way of analysis*. Jones & Bartlett Learning.

A Supremum and Infimum

Following are some definitions and propositions that we use in the notes.[53]

A.1 Characterization

First, we define upper and lower bounds.

Definition A.1.1. A set $A \subset \mathbb{R}$ of real numbers is bounded from above if there exists a real number $M \in \mathbb{R}$, called an *upper bound* of A, such that $x \leq M$ for every $x \in A$. Similarly, A is bounded from below if there exists a real number $m \in \mathbb{R}$, called an *lower bound* of A, such that $x \geq m$ for every $x \in A$. A set is *bounded* if it is bounded from above and below. △

Now, we define infimum and supremum.

Definition A.1.2. Suppose that $A \subset \mathbb{R}$ is a set of real numbers. If $M \in \mathbb{R}$ is an upper bound of A such that $M \leq M'$ for every upper bound M' of A, then M is called the *supremum* of A, denoted $M = \sup A$. Similarly, if $m \in \mathbb{R}$ is an lower bound of A such that $m \geq m'$ for every lower bound m' of A, then m is called the *infimum* of A, denoted $m = \inf A$. △

We sometimes use an alternate characterization of infimum and supremum.

Proposition A.1.1. *If $A \subset \mathbb{R}$, then $M = \sup A$ if and only if (a) M is an upper bound of A; (b) for every $M' < M$, there exists an $a \in A$ such that $a > M'$. Similarly, $m = \inf A$ if and only if (a) m is a lower bound of A; (b) for all $m' > m$, there exists an $a \in A$ such that $a < m'$.*

Proof. We prove the alternate characterization for the supremum only, as the proof for infimum is similar. We only need to show equivalence for the part (b)'s, as the part (a)'s are identical.

We first show that the definition implies part the proposition. We proceed by way of contradiction. Let $M = \sup A$, $M' < M$, and suppose there is no $a \in A : a > M'$. Then M' is an upper bound of A where $M' < M$, contradicting part (b) of the definition of supremum.

Now we show that the proposition implies the definition. Part (b) of the proposition implies that if $M' < M$, then M' is not an upper bound. Thus part (b) of the definition is satisfied. □

Remark A.1.1. The (b) statement in Proposition A.1.1 roughly tell us that any other candidate for a smaller supremum fails, because it will not be an upper bound. Similarly, any other candidate for a larger infimum fails, because it will not be a lower bound. △

Remark A.1.2. Another way to write Proposition A.1.1 is as follows:

If $A \subset \mathbb{R}$, then $M = \sup A$ if and only if (a) M is an upper bound of A; (b) for all $\epsilon > 0$, there exists an $a \in A$ such that $a > M - \epsilon$. Similarly, $m = \inf A$ if and only if (a) m is a lower bound of A; (b) for all $\epsilon > 0$, there exists an $a \in A$ such that $a < m + \epsilon$. △

Remark A.1.3. (*Existence of sequences converging to the infimum and supremum*) By Remark A.1.2, if $A \subset \mathbb{R}$, we can always find a non-decreasing sequence $\{x_n\} \subset A$ such that $\lim_{n\to\infty} x_n = \sup A$, and likewise for the infimum. △

Remark A.1.4. (*When the upper or lower bound is contained in the set itself*) Note that if a set A *contains* an upper bound, then it is automatically a supremum. That is, if M is an upper bound of A such that $M \in A$, then the *least* upper bound property (the second condition in the characterizations above) automatically follows. For instance, condition (b) in Proposition A.1.1 is immediately satisfied by setting $a = M$. When the supremum is contained in the set, it is called a *maximum*. A similar remark holds for infima: when the lower bound is contained in the set, it is automatically the infimum, and it is called the minimum. △

[53] For a nice introductory overview, see `https://www.math.ucdavis.edu/~hunter/m125b/ch2.pdf`.

Remark A.1.5. One way to show that $\sup A = \sup B$ is using breaking the equality into $\boxed{\leq}$, $\boxed{\geq}$. We can then show

- $\boxed{\leq}$ if RHS is an upper bound on the LHS
- $\boxed{\geq}$ if RHS is a supremum over a subset of the LHS (i.e. if $B \subset A$)

$\triangle$

A.2 Properties

The proposition below characterizes the behavior of the infimum and supremum under set containment. Namely, making a set smaller increases its supremum and decreases its infimum.

Proposition A.2.1. *Suppose that A and B are subsets of $\mathbb{R}$ such that $A \subset B$. If $\sup A$ and $\sup B$ exist, then $\sup A \leq \sup B$. If $\inf A$ and $\inf B$ exist, then $\inf A \geq \inf B$.*

Now we characterize the behavior of the infimum and supremum when we multiply a set by a constant.

Definition A.2.1. If $A \subset \mathbb{R}$ and $c \in \mathbb{R}$, we define

$$cA := \{x \in \mathbb{R} : x = ca \text{ for some } a \in A\}$$

$\triangle$

Proposition A.2.2.

If $c \geq 0$ then

$$\sup cA = c \sup A, \qquad \inf cA = c \inf A$$

If $c < 0$ then

$$\sup cA = c \inf A, \qquad \inf cA = c \sup A$$

Proof. If $c = 0$, then the result holds because $cA = \{0\}$ and $\sup\{0\} = \inf\{0\} = 0$ by Remark A.1.4. If $c > 0$, then $M \geq a$ if and only if $cM \geq ca$, so M is an upper bound of A if and only if cM is an upper bound of cA, so $\sup cA = c \sup A$. If $c < 0$, then $M \geq a$ if and only if $cM \leq ca$, so M is an upper bound of A if and only if cM is a lower bound of cA, so $\inf cA = c \sup A$. The remaining results follow similarly. $\square$

Now we characterize the behavior of infimum and supremum over set (Minskowski) sums and differences.

Definition A.2.2. If $A, B \subset \mathbb{R}$ are non-empty, we define the *Minkowski sum* of the two sets, denoted $A + B$, by

$$A + B := \{z : z = x + y \text{ for some } x \in A, y \in B\}$$

Similarly, we define the *Minkowski difference* of two sets, denoted $A - B$, by

$$A - B := \{z : z = x - y \text{ for some } x \in A, y \in B\}$$

$\triangle$

Proposition A.2.3. *If $A, B \subset \mathbb{R}$ are non-empty, then*

$$\sup(A+B) = \sup A + \sup B, \quad \inf(A+B) = \inf A + \inf B$$
$$\sup(A-B) = \sup A - \inf B, \quad \inf(A-B) = \inf A - \sup B$$

Remark A.2.1. Proposition A.2.3 can be informally described as saying that the infimum and supremum distribute over addition and subtraction, but negative signs “flip" infima to suprema, and vice versa. $\triangle$

Proposition A.2.4. *Let $\{x_n\}$ be a sequence of real numbers. If $\{x_n\}$ is increasing, then its supremum is the limit. If $\{x_n\}$ is decreasing, then its infimum is the limit.*

A.3 Limit inferior and limit superior of real valued sequences

Definition A.3.1. If a_n is a sequence of real numbers, then we define its limit inferior and limit superior as

$$\liminf_{n\to\infty} a_n := \lim_{n\to\infty} \inf_{m\geq n} a_m, \qquad \limsup_{n\to\infty} a_n := \lim_{n\to\infty} \sup_{m\geq n} a_m$$

$\triangle$

Remark A.3.1. The definitions can be presented in quantified form. For example, [Apostol, 1974] defines the limit superior as follows. Suppose there is a real number S satisfying the following two conditions

1. $$\forall \epsilon > 0, \exists N \in \mathbb{N} : \forall n \geq N, \\ x_n < S + \epsilon$$

2. $$\forall \epsilon > 0, N \in \mathbb{N}, \exists n \geq N : \\ x_n > S - \epsilon$$

then S is the limit superior of $\{x_n\}$. We can think of (1) as stating that S is an asymptotic upper bound; ultimately *all* terms of the sequence lie to the left of $S + \epsilon$. We can think of the additional condition (2) as then guaranteeing that S is an asymptotic least upper bound; *infinitely many* terms lie to the right of $U - \epsilon$ (so $U - \epsilon$ cannot be an upper bound for any n). $\triangle$

Unlike the limit, the limit inferior and limit superior always exist, although they may be $\pm\infty$.

Proposition A.3.1.

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

Proof.

$$\liminf_{n\to\infty} a_n \overset{\text{def}}{=} \lim_{n\to\infty} \inf_{m\geq n} a_m \overset{*}{\leq} \lim_{n\to\infty} \sup_{m\geq n} a_m \overset{\text{def}}{=} \limsup_{n\to\infty} a_n.$$

(Note: * holds because limits preserve non-strict inequalities.) $\square$

A.4 Functions

The supremum and infimum of functions are the supremum and infimum of its range, and so results about sets translate immediately to results about functions.

Definition A.4.1. If $f : A \to \mathbb{R}$ is a function, then

$$\sup_A f := \sup\{f(x) : x \in A\}, \qquad \inf_A f := \inf\{f(x) : x \in A\}$$

$\triangle$

One useful result is that the limit of an increasing sequence of functions is its supremum.

Proposition A.4.1. *Let $f_n : A \to \mathbb{R}$ be functions for all $n \in \mathbb{N}$. If $f_n \leq f_{n+1}$ for all n, and $f_n \to f$, then $f = \sup_n f_n$.*

If we allow for extended real-valued functions, we can make a stronger proposition.

Proposition A.4.2. *Let $f_n : A \to \mathbb{R}$ be functions for all $n \in \mathbb{N}$. If $f_n \leq f_{n+1}$ for all n, then there exists $f : A \to \overline{\mathbb{R}}$ such that $f_n \to f$ pointwise. In particular, $f = \sup_n f_n$.*

A.4.1 Limit inferior and limit superior of functions

Definition A.4.2. If $\{f_n\}$ is a sequence of functions, then we define its limit inferior and limit superior as

$$\liminf_{n\to\infty} f := \lim_{n\to\infty} \inf_{m\geq n} f_m, \qquad \limsup_{n\to\infty} f := \lim_{n\to\infty} \sup_{m\geq n} f_m$$

$\triangle$

Applying Proposition A.3.1 to a function pointwise, we have

$$\liminf_{n\to\infty} f \leq \limsup_{n\to\infty} f \tag{A.4.1}$$

B Some information relevant to Riemann integrals

Let us present (using [Strichartz, 2000] for reference, along with pp. 56 of [Folland, 1999] and pp.55 of [Ash et al., 2000]) some information relevant to Riemann integration for real-valued functions.

First off, recall that the Riemann integral $\int_a^b f(x)\,dx$, when it exists, is defined only for real-valued functions f whose domain is a compact space $[a, b] \subset \mathbb{R}$.

Now consider the following definitions:

- A *partition* of a compact interval $[a, b]$ is a finite sequence $P = \{x_i\}_{i=0}^n$ such that $a = x_0 < x_1 < ... < x_n = b$.
- Given a partition P, the *upper sum* $S^+(f, P)$ and *lower sum* $S^-(f, P)$ are defined by[54]

$$S^+(f, P) := \sum_{i=1}^n M_i(x_i - x_{i-1})$$

where M_i is the supremum of f on $(x_{i-1}, x_i]$.

$$S^-(f, P) := \sum_{i=1}^n m_i(x_i - x_{i-1})$$

where m_i is the infimum of f on $(x_{i-1}, x_i]$.

[54] References [Strichartz, 2000] and [Folland, 1999] define the supremum and infimum over closed sets $[x_{i-1}, x_i]$, whereas [Ash et al., 2000] defines them over $(x_{i-1}, x_i]$. It presumably doesn't matter.

- The *oscillation* $\text{Osc}(f, P)$ of a function f over a partition P is given by the difference of the upper and lower sums; i.e. $\text{Osc}(f, P) := S^+(f, P) - S^-(f, P)$.
- The *maximum interval length* of P is defined by $\max_{i=1,\dots,n} \big|[x_{i-1}, x_i]\big|$.

Finally, we present a proposition we use in the main text.

Proposition B.0.1. *If f is a bounded real-valued function on $[a, b]$, then f is Riemann integrable iff $Osc(f, P) \to 0$ as the maximum interval length of P goes to 0.*

Proof. See Theorem 6.2.1 of [Strichartz, 2000]. □

C Miscellaneous

C.1 Right semi-closed intervals

Definition C.1.1. A **right semi-closed interval** is a set of the form $(a, b] = \{x : a < x \leq b\}, -\infty \leq a < b < \infty$. By convention, we also count (a, ∞) as right semi-closed for $-\infty \leq a < \infty$. △

C.2 DeMorgan's Law applies to relative complements

Remark C.2.1. DeMorgan's Law also holds for relative complements. That is, given a sequence of sets $A_1, A_2, \dots$ that are subsets of another set X, we have:

$$X - \bigcap_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (X - A_n) \tag{C.2.1}$$

△