Notes on Probability and Measure Theory

January 27, 2023

Contents

1 Overview

1.1 References

These notes are geared towards statisticians, computer scientists, engineers, and others who are interested in measure theory in the service of data science (specifically, statistics or statistical machine learning). The primary reference 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 – 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] for filling in depth. Durrett [2010] has excellent examples and high-level overviews, but in my view does not provide enough detail to become fluent in the topic.

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).[1]
- Measure theory is convenient in unifying various kinds of random variables.[2]
- 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).
- General 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.[3]
- Add sketches to support intuition.

[1]For instance, Gaussian processes do not have densities, and so the common presentation of basic tools (e.g., Bayes theorem, KL divergence) no longer applies.

[2]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.

[3]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.

- Curate the text.[4]
- Provide additional detail in proofs. Sometimes alternate paths have been given that seemed “nicer" to me.
- Add remarks (illustrating the need for propositions, the utility of theorems, or connections between things).
- Incorporate supporting material from outside sources (or worked homework problems).
- Give some examples and notes for data scientists.[5]

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,

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

$\triangle$

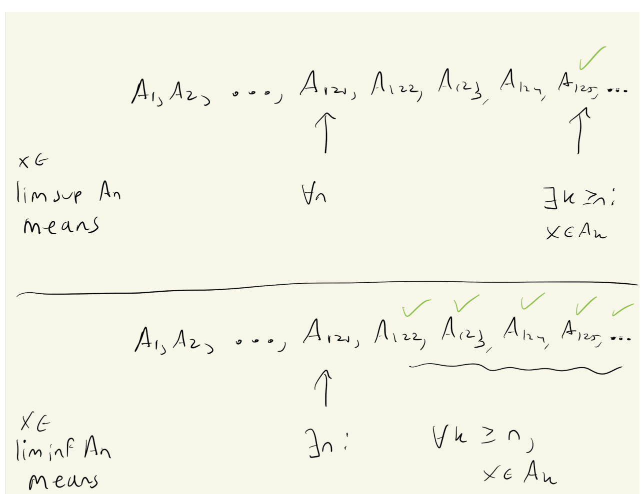
Definition 2.1.2. The **lower limit** of a sequence of sets is given by

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

Alternatively,

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

$\triangle$



[4]We highlight some of the main themes (and cores of proofs), offloading additional detail to the text.
[5]This is a work in progress.

Example 2.1.1. (*A sequence of sets with empty lower limit and non-empty upper limit.*)

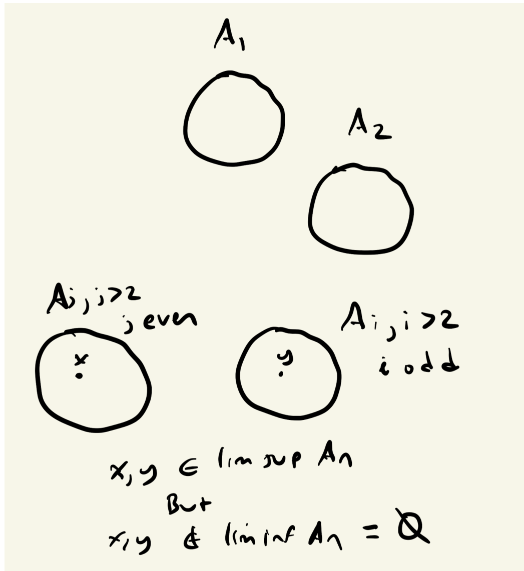


Figure 1: A sequence of sets with empty lower limit and non-empty upper limit.

△

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

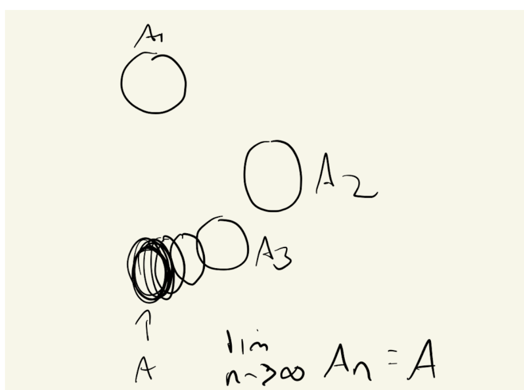


Figure 2: A sequence of sets with a limit.

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 $A_n \downarrow A$. △

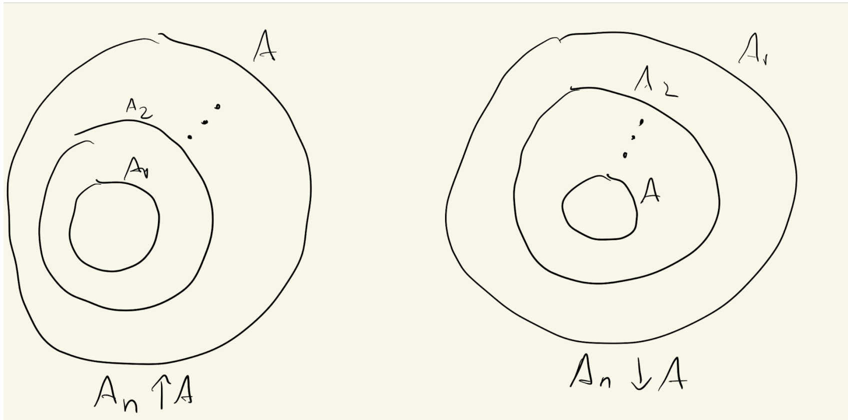


Figure 3: An increasing and decreasing sequence of sets.

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

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

For instance, let $A_n \uparrow A$. So $A_1 \subset A_2 \subset \ldots$ and $A = \bigcup_{n=1}^{\infty} A_n$.

Then

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k \overset{\text{containment}}{=} \bigcup_{n=1}^{\infty} A_n \overset{\text{def.}}{=} A$$

and

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k \overset{\text{containment}}{=} \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_k \overset{\text{constant}}{=} \bigcup_{k=1}^{\infty} A_k \overset{\text{def.}}{=} A$$

So if $A_n \uparrow A$, then $\liminf A_n = \limsup A_n = A$, i.e. the sequence of sets has a limit, and we write $A = \lim_n A_n$.

As shown in Figure 3, limits of increasing and decreasing sequences are very special kinds of limits.

2.2 Representing unions as disjoint unions

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

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

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. $\triangle$

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

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

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

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[6]

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

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

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

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

△

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. △

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*. △

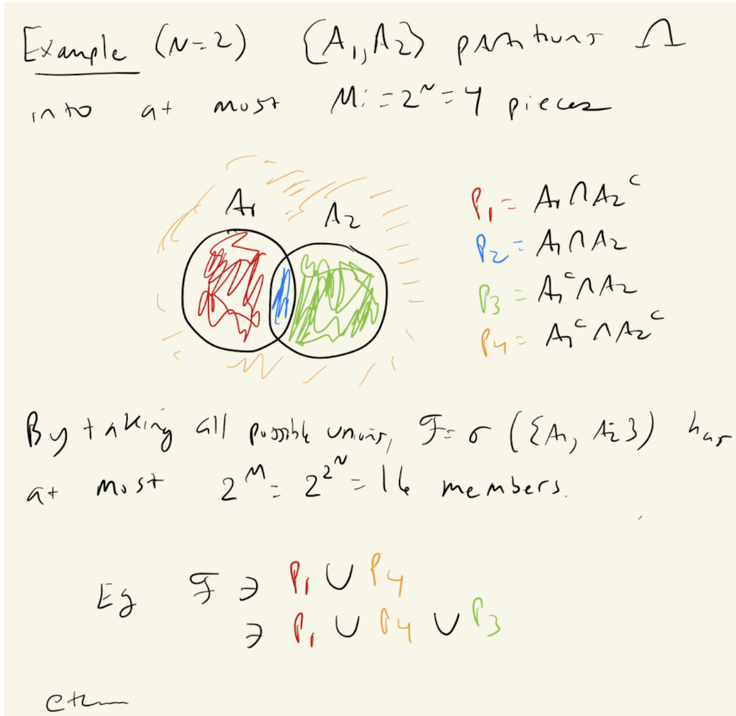
Problem 3.1.1. (*[Ash et al., 2000] Problem 1.2.8*) Let $A_1, ..., A_n$ be subsets of Ω. Describe $\mathcal{F} := \sigma(\{A_1, ..., A_n\})$, the smallest σ-field containing $A_1, ..., A_n$. Also describe the number of sets in $\mathcal{F}$.

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.

□

[6] 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.



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.

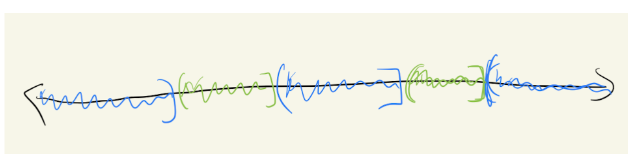
Definition 3.1.2. Let $\mathcal{F}$ be a collection of subsets of a set Ω. Then $\mathcal{F}$ is called a **field** (or *algebra*) if satisfies Definition 3.1.1 after replacing condition c) with

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. $\triangle$

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 $(a, b], a \neq b\}$.[7] Then $\mathcal{F}_0$ is a field, as can be motivated graphically (or see Remark 3.1.2 for a precise argument).



But $\mathcal{F}_0$ is <u>not</u> a σ-field.

Note that if $A_n = (-\frac{1}{n}, 0]$, then $\bigcap_{n=1}^{\infty} A_n = \{0\} \notin \mathcal{F}_0$.

$\triangle$

Now we give a simple way to form fields [Folland, 1999, pp. 23]. We do this in Prop. 3.1.1. First, a definition.

[7] 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. See Def. D.1.1.

Definition 3.1.3. An **elementary family** is a collection $\mathcal{E}$ of subsets of Ω such that

a) $\emptyset \in \mathcal{E}$

b) if $E, F \in \mathcal{E}$ then $\mathcal{E} \cap \mathcal{F} \in \mathcal{E}$

c) if $E \in \mathcal{E}$, then E^c is a finite disjoint union of members of $\mathcal{E}$.

$\triangle$

Proposition 3.1.1. *If $\mathcal{E}$ is an elementary family then the collection*

$$\mathcal{F}_0 := \{\textit{finite disjoint unions of members of } \mathcal{E}\}$$

is a field.

Proof. See [Folland, 1999, pp.24]. $\square$

Remark 3.1.2. (*The collection of disjoint unions of right semi-closed rectangles is a field.*) By Prop. 3.1.1, it is sufficient to show that the right semi-closed rectangles $\{(a, b], a \neq b\}$ (see Def. D.1.1 for a precise definition) are an elementary family. So we verify the conditions:

a) $\emptyset \in \mathcal{E}$? ✓ . Take $b < a$.

b) if $E, F \in \mathcal{E}$ then $\mathcal{E} \cap \mathcal{F} \in \mathcal{E}$? ✓ . Let $E = (a_1, b_1]$ and $F = (a_2, b_2]$.

- If $a_2 \geq b_1$, then $E \cap F = \emptyset \underbrace{\in \mathcal{E}}_{\text{by (a)}}$.
- If $a_2 < b_1$, then $E \cap F = (a_2, b_1] \underbrace{\in \mathcal{E}}_{\text{by def rsc intervals}}$.

c) if $E \in \mathcal{E}$, then E^c is a finite disjoint union of members of $\mathcal{E}$? ✓ . If $E = (a, b]$, then

$$E^c = \underbrace{(-\infty, a]}_{\in \mathcal{E} \text{ by def r.s.c}} \cup \underbrace{(b, \infty)}_{\in \mathcal{E} \text{ by a technicality; see Def. D.1.1}}$$

$\triangle$

Remark 3.1.3. 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

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

- See my handwritten homework exercise for § 1.2, Problem 6.
- Sections of measurable sets are measurable.

Remark 3.1.4. 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. △

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 right semi-closed 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 $(-\infty, b], \ b \in \mathbb{R}$.[8]

i) all open sets of $\mathbb{R}$.[9]

[8] Let $a \in \mathbb{R}$. Intervals of the form (a, ∞) or $(-\infty, a)$ are called open rays. Intervals of the form $[a, \infty)$ or $(-\infty, a]$ are called closed rays.

[9] Recall that an open set is a countable union of open intervals.

j) all closed sets of $\mathbb{R}$.[10]

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

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

For example, the first equation in Eq. (3.1.1) can be verified by the argument

$$x \in \bigcap_{n=1}^{\infty} \left(a, b+\frac{1}{n}\right) \iff x \in (a, b - \frac{1}{n})\ \forall n \iff x > a \text{ and } x < b - \frac{1}{n}\ \forall n \iff x > a \text{ and } x \le b \iff x \in (a,b]$$

.

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 3.1.5 notes that the Cantor set is a Borel set.

Remark 3.1.5. (*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. △

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,

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

△

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

Remark 3.2.1. Ash additionally assumes that a measure does not take $\mu(A) = \infty$ for all $A \in \mathcal{F}$.[11] 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. △

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. △

[10] Recall that a set is open iff its complement is closed.

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

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**. (It gives the number of points in $A \subset \Omega$.)
- If $\sum_i p_i = 1$, then μ is a discrete 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. $\triangle$

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

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

In Table 1, we introduce some notation to try to clarify more immediately when results hold. Note the relations[12]

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

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. $\triangle$

[12]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.

Example 3.3.1. Let $\mathcal{F}_0$ be the field of finite disjoint unions of right semi-closed intervals (see Definition D.1.1), and define the set function $\tilde{\mu}_0$ on $\mathcal{F}_0$ as follows[13]:

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

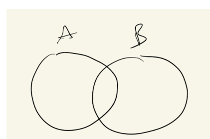
Then $\tilde{\mu}_0$ is finitely additive, but not countably additive on $\mathcal{F}_0$. (Why?) For a proof, see GoodNotes. $\triangle$

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*

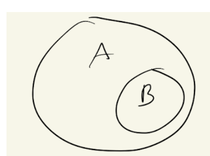
a) $\tilde{\mu}_0(\emptyset) = 0$

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*

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



[14]*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.*

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

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

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

[14]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.*

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

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

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

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

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)

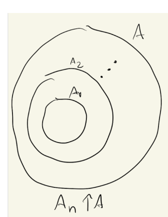
The fact that equivalence holds for signed measures $\widetilde{\mu}$ is surprising. Somehow countable additivity compensates for the signedness. See Section 2.1.3 of the text. △

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



b) (continuity from above) If $A_1, A_2, \ldots \in \mathcal{F}$, $A_n \downarrow A$, 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

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

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 $\widetilde{\mu}(A) = \infty$.[15] 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

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

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. △

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 $\mathcal{F}$. △

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).

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

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

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

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

3.5 Borel-Cantelli Lemma

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

Proof. [16] 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

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

□

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. △

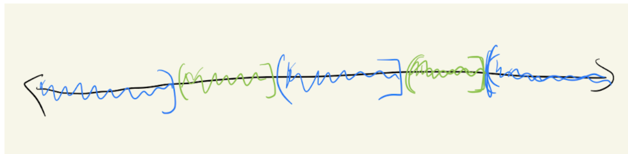
4 § 1.3: Extension of measures

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 $\mathbb{R}$.

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

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 extended 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 ?. 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]. $\square$

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]. $\square$

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 3.1.3.)

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.

△

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. △

For an example of where the strategy in Remark 4.1.2 is used, see the proof of uniqueness in the Carathéodory 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 pre-measure on a 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 Eq. (2.2.1) to re-express the countable union as a disjoint countable union: $\Omega = \cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i$, where $B_i := A_i \cap A_{i-1}^c ... \cap A_1^c$.]

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 $\cup_{i=1}^{\infty} A_i$ is a disjoint union, then so is $\cup_{i=1}^{\infty}(A_i \cap A_n)$, and $\mu(\cup_{i=1}^{\infty}(A_i \cap A_n)) = \sum_{i=1}^{\infty} \mu(A_i \cap A_n)$ 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 $\mu(A) = \mu(A\cap\Omega) = \mu(A\cap(\cup_{n=1}^{\infty} A_n)) = \mu(\cup_{n=1}^{\infty}(A\cap A_n)) = \sum_{n=1}^{\infty} \mu(A\cap A_n) = \mu_n(A)$.]

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:

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

].

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 3.1.3.) 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. (*Folland's statement of Carathéodory's extension theorem, with explicit construction of the measure.*) [Folland, 1999, Thm. 1.14] gives a useful version of Theorem 4.1.3, which provides an explicit construction of the measure extended from a pre-measure. (The constructiveness will be useful e.g. for product measures (see Sec. 13.2; in particular see the product measure of the unit square diagonal under lebesgue and counting measure factors Ex. 13.5.1.)

Folland's construction is as follows: Let $\mathcal{F}_0 \subset \mathcal{P}(\Omega)$ be a field, μ_0 be a premeasure on $\mathcal{F}_0$, and $\mathcal{F} = \sigma(\mathcal{F}_0)$. Then there exists a measure μ on $\mathcal{F}$ such that $\mu|_{\mathcal{F}_0} = \mu_0$, namely for all $A \in \mathcal{F}$,

$$\mu(A) = \inf\left\{ \sum_{j=1}^{\infty} \mu_0(A) : A_j \in \mathcal{F}_0, \bigcup_{j=1}^{\infty} A_j \supset A \right\} \tag{4.1.1}$$

Folland's construction also highlights that the σ-finite assumption in Theorem 4.1.3 is only necessary for uniqueness. △

Remark 4.1.5. (*Folland's statement of Carathéodory's extension theorem does not require σ-finite pre-measures.*) In [Folland, 1999, Thm. 1.14], we see that the extension theorem also applies to non-σ-finite pre-measures, but in that case, we lose uniqueness. △

Remark 4.1.6. 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.

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 $A = \emptyset$). □

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 $B \subset A$. △

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 $\mu(A \cup S) = \mu(A)$. △

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

$$(A \cup S)^c = (A^c \cap S^c) = \underbrace{(A^c \cap N^c)}_{\in \mathcal{F}} \cup \underbrace{(A^c \cap S^c - N^c)}_{\text{has measure 0}}$$

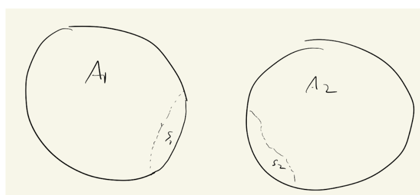
where the term on the right has measure 0 by monotonicity, because $A^c \cap S^c - N^c \subset S^c - N^c = S^c \cap (M^c)^c = S^c \cap N \subset N$.

2. First, we show that countable additivity holds in $\mathcal{F}_\mu$.

$$\mu(\cup_{i=1}^{\infty}(A_i \cup S_i)) \overset{\textit{see below}}{=} \mu(\cup_{i=1}^{\infty} A_i) \overset{\mu \textit{ countably additive on } \mathcal{F}}{=} \sum_{i=1}^{\infty} \mu(A_i) \overset{\textit{construction of extension}}{=} \sum_{i=1}^{\infty} \mu(A_i \cup S_i)$$

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

$$\mu(A_1) \overset{\textit{countable additivity}}{=} \mu(A_1 \cap A_2) + \mu(A_1 \cap A_2^c) \overset{\textit{see below}}{=} \mu(A_1 \cap A_2) \overset{\textit{monotonicity}}{\leq} \mu(A_2)$$

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

$$\mu(C) \overset{\textit{monotonicity}}{\leq} \mu(B) \overset{B \in \mathcal{F}_\mu}{=} \mu(A \cup S) \overset{\textit{subadditivity}}{\leq} \mu(A) + \mu(S) \overset{\textit{see above}}{=} 0 + \mu(S) = 0 + 0 = 0$$

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 $\mathcal{F}_\mu$.

$\triangle$

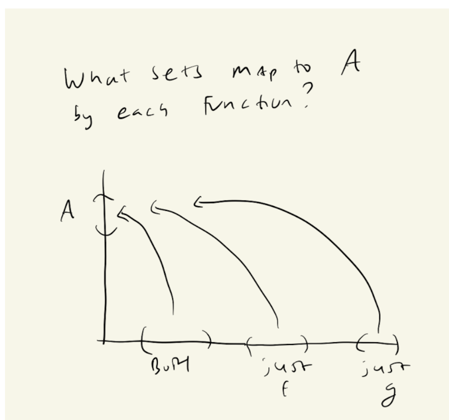
Example 4.2.1. Let us provide a simple example of where completing a measure space generates new measurable sets. Consider a measure space given by $(\Omega, \mathcal{F}, \mu)$ where $\Omega = \mathbb{R}$, $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ where we take $A = [0,1]$ for concreteness, and where 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 $\mathcal{F}_\mu = \{\emptyset, A, A^c, \Omega, \text{ any subset of } [0,1]\}$. $\triangle$

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 $A \in \mathcal{F}'$,

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

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



$\square$

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. $\triangle$

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 [$\lim_{x \downarrow x_0} F(x) = F(x_0)$]. $\triangle$

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. (For reference, Fig. 4 shows a distribution function.)

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 $F : \mathbb{R} \to \mathbb{R}$ 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.

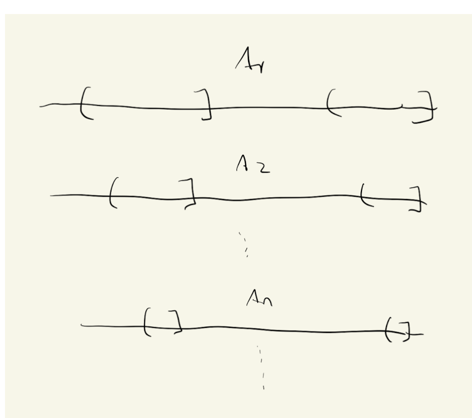
We will temporarily work with $\overline{\mathbb{R}}$, because it is a compact space, and then convert back to $\mathbb{R}$. In $\overline{\mathbb{R}}$, by a similar reasoning as we've seen before (e.g. see Section 4.1), it is straightforward to show that the formula $\mu(a, b] = F(a) - F(b), a, b \in \overline{\mathbb{R}}, a < b$ 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 $\mathcal{F}_0(\overline{\mathbb{R}})$.*

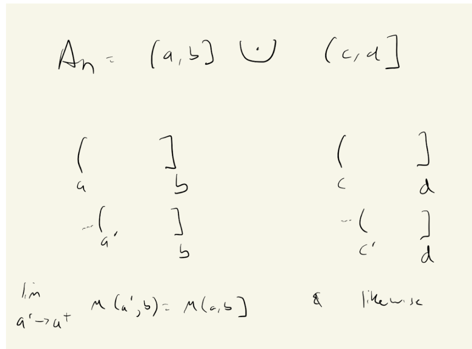
Proof. We assume $F(\infty) - F(-\infty) < \infty$, so that μ is finite. (We leave the case where $F(\infty) - F(-\infty) = \infty$ to the reader, or see the text.) Our strategy will be to show that μ is continuous from above, in which case we can apply Theorem 3.4.2 (b) to show that the set function 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 $\overline{\mathbb{R}} \overset{\text{item a)}}{=} (\overline{\mathbb{R}} - \cap_{n=1}^{\infty} \overline{B}_n) \overset{\text{DeMorgan } Eq. \text{ (D.2.1)}}{=} \cup_{n=1}^{\infty}(\overline{\mathbb{R}} - \overline{B}_n)$. 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 Eq. (D.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)):

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

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. □

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-Stietljes measure on $\mathbb{R}$.*

Proof. See text. □

Remark 5.2.2. The proof of Theorem 5.2.1 essentially directly applies Caratheódory'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 Eq. (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,

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

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

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

□

Remark 5.3.1. (*Distribution is continuous at a point iffi the point has zero measure*)

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 4 puts positive probability mass on the points $\{x_1\}, \{x_2\}, \{x_3\}$ and zero probability mass on all other points.

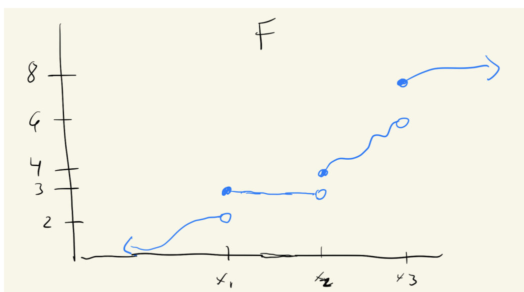


Figure 4: A distribution function with positive mass on points that is not concentrated on a countable set

△

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 Eq. (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 $\pm\infty$. △

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.[17] 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. △

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 $\mu(a,b] = \mu(a,b) = \mu[a,b) = \mu[a,b]$. △

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 $\mu(a,b] = \mu(a,b) = \mu[a,b) = \mu[a,b]$. △

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,

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

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 $(\mathbb{R})$.[18] △

[17] See Theorem 3.3.1.

[18] 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. The term "Lebesgue measure" can be used to refer to

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

as well as

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

[Folland, 1999, pp. 37],[Ash et al., 2000]. △

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 $\mathbb{R}^n$.

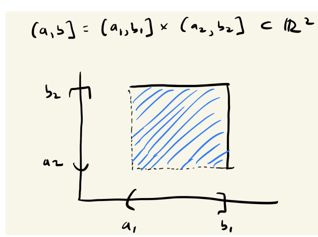
5.6.2 Definitions

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 rectangle 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

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

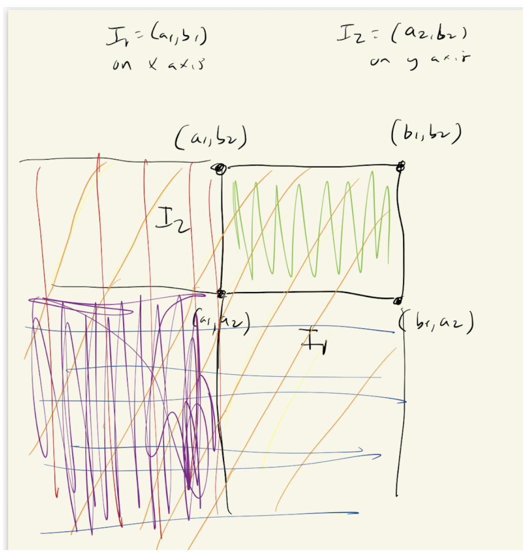


Figure 5: Using a distribution function in $\mathbb{R}^2$ to measure the box $I_1 \times I_2$.

Now consider Figure 5. 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 Eq. (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

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

□

Notation 5.6.1. What I call $\Delta_{(a,b]}F$ in Eq. (5.6.4) is called $F(a,b]$ by [Ash et al., 2000]. See e.g. [Ash et al., 2000, pp.28, or pp.149]. I find Ash's notation completely puzzling. △

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 6.

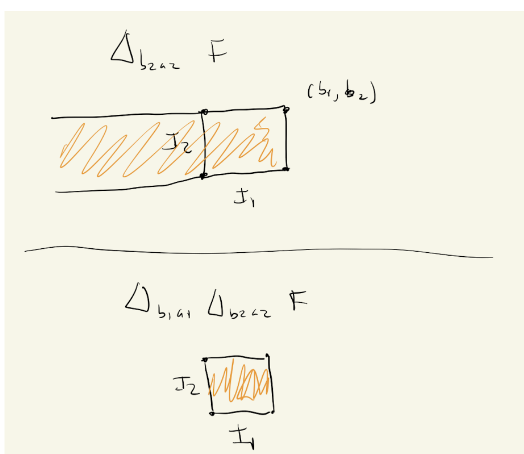


Figure 6: Repeated applications of the difference operator to a distribution function in $\mathbb{R}^2$.

△

Remark 5.6.2. Eq. (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 $i = 1, ..., n$.

$\triangle$

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]. $\triangle$

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. $\square$

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 Eq. (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.[19] 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 functions (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., $\mathcal{F}_0 := \{$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 $\mathcal{F}_0 := \{$disjoint unions of $(a, b], a, b \in \mathbb{R}^n\}$.

[19]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.[20]

- *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

$$\begin{aligned}
\mu(B) - \mu(K) &= \mu(B - K) && \text{piece-and-difference, } \mu \text{ finite}\\
&\overset{1}{\leq} \mu(\cup_{n=1}^{\infty}(B_n - K_n)) && \text{DeMorgan, monotonicity}\\
&\leq \sum_{n=1}^{\infty} \mu(B_n - K_n) && \text{countable subadditivity}\\
&= \sum_{n=1}^{\infty} \mu(B_n) - \mu(K_n) && \text{piece-and-difference, } \mu \text{ finite}\\
&= \epsilon
\end{aligned}$$

[20] 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 $\mathcal{F}_0 :=$ {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 $\sigma(\mathcal{F}_0) = \mathcal{B}(\mathbb{R}^n)$.[21]

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)[22], 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 $B \in (\mathbb{R}^n)$ and $c \in \mathbb{R}^n$, then $B + c \in (\mathbb{R}^n)$ 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 $(\mathbb{R}^n)$ to the reader. We shall use the Monotone Class Theorem as our vehicle for executing the Good Sets Strategy (see Remark

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

[22] 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.[23]
   b) Now we show that $\mathcal{G}$ contains $\mathcal{F}_0$. The property certainly holds for r.s.c intervals; that is, $\mu(a+c, b+c] = \mu(a,b]$. [25] For Lebesgue measure on $\mathbb{R}^n$, the distribution function is

[23] In particular, we write $\Omega = \cup_{n=1}^{\infty} \Omega_n$ where $\mu(\Omega_n) < \infty$.[24] 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
$$\mu(B+c) \overset{(*)}{=} \sum_n \mu_n(B+c) \overset{(+)}{=} \sum_n \mu_n(B) \overset{(*)}{=} \mu(B)$$

[25] 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 $x - y \in \mathbb{Q}$.*[26] *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

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

So

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

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

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

□

[26] 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. △

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[27]

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

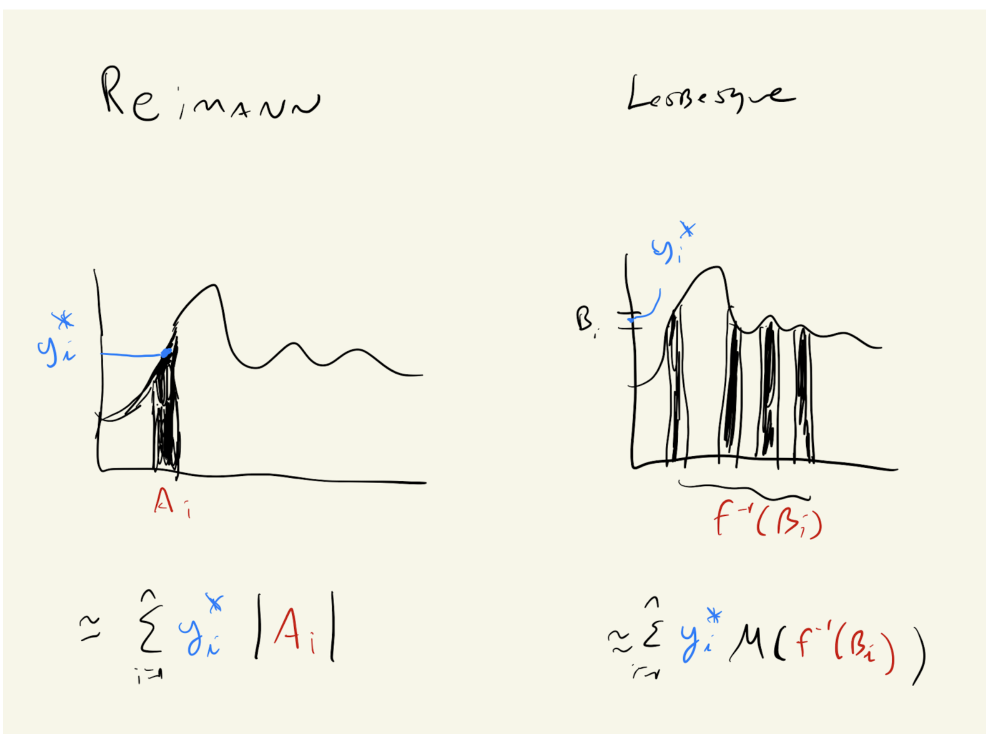


Figure 7: Illustrating the fundamental differences between Riemann and Lebesgue integration

[27] 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:

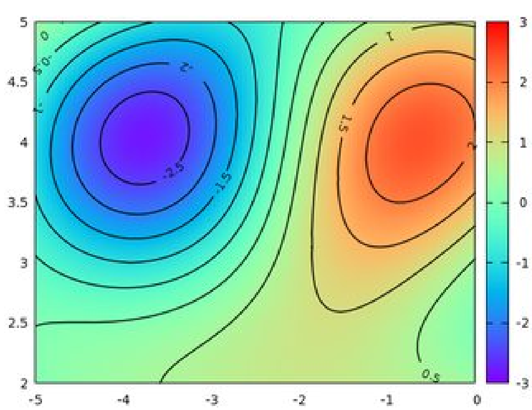
   > I have to pay a certain sum, which I have collected in my pocket. I take the bills and coins out of my pocket and give them to the creditor in the order I find them until I have reached the total sum. This is the Riemann integral. But I can proceed differently. After I have taken all the money out 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*. △

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: $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$. △

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 $h : (\Omega_1, \mathcal{F}_1) \to (\overline{\mathbb{R}}^n, \mathcal{B}(\overline{\mathbb{R}}^n))$. △

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 $B \in \mathcal{C} : \sigma(\mathcal{C}) = \mathcal{F}_2$. △

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.

□

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 $a \in \mathbb{R}$.

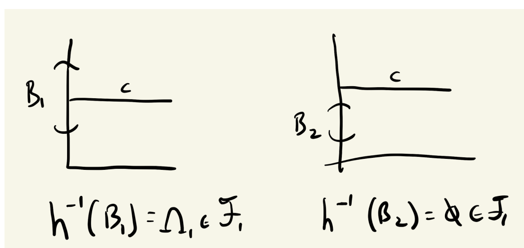
- $h^{-1}\big([a,\infty)\big) \in \mathcal{F}_1$ for each $a \in \mathbb{R}$.
- $h^{-1}\big((a,b)\big) \in \mathcal{F}_1$ for each $a, b \in \mathbb{R}$.
- $h^{-1}\big([a,b]\big) \in \mathcal{F}_1$ for each $a, b \in \mathbb{R}$.
- $h^{-1}(U) \in \mathcal{F}_1$ for each open set $U \subset \mathbb{R}$.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

For a larger list, recall Section 3.1.4. $\triangle$

6.2.3 Examples

Example 6.2.1. (*Constant functions are 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 measurable, since

$$h^{-1}(B) = \begin{cases} \Omega_1, & \text{if } c \in B \\ \emptyset, & \text{if } c \notin B \end{cases}$$



$\triangle$

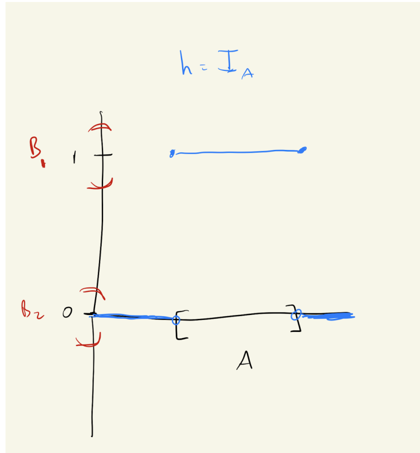
Example 6.2.2. (*Any function is measurable with respect to the trivial σ-field.*) Consider any function $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$, where $\mathcal{F}_2$ is the trivial σ-field: $\mathcal{F}_2 = \{\emptyset, \Omega_2\}$. Then h is measurable since $h^{-1}(\emptyset)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

TODO: Add: If Ω is countable, $\mathcal{F} = 2^{\Omega}$ (i.e. the set of all subsets of Ω) is a sigma-field, and hence any function is measurable. This construction comes up when viewing series as integrals against counting measure.

Example 6.2.3. (*Indicators of Borel sets are Borel measurable.*) Let A be a Borel subset of $\mathbb{R}$,[28] and let $I_A : \mathbb{R} \to \mathbb{R}$ be the indicator of A; that is $I_A(x) = 1$ for $x \in A$ and 0 for $x \notin A$. Then I_A is Borel measurable, since for all $B \in \mathcal{B}(\mathbb{R})$, we have

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

[28] 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.



△

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. △

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. △

6.2.4 Why define measurability this way?

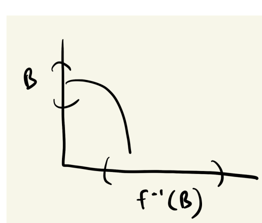
Measurable functions do *not* preserve measurability in *both* directions. That is, if $h : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ is a measurable function, it is not necessarily true that $h(A) \in \mathcal{F}_2$ for all $A \in \mathcal{F}_1$. For a counterexample, we take $\mathcal{F}_2 = \{\emptyset, \Omega_2\}$, 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. See Remark 6.2.2

Remark 6.2.2. In particular, for any function f

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

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



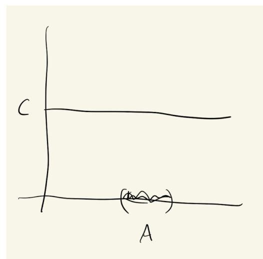
b) Inverse images and unions commute: $f^{-1}(\cup_i B_i) = \cup_i\big(f^{-1}(B_i)\big)$

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 $f(A) = c$ and $f(A^c) = c$. So $\mathbb{R} \setminus c = \big(f(A)\big)^c \neq f(A^c) = c$.



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

$\triangle$

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. $\square$

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*

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

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. $\square$

Proposition 6.2.3. *A composition of Borel measurable functions is measurable.*

Proof. See [Ash et al., 2000] Lemma 1.5.7. $\square$

Below we see that Borel measurability of a multivariate function is equivalent to the Borel measurability of all the component functions.

Proposition 6.2.4. *[Ash et al., 2000, Thm. 1.5.8]. Let $h : \Omega \to \overline{\mathbb{R}}^n$. So $h = (h_1(\omega), \dots, h_n(\omega))$. Then h is Borel measurable iff h_i is Borel measurable for all $i = 1, \dots n$.*

Proof. First, some notation. Let $p_i : \overline{\mathbb{R}}^n \to \overline{\mathbb{R}}$ be the projection map taking $(x_1, \dots, x_n)$ to x_i. Then each component function can be written as $h_i = p_i \circ h$.

Now we prove each direction

- $\boxed{\implies}$. Let h be Borel measurable. We need to show that

  $$h_i^{-1}(B) \in \mathcal{B}(\overline{\mathbb{R}}), \quad \forall B \in \mathcal{B}(\overline{\mathbb{R}}).$$

  By the computational def. of measurable functions (see Remark 6.2.1), it suffices to show that

  $$h_i^{-1}(I) \in \mathcal{B}(\overline{\mathbb{R}}), \quad \forall I = [a, b] \subset \overline{\mathbb{R}}.$$

  So note that each p_i is Borel measurable, since

  $$p_i^{-1}\{x_i \in \overline{\mathbb{R}} : a_i \le x_i \le b_i\} = \{x \in \overline{\mathbb{R}}^n : a_i \le x_i \le b_i, -\infty \le x_j \le \infty, \quad j \ne i\}$$

  which is an interval of $\overline{\mathbb{R}}^n$. So $h_i = p_i \circ h$ is the composition of measurable functions, which is measurable by Proposition 6.2.3.

- $\boxed{\impliedby}$. Let each h_i be Borel measurable. Again using Remark 6.2.1, we have

  $$h^{-1}\{x \in \overline{\mathbb{R}}^n : a \le x \le b\} = \bigcap_{i=1}^{n} \{\omega \in \Omega : a_i \le h_i(\omega) \le b_i\} \underbrace{\in \mathcal{F}}_{\text{by hypoth.}} .$$

$\square$

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:

$$\int_\Omega h \, d\mu \, , \qquad \int_\Omega h(\omega) \, d\mu(\omega) \, , \quad \text{or} \quad \int_\Omega h(\omega)\mu(d\,\omega) \tag{6.3.1}$$

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. $\triangle$

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.

$\triangle$

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. $\triangle$

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

$$\int_\Omega h \; d\mu \; := \sum_{i=1}^{r} y_i \; \mu(A_i). \tag{6.3.2}$$

$\triangle$

The integral of a simple function can also be expressed as

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

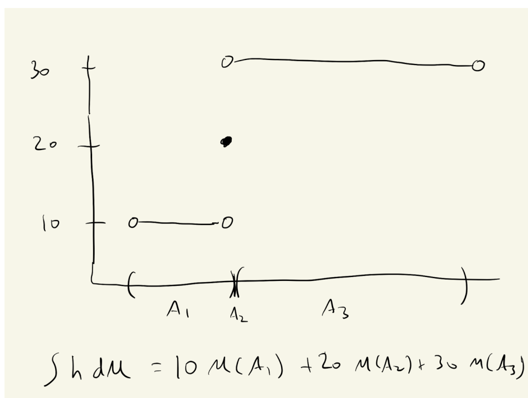
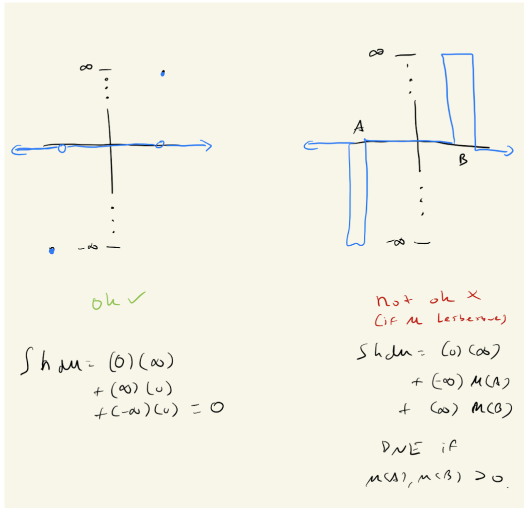


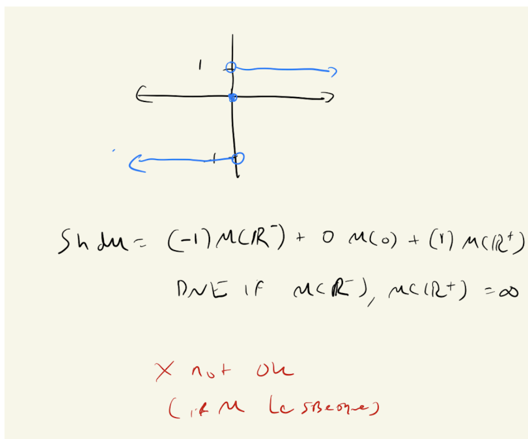
Figure 8: The Lebesgue integral of a simple function. (In this case, the simple function is also a step function.)

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 $A \in \mathcal{F}$ and simple function s, we define

$$\int_A s \, d\mu \; := \int_\Omega s \, 1_A \, d\mu$$

This definition is possible because whenever s is a simple function, then so is $s1_A$ for any measurable set A. Indeed, if we express $s = \sum_{i=1}^r y_i 1_{B_i}$, then $s1_A = \sum_{i=1}^r y_i 1_A 1_{B_i} = \sum_{i=1}^r y_i 1_{A \cap B_i}$, 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 $\Omega := \mathrm{dom}(h)$, and h takes on values y_i on

(x_{i-1}, x_i) for $i = 1, ..., n$. Then

$$\int_\Omega h \; d\mu \; = \sum_{i=1}^n y_i \; \mu(x_{i-1}, x_i) + \sum_{i=1}^n f(x_i)\mu\{x_i\}$$

$$\overset{\text{(if } \mu \text{ is Lebesgue measure)}}{=} \sum_{i=1}^n y_i \; (x_i - x_{i-1}) + \cancelto{0}{\sum_{i=1}^n f(x_i)\mu\{x_i\}}$$

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

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.

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

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.[29] 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.[30] Thus f is not Riemann integrable.

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

[30] To be more explicit, the condition of Proposition B.0.1 is that $\forall \epsilon > 0, \; \exists \delta > 0 : \forall P$ 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: $\exists \epsilon = \frac{b-a}{2} > 0 : \forall P, \text{Osc}(f, P) > \epsilon$.

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 $g \le h$ then* $\displaystyle\int g \; d\mu \;\le \int h \; d\mu$.

b) *(additivity)* $\displaystyle\int (g+h) \; d\mu \; = \int g \; d\mu \; + \int h \; d\mu$.

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

$$g = \sum_{i=1}^{r}\sum_{j=1}^{s} x_i I_{A_i \cap B_j}, \qquad h = \sum_{i=1}^{r}\sum_{j=1}^{s} y_j I_{A_i \cap B_j} \tag{6.3.3}$$

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

So

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

b) We sum the two forms in Eq. (6.3.3) and apply the definition of integral of simple functions to obtain

$$\int (g+h) \; d\mu \; = \sum_{i=1}^{r}\sum_{j=1}^{s} (x_i + y_j) \, \mu(A_i \cap B_j) \tag{6.3.4}$$

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

$$
\begin{aligned}
\int g \; d\mu \; &= \sum_{i=1}^{r} x_i \mu(A_i) = \sum_{i=1}^{r} x_i \Big(\sum_{j=1}^{s} \mu(A_i \cap B_j) \Big) \\
\int h \; d\mu \; &= \sum_{j=1}^{s} y_i \mu(B_i) = \sum_{j=1}^{s} y_i \Big(\sum_{i=1}^{r} \mu(A_i \cap B_j) \Big)
\end{aligned}
$$

and summing these together yields Eq. (6.3.4).

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 $\mathcal{F}$.

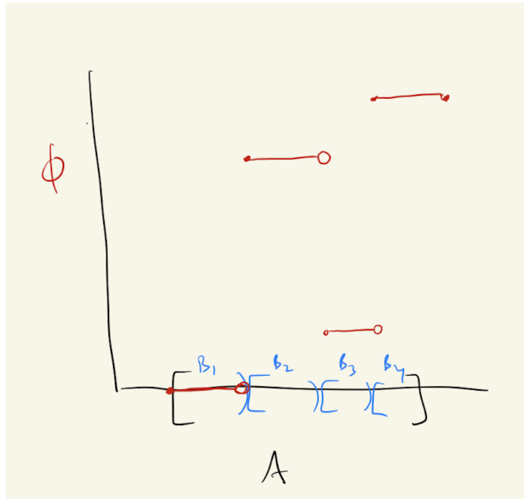
Proof. Weshow countable additivity.[31] So let $A = \dot\bigcup_{i=1}^{\infty} A_i$. Writing the simple function as $s = \sum_{j=1}^{r} y_j 1_{B_j}$, we obtain

$$
\begin{aligned}
\int_A s \; d\mu \; &= \sum_{j=1}^{r} y_j \; \mu(B_j \cap A) && \text{Integrals of simple functions over subsets (see Remark 6.3.4)} \\
&= \sum_{j=1}^{r} \sum_{i=1}^{\infty} y_i \; \mu(B_j \cap A_i) && \text{countable additivity – note } \Big(B_j \cap \dot\bigcup_{i=1}^{\infty} A_i \Big) = \dot\bigcup_{i=1}^{\infty} (B_j \cap A_i) \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{r} y_i \; \mu(B_j \cap A_i) && \text{sum of limit is limit of sum} \\
&= \sum_{i=1}^{\infty} \int_{A_i} s \; d\mu && \text{Remark 6.3.4}
\end{aligned}
$$

□

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.

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



△

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. △

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

$$\int_\Omega h \, d\mu = \sup \left\{ \int_\Omega s \, d\mu \; : s \quad \text{simple}, \quad 0 \le s \le h \right\} \tag{6.3.5}$$

△

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.[32]

[32]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 $a = \int_\Omega h \, d\mu$.

Figure 9: 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. $\triangle$

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)* $\displaystyle\int f \leq \int g$ *whenever* $f \leq g$

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

Proof. a)

$$\begin{aligned} f \leq g &\implies \{s : s \text{ simple }, 0 \leq s \leq f\} \subset \{s : s \text{ simple }, 0 \leq s \leq g\} \\ &\implies \{\int s\ d\mu\ : s \text{ simple }, 0 \leq s \leq f\} \subset \{\int s\ d\mu\ : s \text{ simple }, 0 \leq s \leq g\} && \text{monotonicity for simple fns} \\ &\implies \sup\{\int s\ d\mu\ : s \text{ simple }, 0 \leq s \leq f\} \leq \sup\{\int s\ d\mu\ : s \text{ simple }, 0 \leq s \leq g\} && \text{Prop. A.2.1} \\ &\implies \int f\ d\mu\ \leq \int g\ d\mu && \text{def. integral for non-negative fns} \end{aligned}$$

b)

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

$\square$

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

Proof. Define

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

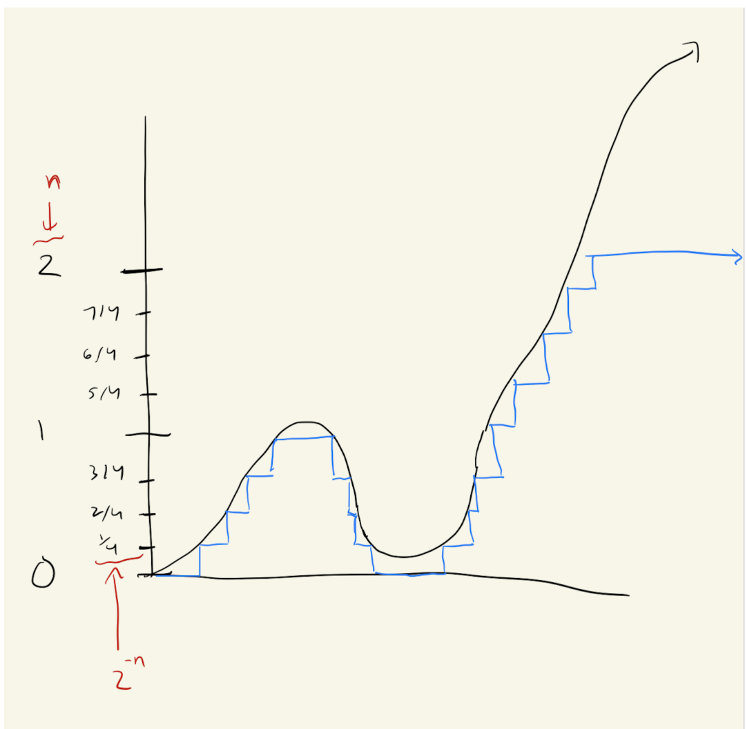


Figure 10: An element of an increasing sequence of simple functions approximating an arbitrary non-negative Borel measurable function. The nth function in the sequence has a maximum value of n and divides the range into bins of size 2^{-n}.

Then

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

$\square$

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? $\triangle$

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 $h_n \leq h_{n+1}$ for all n and $h = \lim_{n\to\infty} h_n (= \sup_n h_n)$, then*

$$\int h_n \, d\mu \uparrow \int h \, d\mu \tag{6.3.6}$$

Proof. We break the proof into four parts. Consider

$$\int h \, d\mu = \lim_{n\to\infty} \int h_n \, d\mu \tag{6.3.7}$$

- We first show that both quantities in Eq. (6.3.7) 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 Eq. (6.3.7); 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 Eq. (6.3.7); 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

  $$\int h_n \overset{\text{monotonicity}}{\geq} \int_{A_n} h_n \overset{\text{def. } A_n\text{, monotonicity}}{\geq} \int_{A_n} \alpha s \overset{\text{linearity}}{=} \alpha \int_{A_n} s \tag{6.3.8}$$

  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 Eq. (6.3.8) 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 Eq. (6.3.7) holds. It just remains to show that Eq. (6.3.7) $\Longrightarrow$ Eq. (6.3.6). 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.3.1), $\int h \, d\mu \geq \int h_n \, d\mu$ 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[33]) 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. ***Additivity of integration with non-negative Borel measurable functions.*** *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

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

□

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*

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

[33] TODO: Justify that each f_n is Riemann integrable.

Then

$$\int_\Omega f \, d\mu = \sum_{n=1}^{\infty} \int_\Omega f_n \, d\mu \tag{6.3.9}$$

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 Eq. (6.3.9) 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 Eq. (6.3.9) is

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

△

Remark 6.3.12. (*Linearity of integration with non-negative Borel measurable functions.*) The scalar multiple property is given by Prop. 6.3.3, and the additivity is given by Prop. 6.3.5. △

6.3.3 Integrals of arbitrary Borel measurable functions

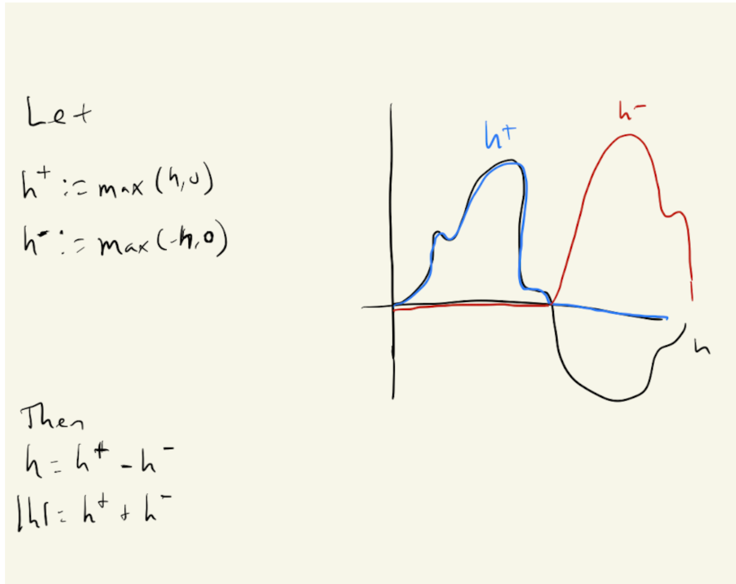
Let h be an arbitrary Borel measurable function. We will express an arbitrary Borel measurable function as as the difference of two non-negative Borel measurable functions.

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*)

$$\int_\Omega h \; d\mu \; = \int_\Omega h^+ \; d\mu \; - \int_\Omega h^- \; d\mu \tag{6.3.10}$$

△

Remark 6.3.13. (*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 6.3.14. (*Integrals on subsets*) For $A \in \mathcal{F}$, we define

$$\int_A h \, d\mu \; = \int_\Omega h 1_A \, d\mu$$

This definition works because whenever h is measurable, then so is $h1_A$:

$$\{\omega : h1_A(\omega) < c\} = \underbrace{\{\omega : h(\omega) < c\}}_{\in \mathcal{F} \text{ since } h \text{ measurable}} \cap \underbrace{\{\omega : \omega \in A\}}_{\in \mathcal{F} \text{ by assumption}}$$

$\triangle$

Example 6.3.3. (*Series as integrals against counting measure.*) [Ash et al., 2000, pp.89 and Problem 1a (real part only) pp.94]. Let $\Omega = \{1, 2, 3, ...\}$, $\mathcal{F} = 2^\Omega$ (i.e. all subsets of Ω), and μ be the counting measure. A real-valued function f on Ω can be written as a sequence of real numbers; we write $f = \{a_n\}, n = 1, 2, \ldots$. We will show that an integral on this space is really a sum:

$$\int_\Omega f \, d\mu \; = \sum_{n=1}^{\infty} a_n \tag{6.3.11}$$

where the series is interpreted as $\sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-$ if this is not of the form $\infty - \infty$ (if it is, the integral does not exist).[34] Note that every f is automatically measurable since $\mathcal{F} = 2^\Omega$.

To justify Eq. (6.3.11), let us first assume that $a_n \geq 0$ for each n. We define a sequence of non-negative functions $\{f_k\}$ by $f_k : n \mapsto f(n)1_{n \leq k}$, i.e. $f_k = (a_1, a_2, \ldots, a_k, 0, 0, \ldots)$.

We know how to integrate each f_k, since by the definition of integrals of simple functions, we have

$$\int f_k \, d\mu \; = \sum_{i=1}^{k} a_i \tag{6.3.12}$$

Since $f_k \uparrow f$, we apply Monotone Convergence Theorem to obtain

$$\int f \, d\mu \overset{\text{def } f}{=} \int \lim_{k \to \infty} f_k \, d\mu \overset{MCT}{=} \lim_{k \to \infty} \int f_k \, d\mu \overset{\text{Eq. (6.3.12)}}{=} \lim_{k \to \infty} \sum_{i=1}^{k} a_i = \sum_{i=1}^{\infty} a_i \tag{6.3.13}$$

Relaxing the non-negativity assumption to allow $a_n \in \mathbb{R}$, we then have

$$\begin{aligned} \int f \, d\mu \; &= \int f^+ \, d\mu \; - \int f^- \, d\mu && \text{Def. integral of arbitrary measurable functions} \\ &= \sum_{i=1}^{\infty} a_i^+ - \sum_{i=1}^{\infty} a_i^- && \text{Result with non-negative functions, Eq. (6.3.13)} \end{aligned}$$

which exists whenever this does not take the form $\infty - \infty$. $\triangle$

Remark 6.3.15. (*Conditionally convergent sums are not integrals.*) When summation is considered from the point of Lebesgue integration theory, series that converge conditionally but not absolutely are ignored.[35]

To see this, note that the expression $\int f \, d\mu \; = \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^-$ from Eq. (6.3.11) yields four cases

---

[34] Recall that $a_n^+ := \max(a_n, 0)$ and $a_n^- := \max(-a_n, 0)$.

[35] The alternating harmonic series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}$ is an example of a series that converges conditionally but not absolutely. It is (potentially) easier for series with both negative and positive terms to converge, because terms with different signs may partially cancel or compensate. See `https://www.sfu.ca/math-coursenotes/Math%20158%20Course%20Notes/sec_AbsoluteConvergence.html`.

1. $\sum_{n=1}^{\infty} a_n^+ = \infty, \sum_{n=1}^{\infty} a_n^- < \infty$. The series diverges to ∞ and the integral is ∞.
2. $\sum_{n=1}^{\infty} a_n^+ < \infty, \sum_{n=1}^{\infty} a_n^- = \infty$. The series diverges to $-\infty$ and the integral is $-\infty$.
3. $\sum_{n=1}^{\infty} a_n^+ < \infty, \sum_{n=1}^{\infty} a_n^- < \infty$. The series is absolutely convergent and the integral equals the sum of the series.
4. $\sum_{n=1}^{\infty} a_n^+ = \infty, \sum_{n=1}^{\infty} a_n^- = \infty$. The series is not absolutely convergent; it may or may not converge conditionally. Whether it does or not, the integral does not exist.

$\triangle$

6.3.4 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 $\displaystyle\int g \, d\mu \geq \int h \, d\mu$. Moreover, if $g \geq h$, $\displaystyle\int h \, d\mu$ exists and $\displaystyle\int h \, d\mu > -\infty$, then $\displaystyle\int g \, d\mu$ exists. And if $g \geq h$, $\displaystyle\int g \, d\mu$ exists and $\displaystyle\int g \, d\mu < \infty$, then $\displaystyle\int h \, d\mu$ exists.*[36]

c) *(Triangle inequality for integrals.) If $\int_\Omega h \, d\mu$ exists, then $| \int_\Omega h \, d\mu | \leq \int_\Omega |h| \, d\mu$.*

d) *(Delayed truncation of simple functions.) If $h \geq 0$ and $B \in \mathcal{F}$, then*[37]

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

e) *(Existence of integral transfers to subsets.)*

$$\int_\Omega h \, d\mu \textit{ exists} \implies \int_A h \, d\mu \textit{ exists} \quad \forall A \in \mathcal{F} \tag{6.3.14}$$

$$\int_\Omega h \, d\mu \textit{ finite} \implies \int_A h \, d\mu \textit{ finite} \quad \forall A \in \mathcal{F} \tag{6.3.15}$$

Proof. a) Since cf is a Borel measurable function[38], we apply Definition 6.3.4.

We have

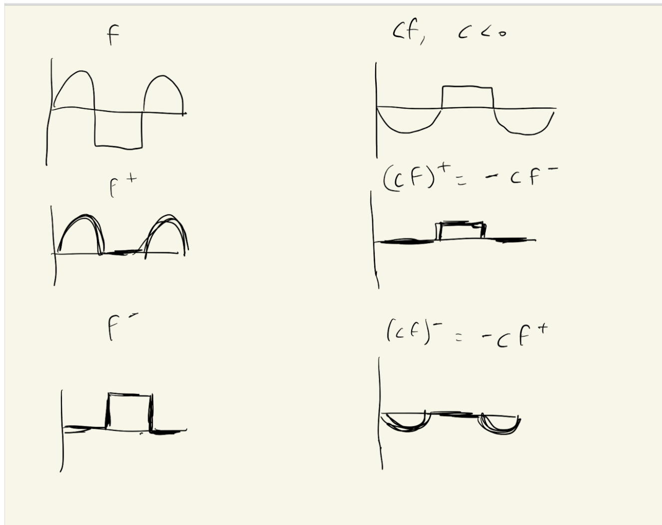
$$\text{if } c \geq 0, \qquad (cf)^+ = cf^+ \qquad (cf)^- = cf^- \tag{6.3.16a}$$

$$\text{if } c < 0, \qquad (cf)^+ = -cf^- \qquad (cf)^- = -cf^+. \tag{6.3.16b}$$

[36] We might consider this as a “dominance criterion for existence." For more on why the monotonicity statement is concerned with existence, see Remark 6.3.16.

[37] For why this needs to be proven, see Remark 6.3.17.

[38] 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.



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)).

So if $c \geq 0$

$$\begin{aligned}
\int cf \, d\mu &= \int (cf)^+ \, d\mu - \int (cf)^- \, d\mu && \text{Def. 6.3.4} \\
&= \int cf^+ \, d\mu - \int cf^- \, d\mu && \text{Eq. (6.3.16a)} \\
&\overset{*}{=} c \int f^+ \, d\mu - c \int f^- \, d\mu && \text{Prop. 6.3.3 (b)} \\
&= c \int f \, d\mu && \text{Def. 6.3.4}
\end{aligned}$$

Likewise if $c < 0$

$$\begin{aligned}
\int cf \, d\mu &= \int (cf)^+ \, d\mu - \int (cf)^- \, d\mu && \text{Def. 6.3.4} \\
&= \int -cf^- \, d\mu - \int -cf^+ \, d\mu && \text{Eq. (6.3.16b)} \\
&\overset{**}{=} -c \int f^- \, d\mu + c \int f^+ \, d\mu && \text{Prop. 6.3.3 (b)} \\
&= c \int f \, d\mu && \text{Def. 6.3.4}
\end{aligned}$$

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

b) First we show that $g \geq h \implies \displaystyle\int g \, d\mu \geq \int h \, d\mu$ 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

$$\int g^+ \, d\mu \;\geq \int h^+ \, d\mu \;, \quad \int g^- \, d\mu \;\leq \int h^- \, d\mu \;. \tag{6.3.17}$$

So

$$\begin{aligned} \int g \, d\mu \;&= \int g^+ \, d\mu \;- \int g^- \, d\mu && \text{(def. integral; existence assumed)} \\ &\geq \int h^+ \, d\mu \;- \int h^- \, d\mu && \text{Eq. (6.3.17)} \\ &= \int g \, d\mu && \text{(def. integral; existence assumed)} \end{aligned}$$

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.[39]

c) We have $-|h| \leq h \leq |h|$. So by monotonicity and the scalar multiple property, $-\int_\Omega |h| \, d\mu \;\leq \int_\Omega h \, d\mu \;\leq \int_\Omega |h| \, d\mu$. By multiplying the left-hand inequality by -1 and keeping the right-hand inequality as is, we see

$$-\int_\Omega h \, d\mu \;\leq \int_\Omega |h| \, d\mu \quad \text{ and } \quad \int_\Omega h \, d\mu \;\leq \int_\Omega |h| \, d\mu$$

which taken together says $| \int_\Omega h \, d\mu \, | \leq \int_\Omega |h| \, d\mu$.

d) 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.

- $\boxed{\leq}$

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

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

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

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

So by monotonicity,

$$\underbrace{\int (h1_A)^+ \, d\mu}_{\text{A1}} \le \underbrace{\int h^+ \, d\mu}_{\text{B1}}$$
$$\underbrace{\int (h1_A)^- \, d\mu}_{\text{A2}} \le \underbrace{\int h^- \, d\mu}_{\text{B1}}$$

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.

□

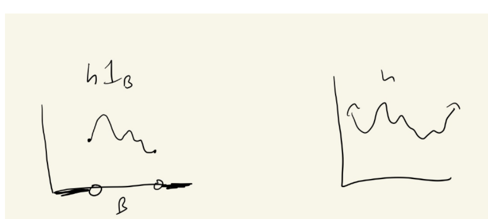
Remark 6.3.16. (*Why the monotonicity property is concerned with existence.*) Why is Proposition 6.3.6 b) concerned with monotonicity? Answer: even if $\int g \, d\mu$ exists and $g \ge h$, we can still have $\int h \, d\mu$ not exist, because of Eq. (6.3.17). For example, we can have

$$\int g^+ \, d\mu = \int h^+ \, d\mu = \infty \tag{6.3.18}$$
$$\int g^- \, d\mu < \int h^- \, d\mu = \infty \tag{6.3.19}$$

and so $\int h \, d\mu$ DNE. △

Remark 6.3.17. (*Why delayed truncation of simple functions is something that needs to be proven.*) Proposition 6.3.6 d) 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

$$\int_\Omega h \, d\mu = \sup\left\{ \int s \, d\mu \ : \ 0 \le s \le h1_B, \ s \text{ simple} \right\} \qquad \text{def. integral} \tag{6.3.20a}$$
$$\int_\Omega h \, d\mu = \sup\left\{ \int s1_B \, d\mu \ : \ 0 \le s \le h, \ s \text{ simple} \right\} \qquad \text{Prop. 6.3.6 d).} \tag{6.3.20b}$$

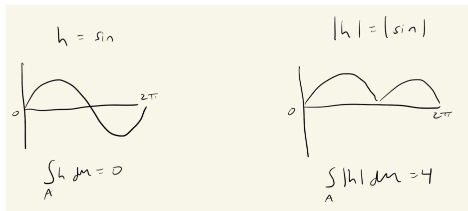


What does each say?

- Eq. (6.3.20a) : truncate first, then “simplify"
- Eq. (6.3.20b) “simplify" first, then truncate

. △

Example 6.3.4. (*A simple example of triangle inequality for integrals.*) Let $h(x) = \sin(x)$, $A = [0, 2\pi]$, and μ be Lebesgue measure. Then



The basic idea is that integrating over a function h that can take on both positive and negative values can lead to cancelations, which explains the result. △

Proposition 6.3.7. ***(A Borel measurable function is the limit of a sequence of simple functions which it dominates.)*** *An arbitrary Borel measurable function f is the limit of a sequence of finite-valued simple functions s_n, with $|s_n| < |f|$ for all n.*

Proof. This follows straightforwardly from Prop 6.3.4. Write $f = f^+ - f^-$. By Prop 6.3.4, we can find sequences of simple functions $(s_n^+)_n$ and $(s_n^-)_n$ such that $s_n^+ \uparrow f^+$ and $s_n^- \uparrow f^-$. Therefore,

$$\begin{aligned} s_n \triangleq (s_n^+ - s_n^-) \to f^+ - f^- = f \qquad & \text{by additivity of limits} \\ |s_n| \leq |f| \text{ for all } n \qquad & \text{by definition of absolute value} \end{aligned}$$

□

7 § 1.6 Basic Integration Theorems

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

7.1 Indefinite integrals as countably additive set functions

Definition 7.1.1. We say that λ is an *indefinite integral*[40] with respect to μ if for any $A \in \mathcal{F}$ we have

$$\lambda(A) = \int_A f \, d\mu \tag{7.1.1}$$

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

Theorem 7.1.1. ***Indefinite integrals are countably additive set functions.*** *Let f be a Borel measurable function on $(\Omega, \mathcal{F})$ 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 $f \geq 0$, 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 $\lambda(B) = \sum_{n=1}^{\infty} \lambda(B_n)$.

[40] 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 d), 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 d) 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

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

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 7.1.1. ***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 7.1.1. The non-negativity of λ is immediate from Definition 6.3.3. □

Remark 7.1.1. Indefinite integrals are signed measures. As will become clear in Section 9, Theorem 7.1.1 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 7.1.2. Change of measure and differential notation.

By Cor. 7.1.1, given Borel measurable $f \geq 0$, the indefinite integral

$$\lambda(A) = \int_A f \, d\mu \tag{7.1.2}$$

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]:

$$d\lambda = f \, d\mu \tag{7.1.3}$$

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]:

$$\int_\Omega g \, d\lambda \; = \int_\Omega gf \, d\mu \tag{7.1.4}$$

for every measurable function g on Ω.[41] (See Proposition 7.2.1 for a proof.)

As pointed out by [Rudin, 1987, pp. 24], we assign no independent meaning to the symbols $d\lambda$ and $d\mu$; Eq. (7.1.3) simply means that Eq. (7.1.2) (and therefore Eq. (7.1.4)) holds for every measurable $f \geq 0$.

$\triangle$

Example-for-data-scientists 7.1.1. (*Jeffreys prior for observation noise in linear regression*) An example for statisticians: Jeffreys prior for observation noise in linear regression (see e.g. [Carvalho et al., 2010] or [Makalic and Schmidt, 2015]) is sometimes written as

$$\sigma^2 \sim \sigma^{-2} \, d\sigma^2 \tag{7.1.5}$$

What does this mean? First off, the $\sim$ notation means that the random variable σ^2 has the probability distribution P described by

$$dP = \sigma^{-2} \, d\mu(\sigma^2)$$

where μ is Lebesgue measure. This is differential notation of the form Eq. (7.1.3). Unpacking back to Eq. (7.1.2), this is saying that

$$P(A) = \int_A \sigma^{-2} \, d\mu \; = \int_A \sigma^{-2} \, d\sigma^2$$

for all measurable sets A. $\triangle$

7.2 Additivity theorem

Theorem 7.2.1. ***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*

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

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

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

Proof. See [Ash et al., 2000], Theorem 1.6.3. □

Remark 7.2.1. (*Additivity holds automatically for integrable functions.*) If f and g are integrable, the conditions of Theorem 7.2.1 are always met.

Moreover, in this situation, the proof of Theorem 7.2.1 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

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

Rearranging (possible by integrability), we get

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

△

Non-example 7.2.1. Let us demonstrate where the Additivity Theorem fails to apply. Let $f \equiv 1, g \equiv -1$ and μ be Lebesgue measure. Then $\int f \, d\mu = \infty$ and $\int g \, d\mu = -\infty$. But

$$\underset{0}{\int (f+g)} \quad d\mu \neq \underset{\infty}{\int f \, d\mu} + \underset{-\infty}{\int g \, d\mu}$$

Because $\infty - \infty$ is undefined. (To reinforce the undefinedness of $\infty - \infty$, note that the LHS could be $0, \infty$ or $-\infty$ by setting $f \equiv a$, $g \equiv -b$, by choosing $a < b$, $b > a$, or $a = b$.) △

Remark 7.2.2. (*From additivity to linearity.*) 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. △

Proposition 7.2.1. *(Change of differential.)*[42] *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $f \geq 0$ a non-negative Borel measurable function on Ω. Recalling Cor. 7.1.1, 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

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

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

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 $\int s_n f d\mu \uparrow \int f g d\mu$. 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 $\int g d\lambda = \int g f d\mu$.

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

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

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. 7.2.1) can be applied.

$\square$

Corollary 7.2.1. *(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.

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

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

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

$$\left| \int |h| \, d\mu \right| \overset{\textit{additivity (Theorem 7.2.1)}}{=} \left| \int h^+ \, d\mu \; + \int h^- \, d\mu \right| < \infty$$

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

b)

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

$\square$

7.3 Almost everywhere theorems

Definition 7.3.1. 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. $\triangle$

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 7.3.1. ***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

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

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 7.2.2) 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

$$h = h 1_A + h 1_{A^c} \overset{\text{def. } A}{=} g 1_A + h 1_{A^c} \tag{7.3.1}$$
$$g = g 1_A + g 1_{A^c} \tag{7.3.2}$$

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

$$\int_{A^c} g\, d\mu\ = 0, \quad \int_{A^c} h\, d\mu\ = 0 \tag{7.3.3}$$

And so we can apply additivity to Eq. (7.3.1) and Eq. (7.3.2), since:

- $\int g\ d\mu$, $\int h\ d\mu$ exist, so since existence transfers to subsets (see Prop. 5.3.1e)), $\int_A g\ d\mu$, $\int_{A^c} g\ d\mu$, $\int_A h\ d\mu$, $\int_{A^c} h\ d\mu$ exist.
- By Eq. (7.3.3),

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

So applying linearity to Eq. (7.3.1) and Eq. (7.3.2), we get

$$\int h\ d\mu\ = \int_A g\ d\mu\ + \cancelto{0}{\int_{A^c} h\ d\mu} \qquad \text{cancelation by Eq. (7.3.3).}$$
$$\int g\ d\mu\ = \int_A g\ d\mu\ + \cancelto{0}{\int_{A^c} g\ d\mu} \qquad \text{cancelation by Eq. (7.3.3).}$$

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

□

Remark 7.3.1. Thanks to Theorem 7.3.1, 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 7.3.1 b), $\int_\Omega h_n^*\ d\mu\ = \int_\Omega h_n\ d\mu$ for all n, and so the conclusion holds.)

△

Theorem 7.3.2. *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 $|h1_B| = \infty$. Then

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

So $|h|$ is not integrable. So by Corollary 7.2.1 b), h is not integrable.

b) Let $B_n := \{\omega : h(\omega) \geq \frac{1}{n}\}$. Then $B_n \uparrow B := \{\omega : h(\omega) > 0\}$.[43] ①

Now

$$0 \overset{\textit{first hypothesis}}{\leq} h1_{B_n} \overset{B_n \subset B}{\leq} h1_B \overset{\textit{def. } B}{=} h$$

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

$$\overset{\textit{first hypothesis, monotonicity}}{\Longrightarrow} \int_{B_n} h \; d\mu = 0 \qquad ②$$

Now ① and monotonicity again give

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

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

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

□

We can now construct a converse to monotonicity.

Theorem 7.3.3. ***Monotonicity Converse.*** *If μ is σ-finite on $\mathcal{F}$, g and h are Borel measurable, $\int_\Omega g \; d\mu$ and $\int_\Omega h \; d\mu$ exist, and $\int_A g \; d\mu \leq \int_A h \; d\mu$ for all $A \in \mathcal{F}$, then $g \leq h$ a.e. $[\mu]$.*

Proof. We prove the theorem assuming that at least one of $\{\int_\Omega g \; d\mu, \int_\Omega h \; d\mu\}$ is finite (so that we can apply the Additivity Theorem). Note that in this special case, we need not assume that μ is σ-finite on $\mathcal{F}$. For a full proof, see [Ash et al., 2000] Theorem 1.6.11.

Let $A := \{x : g(x) < h(x)\}$. (Note that A is the set where the desired conclusion fails.) Then

$$\int_A g \; d\mu \overset{\textit{monotonicity on } A}{\leq} \int_A h \; d\mu \overset{\textit{hypothesis}}{\leq} \int_A g \; d\mu$$

So by sandwiching,

$$\int_A g \; d\mu = \int_A h \; d\mu .$$

Now

$$0 \overset{\textit{subtraction}}{=} \int_A h \; d\mu - \int_A g \; d\mu \overset{\textit{linearity}}{=} \int_A (h-g) \; d\mu$$

where linearity applies under the assumptions of the theorem. (It wouldn't apply if $\int_A g \; d\mu = \infty$, $\int_A h \; d\mu = -\infty$, or vice versa, which would violate the conditions of the additivity theorem.)

So

$$0 = \int \underbrace{(h-g)\,1_A}_{\text{non-negative by def } A} \; d\mu \overset{\textit{Theorem 7.3.2 b)}}{\Longrightarrow} (h-g)1_A = 0 \text{ a.e.}$$

$$\implies (h-g) = 0 \quad \text{a.e. on } A$$

$$\implies h = g \quad \text{a.e. on } A$$

This gives an a.e. contradiction to the definition of A. So $\mu(A) = 0$. □

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

Remark 7.3.2. (*On the assumptions of the monotonicity converse.*) By monotonicity (Prop. 6.3.6), $g \le h$ implies $\int_\Omega g\ d\mu\ \le \int_\Omega h\ d\mu$, and in fact, $\int_A g\ d\mu\ \le \int_A h\ d\mu$ for all $A \in \mathcal{F}$ (This holds by monotonicity again, since $g \le h \implies g1_A \le h1_A$.) Now note that $\int_\Omega g\ d\mu\ \le \int_\Omega h\ d\mu \not\Rightarrow \int_A g\ d\mu\ \le \int_A h\ d\mu$ for all $A \in \mathcal{F}$. (To see this, consider that g, h are not necessarily non-negative. So for a counter-example, just take $g = \sin$ and $h = 0$ when integrating with respect to Lebesgue measure. See also Example 6.3.4.) So we obtain the converse by imposing the condition that the integral inequality holds over all measurable sets. △

Remark 7.3.3. In the proof of the monotonicity converse, we derived an "a.e. contradiction" to a set to show it has measure 0 seems like a fun proof technique. I haven't seen this before. △

Remark 7.3.4. This remark could be made in many places, but note from the proof of the monotonicity converse how convenient it is to work with Lebesgue integrals rather than than Riemann integrals for proving integral properties. We can define some set A in the domain to have any desired property, and then proceed to work with it. △

Corollary 7.3.1. *If μ is σ-finite on $\mathcal{F}$, g and h are Borel measurable, $\int_\Omega g\ d\mu$ and $\int_\Omega h\ d\mu$ exist, and $\int_A g\ d\mu\ = \int_A h\ d\mu$ for all $A \in \mathcal{F}$, then $g = h$ a.e. $[\mu]$.*

Proof. Since $\int_A g\ d\mu\ = \int_A h\ d\mu$ for all $A \in \mathcal{F}$, we have

$$\begin{aligned} \int_A g\ d\mu\ &\le \int_A h\ d\mu \overset{\text{Thm. 7.3.3}}{\Longrightarrow} g \le h \text{ a.e.} \\ \text{and} \int_A h\ d\mu\ &\le \int_A g\ d\mu \overset{\text{Thm. 7.3.3}}{\Longrightarrow} h \le g \text{ a.e.} \end{aligned}$$

Hence $g = h$ a.e. □

Remark 7.3.5. Cor. 7.3.1 gives a statent in Prop 2.2.3 b) of [Folland, 1999], but that statement assumes that both g and h are both integrable, whereas here we only need to assume that the integrals exist. △

7.4 Extended monotone convergence theorem

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

Theorem 7.4.1. ***Extended Monotone Convergence Theorem.*** *Let $f_1, f_2, ..., f, g$ be Borel measurable*

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

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

b) *If $f_n \downarrow f$ and $f_n \le g$ 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), $\int f\ d\mu\ \ge \int f_n\ d\mu\ \ge \int g\ d\mu$, 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

$$
\begin{aligned}
& f_n - g \geq 0 \quad \text{and} \quad f_n - g \uparrow f - g && \text{Hypothesis (and Prop A.3.1)} \\
\implies & \int (f_n - g)\, d\mu \uparrow \int (f - g)\, d\mu && \text{Monotone Convergence Theorem} \\
\overset{1}{\implies} & \int f_n\, d\mu - \int g\, d\mu \uparrow \int f\, d\mu - \int g\, d\mu && \text{Linearity} \\
\implies & \int f_n\, d\mu \uparrow \int f\, d\mu && \text{Since } \int g \text{ is finite}
\end{aligned}
$$

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

$$
\begin{aligned}
& -f_n \geq -g \quad \text{and} \quad -f_n \uparrow -g && \text{Hypothesis} \\
\implies & -\int f_n\, d\mu \uparrow -\int f\, d\mu && \text{Part (a) (and constant multiple property; Prop 6.3.6 a)} \\
\implies & \int f_n\, d\mu \downarrow \int f\, d\mu
\end{aligned}
$$

□

7.5 Fatou's Lemma

Theorem 7.5.1. ***Extended Fatou's Lemma***[44] *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

$$\int_\Omega \left(\liminf_{n \to \infty} f_n \right) d\mu \leq \liminf_{n \to \infty} \int_\Omega f_n\ d\mu \tag{7.5.1}$$

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

$$\int_\Omega \left(\limsup_{n \to \infty} f_n \right) d\mu \geq \limsup_{n \to \infty} \int_\Omega f_n\ d\mu \tag{7.5.2}$$

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).

[44]We refer to Theorem 7.5.1 as *extended* Fatou's lemma in parallel with the extended monotone convergence theorem (Theorem 7.4.1). 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 $g \equiv 0$.

Hence, by the extended Monotone Convergence Theorem (Thm. 7.4.1)

$$\underbrace{\lim_{n\to\infty}\int h_n\, d\mu}_{=\liminf_{n\to\infty}\int h_n\, d\mu} \overset{\text{(MCT)}}{=} \int \lim_{n\to\infty} h_n\, d\mu \qquad (1)$$

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 7.5.1. (*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

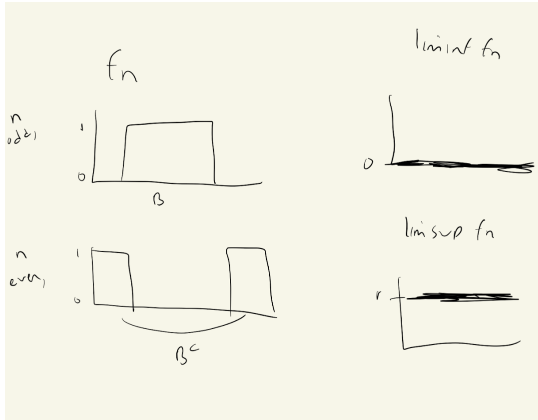
$$\int_\Omega \left(\liminf_{n\to\infty} f_n\right) d\mu \overset{\text{Eq. (7.5.1)}}{\leq} \liminf_{n\to\infty}\int_\Omega f_n\, d\mu \overset{\text{Eq. (A.5.1)}}{\leq} \limsup_{n\to\infty}\int_\Omega f_n\, d\mu \overset{\text{Eq. (7.5.2)}}{\leq} \int_\Omega \left(\limsup_{n\to\infty} f_n\right) d\mu \tag{7.5.3}$$

△

Example 7.5.1. (*Strict inequalities can occur in Fatou's lemma*) We show that that strict inequalities can occur in the expanded view of Fatou's lemma, Eq. (7.5.3).

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, Eq. (7.5.3).

$$\underbrace{\int_\Omega \left(\liminf_{n\to\infty} f_n \right) d\mu}_{0} < \underbrace{\liminf_{n\to\infty} \int_\Omega f_n \, d\mu}_{\mu(B^c)} < \underbrace{\limsup_{n\to\infty} \int_\Omega f_n \, d\mu}_{\mu(B)} < \underbrace{\int_\Omega \left(\limsup_{n\to\infty} f_n \right) d\mu}_{\mu(\Omega)}$$

The example captures the fact that *integrals of asymptotics give more extreme values than asymptotics of integrals*.

$\triangle$

Remark 7.5.2. (*Fatou's Lemma for series.*) In the special case where $\Omega = \{1, 2, 3, ...\}$, $\mathcal{F} = 2^\Omega$ (i.e. all subsets of Ω), and μ is the counting measure, Example 6.3.3 tells us that the integrals are series. That is, we can write $\sum_{k=1}^\infty x_k = \int f \, d\mu$, where $f : \Omega \to \mathcal{X}$ maps $\omega_1, \omega_2, \ldots$ to $x_1, x_2, \ldots$.

In this setting, the big picture conclusion of Fatou's lemma Eq. (7.5.3) specializes to: *~~integrals~~ series of asymptotics give more extreme values than asymptotics of series ~~integrals~~*:

$$\sum_{k=1}^\infty \liminf_{n\to\infty} x_{nk} \leq \liminf_{n\to\infty} \sum_{k=1}^\infty x_{nk} \leq \limsup_{n\to\infty} \sum_{k=1}^\infty x_{nk} \leq \sum_{k=1}^\infty \limsup_{n\to\infty} x_{nk} \qquad (7.5.4)$$

This observation will be useful when we discuss weak convergence.

$\triangle$

7.6 Dominated Convergence Theorem

The dominated convergence theorem justifies exchanging integrals and limits. We begin with a motivating example where the exchange is not warranted.

Example 7.6.1. (*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 $\int_{\mathbb{R}} f = \int_{\mathbb{R}} 0 = 0$.

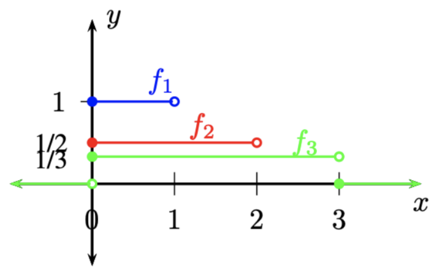


Figure 11: The first 3 functions in a sequence f_n where limit and integral (against Lebesgue measure) cannot be exchanged.

△

The dominated convergence theorem says that if all functions in a sequence are dominated by an integrable function ($|f_n| \leq g$), then the pathology of Example 7.6.1 cannot happen, and limits and integrals can be safely exchanged.

Theorem 7.6.1. ***Dominated Convergence Theorem.**.* *If $f_1, f_2, ..., f, g$ are Borel measurable, $|f_n| \leq g$ 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 $|f| \leq g$ a.e. (In detail, $|f_n| \leq g$ can be unpacked into $f_n \leq g$ and $f_n \geq -g$. Since limits preserve non-strict inequalities, we have ($f \leq g$ and $f \geq -g$) a.e. So repacking the absolute value operator, the conclusion follows.), hence f is integrable by Cor. 7.2.1 (b).

By hypothesis, both sides of the expanded Fatou's lemma apply, and so we have Eq. (7.5.3):

$$\underbrace{\int_\Omega \left(\liminf_{n\to\infty} f_n \right) d\mu}_{1} \leq \underbrace{\liminf_{n\to\infty} \int_\Omega f_n \, d\mu}_{2} \leq \underbrace{\limsup_{n\to\infty} \int_\Omega f_n \, d\mu}_{3} \leq \underbrace{\int_\Omega \left(\limsup_{n\to\infty} f_n \right) d\mu}_{4}$$

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. 7.3.1), 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

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega \lim_{n\to\infty} f_n \, d\mu .$$

□

7.7 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.[45] The following theorem describes continuity and the computation of derivative for such functions.

[45] Here, a "parameter" refers to a variable in the domain of the function that is not integrated over.

Theorem 7.7.1. ***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

$$\begin{aligned} \lim_{t \to t_0} F(t) &= \lim_{t \to t_0} \int_{\mathcal{X}} f(x, t)\ d\mu\ (x) && \text{definition of } F \\ &= \int_{\mathcal{X}} \lim_{t \to t_0} f(x, t)\ d\mu\ (x) && \text{Dominated Convergence Thm} \\ &= \int_{\mathcal{X}} f(x, t_0)\ d\mu\ (x) && \text{hypothesis} \\ &= F(t_0) && \text{definition of } F \end{aligned}$$

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

$$|h_n| \overset{\textit{(def)}}{=} \left|\frac{f(x, t_n) - f(x, t_0)}{t_n - t_0}\right| \overset{\textit{(MVT)}}{=} \left|\frac{\partial f}{\partial t}(x, s_n)\right| \overset{\textit{(hypothesis)}}{\le} g(x)$$

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 7.7.1. (*Extensions to real-valued parameters with unbounded support.*) Theorem 7.7.1 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*.[46]

[46] TODO: integrate this with my review of the Riemann integral in the appendix.

The criterion for Riemann integrability criterion in terms of Lebesgue integration. Now consider the measure space $(\Omega, \mathcal{F}, \mu) = ([a, b], ([a, b]), \text{Lebesgue measure})$, where $([a, b])$ 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 Eq. (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 $|f| \leq M$, and therefore $|\alpha_k|, |\beta_k| \leq M$. Moreover M is integrable, since $\int_{[a,b]} M \, d\mu = M(b-a) < \infty$.) Moreover, (by Proposition A.2.1), we have

$$\alpha_1 \geq \alpha_2 \geq ... \geq f \geq ... \geq \beta_2 \geq \beta_1$$

so α_k and β_k approach limit functions α and β. Thus, we can apply Lebesgue dominated convergence theorem to obtain

$$\lim_{k \to \infty} U(P_k) = \lim_{k \to \infty} \int_{[a,b]} \alpha_k \, d\mu \overset{\textit{(LDCT)}}{=} \int_{[a,b]} \alpha \, d\mu$$
$$\lim_{k \to \infty} L(P_k) = \lim_{k \to \infty} \int_{[a,b]} \beta_k \, d\mu \overset{\textit{(LDCT)}}{=} \int_{[a,b]} \beta \, d\mu$$

Thus, we can write the criterion for Riemann integrability Eq. (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 $\{P_k\}$.

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*

$$f \textit{ is continuous at } x \quad \textit{iff} \quad \alpha(x) = f(x) = \beta(x)$$

Proof.

- $\boxed{\implies}$. f is continuous at x means that $\forall \epsilon > 0, \exists \delta > 0$:

$$\begin{aligned} |y - x| < \delta &\implies |f(y) - f(x)| < \epsilon \\ &\implies f(y) \leq f(x) + \epsilon && \forall y \in B_x(\delta) \\ &\implies \sup_{y \in B_x(\delta)} f(y) \leq f(x) + \epsilon && 1. \end{aligned}$$

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.

- $\boxed{\Longleftarrow}$. By hypothesis,

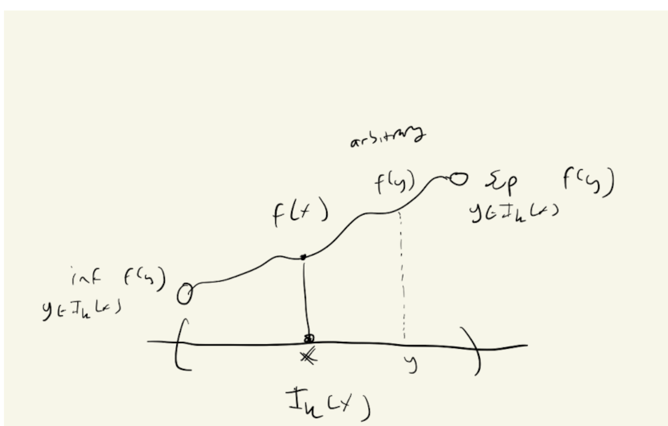
  $\forall \epsilon > 0, \exists K : \forall k \geq K,$

$$\begin{aligned} &|\sup_{y \in I_k(x)} f(y) - f(x)| \leq \epsilon \quad \text{and} \quad |\inf_{y \in I_k(x)} f(y) - f(x)| \leq \epsilon \\ \overset{1}{\Longrightarrow}\ & f(y) - f(x) < \epsilon \quad \text{and} \quad f(y) - f(x) > -\epsilon \quad \forall y \in I_k(x) \end{aligned}$$

  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.



□

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) • $\boxed{\implies}$. By Eq. (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 7.2.1),

$$\int_{[a,b]} (\alpha - \beta) \; d\mu \; = 0$$

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 7.3.2 (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.[47]

• $\boxed{\impliedby}$. 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 ($[a, b]$) are complete, f must be measurable by Problem 4.2.1.)[48] Since $\alpha = f = \beta$ a.e., by Theorem 7.3.1 (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 Eq. (8.0.3).[49]

b) If f is Riemann integrable, then f is continuous a.e. by part (a). So Eq. (8.0.4) holds. Then, by Eq. (8.0.3), $\int_{[a,b]} f \; d\mu \; = r$, the value of the Riemann integral.

□

Notation 8.0.1. (*Reusing Riemann notation for Lebesgue integrals.*) Let f be a measurable function on a measure space $(\Omega, \mathcal{F}, \lambda)$, where λ is Lebesgue measure. Although this integral exists for a broader class of functions than Riemann integrable functions (see Theorem 8.0.1), it is nevertheless common to designate it as

$$\int f(x) \; dx \; , \tag{8.0.5}$$

although of course one could also use $\int f \; d\lambda$ or any of the other notations given in (6.3.1). The interpretation of Eq. (8.0.5) is a Lebesgue integral whenever the integrand is not Riemann but still Lebesgue integrable. △

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].

---

[47] 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 $\mu(\lim_{k\to\infty} E_k) \stackrel{1}{=} \lim_{k\to\infty} \mu(E_k) = \lim_k \mu(\cup_{i=1}^{N_k} E_{ik}) \stackrel{2}{=} 0$, 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.

[48] 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.

[49] Presumably the "independent of the sequence of partitions" condition of Eq. (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.

9.1 Overview of signed measures

Definition 9.1.1. Let $(\Omega, \mathcal{F})$ be a measurable space.[50] 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 $\nu(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \nu(A_j)$.

△

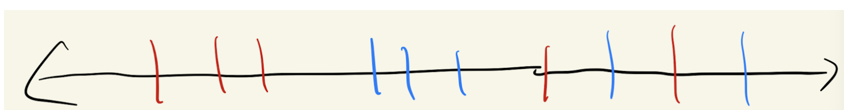
Remark 9.1.1. Every measure is a signed measure. △

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.[51]



△

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:

	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). △

---

[50] Recall Definition 6.2.1.

[51] 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.

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 $B \subset A$.

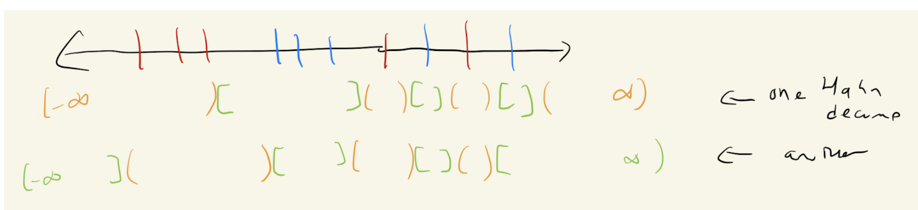
$\triangle$

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. $\triangle$

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. $\square$

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.]



$\triangle$

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. $\triangle$

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. $\triangle$

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 $\Omega = P \cupdot N$ 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 = $(E \cap P) \cupdot (E \cap N)$, 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 $\Omega = P \cupdot N$ 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[52]

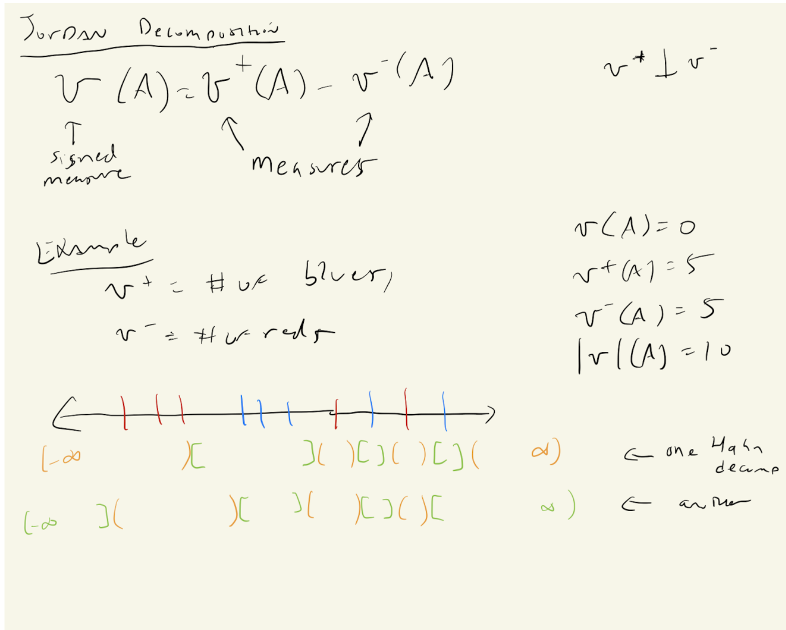
$$\begin{aligned}
|\nu|\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \nu^+\Big(\bigcup_{i=1}^{\infty} A_i\Big) + \nu^-\Big(\bigcup_{i=1}^{\infty} A_i\Big) \\
&= \sum_{i=1}^{\infty} \nu^+(A_i) + \sum_{i=1}^{\infty} \nu^-(A_i) && \nu^+, \nu^- \text{ are measures} \\
&= \sum_{i=1}^{\infty} \nu^+(A_i) + \nu^-(A_i) && \text{Limits and sums commute; see [Strichartz, 2000] Theorem 2.3.2 and apply it to partial sums} \\
&= \sum_{i=1}^{\infty} |\nu|(A_i) && \text{def. of } |\nu|
\end{aligned}$$

△

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

[52] TODO: Argue instead by appealing to a more general theorem.

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

- $\boxed{\implies}$. 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
  $$|\nu|(E) = \nu^+(E) + \nu^-(E) = 0.$$

- $\boxed{\Longleftarrow}$.

$$
\begin{aligned}
|\nu|(E) = 0 &\iff \nu^+(E) + \nu^-(E) = 0 && \text{def. } |\nu| \\
&\implies \nu^+(E) = 0, \nu^-(E) = 0 && \text{non-negativity of measure} \\
&\implies \nu^+(F) = 0, \nu^-(F) = 0 \quad \forall F \subset E && \text{monotonicity of measure} \\
&\implies \nu^+(F) + \nu^-(F) = 0 \quad \forall F \subset E \\
&\implies \nu(F) = 0 \quad \forall F \subset E
\end{aligned}
$$

□

9.4 Integrating against signed measure

We can define integration against a signed measure ν in terms of integrals against measure by using the Jordan decomposition $\nu = \nu^+ - \nu^-$.

Definition 9.4.1. Integration against signed measure Let ν be a signed measure and let f be a Borel measurable function. Then we define

$$\int f \, d\nu := \int f \, d\nu^+ - \int f \, d\nu^- \tag{9.4.1}$$

△

The integral is finite if both $\int f d\nu^+$ and $\int f d\nu^+$ are finite.

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.[53]

$$\text{i.e. } \nu(A) = \int_A h \, d\mu \,, \qquad \forall A \in \mathcal{F}$$

Then $\nu \ll \mu$

$$\text{i.e. } \mu(A) = 0 \implies \nu(A) = 0 \qquad \forall A \in \mathcal{F}$$

△

Proof. [54]

If $\mu(A) = 0$, $h1_A = 0$ a.e. So by Theorem 7.3.1 (a),

$$\int_A h \, d\mu \, = 0$$

□

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. △

Remark 10.1.3. (*A signed measure is absolutely continuous iff its total variation is absolutely continuous.*)[55] Let ν be a signed measure and μ a measure on a measurable space $(\Omega, \mathcal{F})$. Then

$$\nu \ll \mu \quad \text{iff} \quad |\nu| \ll \mu$$

△

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

$$\nu^+(A), \; \nu^-(A) = 0$$

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

[54]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 7.3.1 (a), so we might as well rely on the Theorem to do that dirty work.

[55]This is [Folland, 1999] Problem 3.2.8a

- $\boxed{\Longrightarrow}$.

$$\begin{aligned}
\mu(A) = 0 &\overset{\text{monotonicity}}{\Longrightarrow} A \text{ is } \mu\text{-null} \\
&\qquad \text{i.e. } \mu(B) = 0 \quad \forall B \subset A \\
&\overset{\text{hypothesis}}{\Longrightarrow} \nu(B) = 0 \quad \forall B \subset A \\
&\overset{\text{null sets have zero total variation (Problem 9.3.1)}}{\Longrightarrow} |\nu|(A) = 0
\end{aligned}$$

□

Remark 10.1.4. (*Absolute continuity is the antithesis of mutual singularity.*)

This statement (from [Folland, 1999]) captures the fact that

$$\begin{array}{ll}
\text{if} & \nu \text{ is a signed measure} \\
 & \mu \text{ is a measure} \\
 & \nu \ll \mu \\
 & \nu \perp \mu \\
\text{then} & \nu \equiv 0
\end{array}$$

△

Proof. First recall

$$\nu \ll \mu \overset{\text{Remark 10.1.3}}{\Longleftrightarrow} |\nu| \ll \mu$$

Now by mutual singularity

$$\begin{array}{llllll}
\exists A, B \in \mathcal{F}: & \Omega = A \cup B & & & & \\
 & \nu \text{ is null for } A & \overset{\text{Problem 9.3.1}}{\Longleftrightarrow} & |\nu|(A) = 0 & & \\
 & \mu \text{ is null for } B & \overset{\text{special case}}{\Longrightarrow} & \mu(B) = 0 & \overset{|\nu| \ll \mu}{\Longrightarrow} & |\nu|(B) = 0
\end{array}$$

So

$$\begin{aligned}
|\nu|(\Omega) &\overset{\text{countable additivity}}{=} |\nu|(A) + |\nu|(B) \overset{\text{see above}}{=} 0 \\
&\overset{\text{monotonicity, since } |\nu| \text{ is a measure}}{\Longrightarrow} |\nu| \equiv 0 \\
&\overset{*}{\Longrightarrow} \nu \equiv 0.
\end{aligned}$$

Implication (*) holds since

$$\begin{aligned}
|\nu| \equiv 0 &\overset{\text{def}}{\Longleftrightarrow} \nu^+ + \nu^- \equiv 0 \\
&\overset{\text{non-negativity}}{\Longrightarrow} \nu^+ \equiv 0, \nu^- \equiv 0 \\
&\Longrightarrow \nu^+ - \nu^- \equiv 0 \\
&\overset{\text{def.}}{\Longleftrightarrow} \nu \equiv 0.
\end{aligned}$$

□

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*

$$\begin{array}{lll}
\nu \ll \mu & \textit{iff} & \forall \epsilon > 0,\ \exists \delta > 0 : \forall A \in \mathcal{F}, \\
(\textit{i.e. } \mu(A) = 0 \implies \nu(A) = 0) & & \mu(A) < \delta \implies |\nu(A)| < \epsilon \\
A & & B
\end{array}$$

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

$$\begin{gathered} |\nu(A)|\le|\nu|(A) \\ |\nu^+(A)-\nu^-(A)|\le|\nu^+(A)+\nu^-(A)| \\ |x-y|\le|x+y| \end{gathered}$$

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

$$|\nu|(B)\overset{\text{cty from above}}{=}\lim_{k\to\infty}|\nu|(B_k)\overset{\text{monotonicity, } +}{\ge}\lim_{k\to\infty}\epsilon=\epsilon.$$

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 $E\in\mathcal{F}$,

$$\begin{gathered} \nu(E)=\nu^+(E)-\nu^-(E)\notin\{\infty,-\infty\} \\ \iff \\ |\nu|(E)=\nu^+(E)+\nu^-(E)\notin\{\infty,-\infty\}. \end{gathered}$$

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 7.1.1, 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

$$\begin{aligned}
\int_E h\, d\mu = \int_\Omega 1_E h\, d\mu &= \int_\Omega \Big(1_{E\cap A} + 1_{E\cap A^c}\Big) h\, d\mu \\
&= \int_{E\cap A} h\, d\mu + \int_{E\cap A^c} h\, d\mu \\
&= \int_{E\cap A} f\, d\mu + \int_{E\cap A^c} g\, d\mu \\
&\overset{\textit{(since } f, g \in \mathcal{S}\textit{)}}{\leq} \nu(E\cap A) + \nu(E \cap A^c) \\
&\overset{\textit{countable additivity}}{=} \nu(E).
\end{aligned}$$

Now let $a := \sup\{\int_\Omega f\, d\mu \ : f \in \mathcal{S}\}$, and note that $a \overset{\textit{sup is LUB}}{\leq} \nu(\Omega) \overset{\textit{assumption}}{<} \infty$. 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.5.2),[56] 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 7.3.2 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).
- $\lambda \perp \mu$? ✓ 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

	$d\lambda - d\lambda'$	$= (f - f')d\mu$	
But	$\lambda - \lambda'$	$\perp \mu$	Justified below
And	$\lambda - \lambda'$	$\ll \mu$	Since $d(\lambda - \lambda') = (f - f') \; d\mu$; indefinite integrals are absolutely continuous (Remark 10.1.1)
So	$\lambda - \lambda'$	$= 0$	$\ll$ and $\perp$ are antitheses; see Remark 10.1.4
So	λ	$= \lambda'$	
So	f	$= f' \quad [\mu] - a.e.$	Cor. 7.3.1

Let us show that if $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 - \lambda_2 \perp \mu$.

$\lambda_1 \perp \mu$	means	$\exists E_1, F_1 : E_1 \cup F_1 = \Omega$
		E_1 null for λ_1
		F_1 null for μ
$\lambda_2 \perp \mu$	means	$\exists E_2, F_2 : E_2 \cup F_2 = \Omega$
		E_2 null for λ_2
		F_2 null for μ

Now let $\lambda = \lambda_1 - \lambda_2$.

Let	$E = E_1 \cap E_2$
	Then E null for λ_1 and λ_2, so E null for $\lambda = \lambda_1 - \lambda_2$
Let	$F := E^c = (E_1 \cap E_2)^c = E_1^c \cup E_2^c = F_1 \cup F_2$
	So F null for μ.

□

[56] TODO: Show specifically that $f = \sup_n f_n$. The proposition gives us that $g = \sup_n g_n = \sup_n \max_{m \leq n} f_n$.

10.3 Focus on Radon-Nikodym

Corollary 10.3.1. ***Radon-Nikodym Theorem.*** *Let ν be a σ-finite signed measure and μ a σ-finite (positive) measure on $(\Omega, \mathcal{F})$. Let $\nu \ll \mu$. Then $d\nu = f d\mu$ for some (extended μ-integrable) function $f : \Omega \to \mathbb{R}$. In other words,*

$$\nu(A) = \int_A f \, d\mu \quad \text{for all} \quad A \in \mathcal{F} \tag{10.3.1}$$

Proof. In the Lebesgue-Radon-Nikodym theorem, set $\rho = \nu$ and $\lambda = 0$. □

Terminology 10.3.1. We refer to f in Eq. (10.3.1) as the **Radon-Nikodym derivative** or **density** of ν with respect to μ. We denote it by $d\nu/d\mu$. (Strictly speaking, $d\nu/d\mu$ should be construed as a class of functions equal to f μ-a.e.). When μ is the Lebesgue measure, then f is often simply called the density of ν. △

Below we note that in the special case where ν is a measure, the Radon-Nikodym derivative must be non-negative almost everywhere. (Hence, it can be taken to be non-negative everywhere.)

Corollary 10.3.2. *Let the hypotheses of the Radon-Nikodym Theorem (Cor. 10.3.1) hold. If $\nu \geq 0$, then $f \geq 0$ a.e. $[\mu]$.*

Proof. Let $A := \{x : f(x) < 0\}$. Now

$$0 \overset{\text{measure}}{\leq} \nu(A) = \int_A f \, d\mu \overset{\text{monotonicity}}{\leq} 0$$

So $\int_A f \, d\mu = 0$. Negating this and applying the scalar multiple property, we have that a non-negative function integrates to zero : $\int -f 1_A \, d\mu = 0$. Thus, by Theorem 7.3.2 b), the integrand $-f 1_A = 0$ a.e. $[\mu]$, and so negating again, $f 1_A = 0$ a.e. $[\mu]$. But by construction, $f < 0$ on A, so $\mu(A) = 0$. □

Proposition 10.3.1. *(Calculus with Radon-Nikodym Derivatives.)*[57]

a) (Additivity.) *Let $\nu = \nu_1 + \nu_2$ be a σ-finite signed measure on $(\Omega, \mathcal{F})$. If $\nu_1, \nu_2 \ll \mu$, then $\nu_1 + \nu_2 \ll \mu$, and*

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu} \quad \mu - a.e.$$

b) (Change of variables.) *Suppose that ν is a σ-finite signed measure and μ is a σ-finite measure on $(\Omega, \mathcal{F})$ such that $\nu \ll \mu$. If g is integrable, then $g(d\nu/d\mu)$ is integrable, and*

$$\int g \, d\nu = \int g \frac{d\nu}{d\mu} \, d\mu$$

c) (Chain rule.) *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 $\mu \ll \lambda$. We have $\nu \ll \lambda$, and*

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \quad \lambda - a.e. \tag{10.3.2}$$

[57] The terminology is borrowed from `https://pages.stat.wisc.edu/~doksum/STAT709/n709-5.pdf`.

Proof. a) For all $A \in \mathcal{F}$,

$$\begin{aligned}
\nu(A) &= \nu_1(A) + \nu_2(A) \\
&= \int_A f_1 \, d\mu + \int_A f_2 \, d\mu && \text{Radon-Nikodym derivatives} \\
&\overset{*}{=} \int_A (f_1 + f_2) \, d\mu && \text{linearity applies (see below)}
\end{aligned}$$

So $f_1 + f_2$ must be a Radon-Nikodym derivative of ν w.r.t to μ. Linearity applies in (*) because if not, then ν would fail to map some $A \in \mathcal{F}$ to a number, which would violate the definition of signed measure.

b) See [Folland, 1999, pp.91]. The proof is similar to that of Prop. 7.2.1, except that now we do not assume that $f := \frac{d\nu}{d\mu} \geq 0$.

Let us show that if the identity holds in the $\nu \geq 0$ case, it must hold for general ν. We have

$$\begin{aligned}
\int g \, d\nu &= \int g \, d\nu^+ - \int g \, d\nu^- && \text{Def. integrals of signed measure Eq. (9.4.1)} \\
&\overset{1}{=} \int g f^+ d\mu - \int g f^- d\mu && \text{by hypothesis, where } f^+, f^- \text{ are RN derivatives of } \nu^+, \nu^- \\
&\overset{2}{=} \int g(f^+ - f^-) d\mu && \text{linearity applies (see below)}
\end{aligned}$$

so $f^+ - f^-$ is a Radon-Nikodym derivative of ν w.r.t μ. The Radon-Nikodym derivatives in (1) are possible since $\nu \ll \mu$ implies $\nu^+, \nu^- \ll \mu$. (See Remark 10.1.3. In particular, since ν^+ and ν^- are both measures and therefore non-negative, $|\nu|(A) = \nu^+(A) + \nu^-(A) = 0$ implies that $\nu^+(A), \nu^-(A) = 0$.) Linearity applies in (2) since $\int g d\nu$ exists by hypothesis, so the expansion in the second line cannot take the form of $\infty - \infty$ or $-\infty + \infty$.[58]

c) For all $A \in \mathcal{F}$, we have

$$\begin{aligned}
\nu(A) &= \int_A d\nu && \text{Def. } \textstyle\int \text{ simple functions, since } \int_A d\nu := \int 1_A \, d\nu \\
&= \int_A \frac{d\nu}{d\mu} \, d\mu && \nu \ll \mu, \text{ Radon-Nikodym} \\
&= \int_A \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \, d\lambda && \mu \ll \lambda \text{ and part (a) (replacing } \nu, \mu \text{ by } \mu, \lambda\text{) where } g = \frac{d\nu}{d\mu} 1_A.
\end{aligned} \tag{10.3.3}$$

On the other hand, we have $\nu \ll \lambda$. ($\ll$ can be seen to be transitive immediately from the definition). Thus, by Radon-Nikodym,

$$\nu(A) = \int_A \frac{d\nu}{d\lambda} \, d\lambda \tag{10.3.4}$$

Since Eq. (10.3.3) and Eq. (10.3.4) both hold for all $A \in \mathcal{F}$, then by Cor. 7.3.1 ,

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \quad \lambda - \text{a.e.}$$

□

Although we can *generally* treat the differentials as fractions, care must be taken, as shown in the Remark below.

[58] TODO: This argument is essentially the additivity argument all over again. Can I just refer to that directly?

Remark 10.3.1. (*Problems with treating differentials as fractions.*) Let $a, b, c, d, e, f \in \mathbb{R} \setminus \{0\}$. Then by cross-multiplying, we of course have

$$\frac{a}{b} = \frac{c}{d}\frac{e}{f} \iff \frac{b}{a} = \frac{d}{c}\frac{f}{e}$$

However, note that for Radon-Nikodym derivatives, we have

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \not\Longleftrightarrow \frac{d\mu}{d\nu} = \frac{d\mu}{d\lambda}\frac{d\lambda}{d\nu}$$

The first trouble is that for the expression on the left side of $\not\Longleftrightarrow$ to exist, the chain rule Eq. (10.3.2) requires that $\nu \ll \mu, \mu \ll \lambda$. But absolute continuity is not a symmetric relation (recall Example 10.1.1), and if the dominance relations do not hold in the other direction, then the Radon-Nikodym derivatives on the right hand side of $\not\Longleftrightarrow$ may not exist. A second trouble is that the left side does not apply everywhere, but λ-a.e.; one might be tempted to flip the fractions on each side of the equality without taking care of the constraints on where the equality applies. $\triangle$

11 § 2.4 L^p spaces

11.1 Integrating complex-valued functions

We follow [Ash et al., 2000, pp. 83]. Let $(\Omega, \mathcal{F})$ be a measurable space, and let f be a complex-valued function on Ω, so that $f = \text{Re}f + i\,\text{Im}f$. We say that f is a *complex-valued Borel measurable function* on $(\Omega, \mathcal{F})$ if both $\text{Re}f$ and $\text{Im}f$ are real-valued Borel measurable functions. If μ is a measure on $\mathcal{F}$, we define

$$\int_\Omega f\; d\mu \;= \int_\Omega \text{Re}f\; d\mu \;+ i \int_\Omega \text{Im}f\; d\mu \tag{11.1.1}$$

provided $\text{Re}f$ and $\text{Im}f$ are *both* finite. In this case, we say that f is μ-integrable. Thus, in working with complex-valued functions, we do not consider any cases in which integrals exist but are finite.

11.2 What properties are preserved?

Many standard properties of the integral carry over to the complex case. (For a complete list, see [Ash et al., 2000, pp. 84].) In almost all cases, the result is an immediate consequence of the fact that integrating a complex function is equivalent to integrating the real and complex parts separately.

The only theorems that require additional comment are the triangle inequality (Prop 6.3.6 c), dominated convergence theorem, the Radon-Nikodym theorem, and transfer of integrability (Cor 7.2.1 (a)). For the first two, see [Ash et al., 2000, pp. 84]. For the third, see [Ash et al., 2000, Problem 10, pp. 95]. We show the fourth here.

Proposition 11.2.1. *Let f be a complex-valued Borel measurable function on $(\Omega, \mathcal{F})$. Then f is integrable iff $|f|$ is integrable.*

Proof. First, a note. By definition, $|f|$ being integrable means that $|\int |f|\; d\mu\;| < \infty$. But the integral of a non-negative function can only be non-negative (by monotonicity of the integral of real valued functions), so this is equivalent to $\int |f|\; d\mu\; < \infty$.

We will apply the strategy of bounding the complex modulus by real moduli Eq. (C.2.1), which in the case of functions can be written as

$$|\text{Re}f|, |\text{Im}f| \;\overset{A}{\leq} |f| \;\overset{B}{\leq} |\text{Re}f| + |\text{Im}f| \tag{11.2.1}$$

- $\boxed{\Longleftarrow}$. To show that f is integrable, we need to show that $|\int_\Omega \text{Re}f \; d\mu \; |, |\int_\Omega \text{Im}f \; d\mu \; | < \infty$.

$$\begin{vmatrix} \int_\Omega \text{Re}f \; d\mu \\ \int_\Omega \text{Im}f \; d\mu \end{vmatrix} \quad \begin{matrix} \overset{1}{\leq} \int_\Omega |\text{Re}f| \; d\mu & \overset{\text{by A}}{\leq} \int_\Omega |f| \; d\mu & \overset{\text{hypoth}}{<} \infty \\ \overset{1}{\leq} \int_\Omega |\text{Im}f| \; d\mu & \overset{\text{by A}}{\leq} \int_\Omega |f| \; d\mu & \overset{\text{hypoth}}{<} \infty \end{matrix}$$

where (1) is the triangle inequality for integrals of real-valued functions (Prop 6.3.6 c).

- $\boxed{\Longrightarrow}$. We have

$$\int_\Omega |f| \; d\mu \overset{\text{by B}}{\leq} \int_\Omega |\text{Im}f| + |\text{Re}f| \; d\mu \overset{\text{linearity}}{=} \int_\Omega |\text{Im}f| \; d\mu \; + \int_\Omega |\text{Re}f| \; d\mu \overset{\text{hypothesis}}{<} \infty$$

where in applying both linearity and the hypothesis, we utilize from (Cor 7.2.1 (a)) that the proposition is true for real-valued functions ($| \int f \; d\mu \; | < \infty$ iff $\int |f| \; d\mu \; < \infty$).

□

Note that monotonicity does not transfer over, because the complex numbers cannot be ordered [Stewart and Tall, 2018, Sec 1.8].

11.3 L^p spaces $(0 < p < \infty)$

We begin with a definition of L^p spaces when p is a finite positive real number.

Definition 11.3.1. (L^p **space,** $0 < p < \infty$.) If $0 < p < \infty$, we define the space $L^p = L^p(\Omega, \mathcal{F}, \mu)$ as the collection of complex-valued Borel measurable functions f such that $\int_\Omega |f|^p \; d\mu \; < \infty$. △

We defer defining L^∞ spaces to later.

11.4 L^p spaces $(1 \leq p < \infty)$ as normed vector spaces

When $1 \leq p \leq \infty$, L^p is a *vector space* (see Sec. D.3). To show this, we will use Minkowski's inequality.

Theorem 11.4.1. ***Minkowski's inequality*** *If $f, g \in L^p (1 \leq p < \infty)$, then $f + g \in L^p$ and $||f + g||_p \leq ||f||_p + ||g||_p$.*

Proof. See [Ash et al., 2000, pp.86]. □

Now let us check that L^p is a *vector space* (see Sec. D.3) for $1 \leq p < \infty$. (For the $p = \infty$ case, see [Ash et al., 2000, pp.93].)

- *Closure under scalar multiplication?* ✓ . This is given by the scalar multiple property of integrals (Prop. 6.3.6 a) (which extends immediately to the complex case by definition of the integral of a complex-valued function).

- *Closure under addition?* ✓ . Given by Minkowski's inequality.

Moreover, by defining

$$||f||_p := \left(\int_\Omega |f|^p \; d\mu \right)^{1/p}, \quad f \in L^p. \tag{11.4.1}$$

we see that the vector space $L^p (1 \leq p < \infty)$ has a *semi-norm* (Def. D.4.1). We check:

- *Non-negativity?* ✓ . Immediate from the definition.
- *Scalar multiple property?* ✓ . Immediate from the scalar multiple property of integrals (Prop. 6.3.6 a)).
- *Triangle inequality?* ✓ . Given by Minkowski's inequality.

The semi-norm immediately induces a *psuedo-distance* (Def. D.5.1) by $d(f, g) := ||f - g||_p$.

If we define equivalence classes $f \sim g$ iff $f = g$ a.e. $[\mu]$, then $|| \cdot ||_p$ becomes a *norm* (Def. D.4.1) and d becomes a *distance* (Def. D.5.1).

11.5 Approximation Theorems on L^p spaces $(0 < p < \infty)$

Functions in L^p for $0 < p < \infty$ can be approximated nicely. We point out two basic approximation theorems.

- Finite valued simple functions are dense in L^p for $0 < p < \infty$. (see Theorem 12.5.2). Here $L^p = L^p((\Omega, \mathcal{F}, \mu))$, where $(\Omega, \mathcal{F}, \mu)$ is any measure space.
- Continuous functions are dense in $L^p(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$ for $0 < p < \infty$, where μ is a Lebesgue-Stieltjes measure. (See [Ash et al., 2000, Thm 2.4.14].)

Exercise 11.5.1. How can these approximation theorems hold any meaning if $||f||_p$ is not a norm for $0 < p < 1$? △

11.6 L^∞ spaces

We may also define L^∞ spaces, but we must proceed differently. We first define the essential supremum.

Definition 11.6.1. If f is a real-valued Borel measurable function on $(\Omega, \mathcal{F}, \mu)$, we define its **essential supremum** as:

$$\operatorname{ess\,sup} f := \inf \left\{ c \in \overline{R} : \mu\{\omega : f(\omega) < c\} = 0 \right\}$$

that is, the smallest number c such that $g \leq c$ a.e. $[\mu]$. △

We may now define L^∞ spaces.

Definition 11.6.2. (L^∞ **space**.) If f is a complex-valued Borel measurable function on $(\Omega, \mathcal{F}, \mu)$, define

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

Then the space $L^\infty = L^\infty(\Omega, \mathcal{F}, \mu)$ is the collection of complex-valued Borel measurable functions f such that $||f||_\infty < \infty$. △

Thus $f \in L^\infty$ if f is essentially bounded; that is, bounded outside a set of measure 0.

L^∞ is a normed vector space. As with $0 < p < \infty$, L^∞ is a vector space, and $|| \cdot ||_\infty$ is a semi-norm, which becomes a norm if we pass to equivalence classes as before. For details, see [Ash et al., 2000, pp.93].

Convergence in L^∞. If $f, f_1, f_2, \ldots \in L^\infty$ and $||f_n - f||_\infty \to 0$, we write $f_n \overset{L^\infty}{\to} f$.

Remark 11.6.1. L^∞ convergence is equivalent to uniform convergence a.e. [Ash et al., 2000, pp.93]. △

12 § 2.5 Convergence of sequences of measurable functions

12.1 Definitions

Definition 12.1.1. (Modes of convergence.) A sequence of complex-valued Borel functions $f_1, f_2, \dots$ on $(\Omega, \mathcal{F}, \mu)$ is said to

1. **Converge almost everywhere**, written $f_n \to f$ a.e., if

$$\begin{aligned} &\exists A \in \mathcal{F} : \\ &\mu(A) = 0 \quad \text{and} \quad f_n \to f \quad \text{on } A^c \end{aligned}$$

   In fully quantified form, we have

$$\begin{aligned} &\exists\, A \in \mathcal{F} : \mu(A) = 0 \quad \text{and} \quad \forall\, \omega \in A^c, \epsilon > 0, \exists\, N : \forall\, n \geq N \\ &|f_n(\omega) - f(\omega)| < \epsilon \end{aligned}$$

2. **Converge almost uniformly**, written $f_n \xrightarrow{AU} f$ if

$$\begin{aligned} &\forall \epsilon > 0,\ \exists A_\epsilon \in \mathcal{F} : \\ &\mu(A_\epsilon) < \epsilon \quad \text{and} \quad f_n \to f \quad \text{uniformly on } A_\epsilon^c \end{aligned}$$

   In fully quantified form[59], we have

$$\begin{aligned} &\forall\, \epsilon < 0, \exists\, A_\epsilon \in \mathcal{F} : \mu(A_\epsilon) < \epsilon \quad \text{and} \quad \forall\, \widetilde{\epsilon} > 0, \exists\, N : \forall\, n \geq N, \omega \in A_\epsilon^c \\ &|f_n(\omega) - f(\omega)| < \widetilde{\epsilon} \end{aligned} \tag{12.1.1}$$

3. **Converge in measure**, written $f_n \xrightarrow{\mu} f$, if

$$\forall\, \epsilon > 0, \qquad \mu\{\omega : |f_n(\omega) - f(\omega)| \geq \epsilon\} \to 0 \quad \text{as} \quad n \to \infty$$

   We can write this in fully quantified form as

$$\begin{aligned} &\forall\, \widetilde{\epsilon} > 0, \epsilon > 0, \exists\, N : \forall\, n \geq N : \\ &\mu\{\omega : |f_n(\omega) - f(\omega)| \geq \widetilde{\epsilon}\} < \epsilon \end{aligned}$$

4. **Converge in L^p** (here assumed $0 < p < \infty$), written $f_n \xrightarrow{L^p} f$, if

$$\left(\int |f_n - f|^p \, d\mu \right)^{\cancel{1/p}} \to 0 \quad \text{as} \quad n \to \infty$$

   In quantified form:

$$\begin{aligned} &\forall\, \epsilon > 0, \exists\, N : \forall\, n \geq N : \\ &\left(\int |f_n - f|^p \, d\mu \right)^{\cancel{1/p}} < \epsilon \end{aligned}$$

   Note that we have canceled out the $1/p$ exponent since convergence holds with it iff it holds without it.

[59] Recall that $f_n \to f$ uniformly means

$$\begin{aligned} &\forall \epsilon < 0, \exists N : \forall n \geq N, \omega \in \Omega \\ &|f_n(\omega) - f(\omega)| < \epsilon \end{aligned}$$

△

Remark 12.1.1. (*On the asymmetry in the definitions of convergence a.e. and convergence almost uniformly.*) It is regrettable [Bartle, 2014, pp.72] that the modifier "almost" has different meanings in *almost everywhere* convergence and *almost uniform* convergence. For AU convergence, what is required is uniform convergence outside a set of *arbitrarily large measure*, rather than uniform convergence outside a set of *zero measure*. The latter is a stronger condition (see [Bartle, 2014, Exercise 7J]), which presumably Egoroff's Theorem (which motivated the construction of the AU definition) cannot attain. △

12.2 Relations between modes of convergence: Overview

In the Figures below, we summarize the relationship between modes of convergence. The Figures borrow from the charts of [Bartle, 2014, pp.75]. In each chart, a solid line means that convergence in the mode at the tail of the arrow implies convergence in the mode at the head.

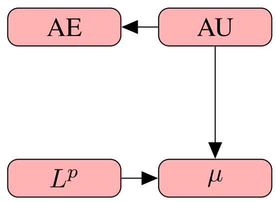


Figure 12: *Relationship between modes of convergence in a general measure space.*

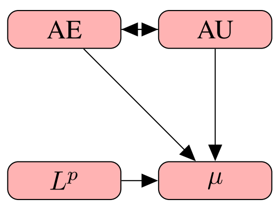


Figure 13: *Relationship between modes of convergence in a finite measure space.*

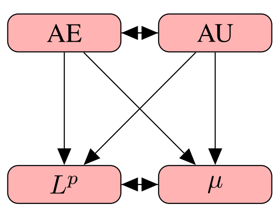


Figure 14: *Relationship between modes of convergence under a dominating function.* Here we assume the sequence (f_n) is dominated by a function g in L^p. The measure space is general (not necessarily finite.)

We note:

- In a finite measure space, or when a dominating function exists, a.e. convergence becomes stronger than it is in general measure spaces. The additional two edges in moving from measure general spaces (Fig. 12) to finite measure spaces (Fig. 13) are obtained from Egoroff's Theorem.

- When a dominating function exists, the same relations exist as in a finite measure space, along with three new ones. In particular, the three other modes of convergence becomes stronger relative to L^p convergence. We note that in this context, *a.e. convergence implies all three other modes of convergence*!

12.3 Convergence on general measure spaces: Implications

The implications in the section are powerful because they *always* hold. In some special case, if one can prove that the antecedent holds, then the consequent automatically follows as well.

12.3.1 Convergence in L^p implies convergence in measure

To relate $\stackrel{L^p}{\to}$ and $\stackrel{\mu}{\to}$, we can will utilize Chebyshev's Inequality.

Theorem 12.3.1. ***(Chebyshev's Inequality.)*** *Let $f : \Omega \to [0, \infty]$ be a Borel measurable function on $(\Omega, \mathcal{F}, \mu)$.*

If $0 < p < \infty$ and $0 < \epsilon < \infty$, then

$$\mu\{\omega : f(\omega) \geq \epsilon\} \leq \frac{1}{\epsilon^p} \int f^p \, d\mu$$

Proof. Using monotonicity, we can obtain a lower bound the integral by the integral of a simple function.

$$\int_\Omega f^p \, d\mu \stackrel{\text{monotonicity}}{\geq} \int_{\{\omega : f(\omega) > \epsilon\}} f^p \, d\mu \stackrel{\text{monotonicity}}{\geq} \int_{\{\omega : f(\omega) > \epsilon\}} \epsilon^p \, d\mu \stackrel{\text{int. simple func.}}{=} \epsilon^p \mu\{\omega : f(\omega) > \epsilon\}$$

□

Remark 12.3.1. The proof of Chebyshev's theorem highlights yet again the niceities of Lebesgue integration - we can integrate over any set we want (so long as the set is measurable), and those sets can be defined in terms of conditions in the range. △

Now we can show that convergence in L^p is stronger than convergence in measure

Theorem 12.3.2. *If $f_1, f_2, \ldots \in L^p$ $(0 < p < \infty)$, then $f_n \stackrel{L^p}{\to} f \implies f_n \stackrel{\mu}{\to} f$.*

Proof. We have

$$\begin{aligned}
&\forall \epsilon < 0, \exists N : \forall n \geq N, \\
&\mu\{\omega : |f_n(\omega) - f(\omega)| \geq \widetilde{\epsilon}\} \stackrel{(1)}{\leq} \frac{1}{\widetilde{\epsilon}^p} \int |f_n - f|^p \, d\mu \stackrel{(2)}{<} \epsilon
\end{aligned}$$

where (1) is obtained by applying Chebyshev's Inequality to the non-negative function $|f_n - f|$, and (2) is obtained by the hypothesis of L^p convergence. (More precisely, L^p convergence can make $\int |f_n - f|^p \, d\mu < \epsilon_1$ for any desired $\epsilon_1 > 0$ for sufficiently large n. So we decompose we write $\epsilon = \frac{1}{\widetilde{\epsilon}^p} \epsilon_1$. Since $\widetilde{\epsilon}^p$ is fixed, we can choose ϵ_1 to get any desired ϵ.) □

12.3.2 Almost uniform convergence implies convergence almost everywhere and convergence in measure

Now we show that almost uniform convergence is stronger than convergence almost everywhere and convergence in measure.

Theorem 12.3.3. *We have*

a) $f_n \overset{AU}{\to} f \implies f_n \overset{a.e.}{\to} f$

b) $f_n \overset{AU}{\to} f \implies f_n \overset{\mu}{\to} f$

Proof. a) TBD.

b) We take the definition of $\overset{AU}{\to}$ in quantified form Eq. (12.1.1)

$$\forall \widetilde{\epsilon} > 0, \exists A \in \mathcal{F} : \mu(A) < \widetilde{\epsilon} \quad \text{and} \quad \forall \epsilon > 0, \exists N : \forall n \geq N, \omega \in A^c$$
$$|f_n(\omega) - f(\omega)| < \epsilon$$

and remove the explicit reference to the set A. In particular, for fixed $\widetilde{\epsilon}, \epsilon > 0$ and sufficiently large n, we have

$$\{\omega : |f_n(\omega) - f(\omega)| \geq \epsilon\} \subset A$$
$$\implies \mu\{\omega : |f_n(\omega) - f(\omega)| \geq \epsilon\} \overset{\textit{monotonicity of measure}}{\leq} \mu(A) \overset{\textit{hypoth}}{<} \widetilde{\epsilon}$$

So the original quantified statement becomes

$$\forall \widetilde{\epsilon} > 0, \epsilon > 0, \exists N : \forall n \geq N$$
$$\mu\{\omega : |f_n(\omega) - f(\omega)| \geq \epsilon\} < \widetilde{\epsilon}$$

which is the definition of convergence in measure.

□

12.4 Convergence on finite measure spaces: Implications

As noted in Sec. 12.2, finite measure spaces have two implications that doesn't exist for general measure spaces: $f_n \overset{a.e.}{\to} f \implies f_n \overset{AU}{\to} f, f_n \overset{\mu}{\to} f$. We obtain the former implication via Egoroff's Theorem, and then the latter follows immediately from Thm. 12.3.3.

Theorem 12.4.1. ***(Egoroff's Theorem.)*** *If μ is finite and $f_1, f_2, \ldots$ and f are measurable complex-valued functions, then $f_n \overset{a.e.}{\to} f \implies f_n \overset{AU}{\to} f$.*

Proof. Without loss of generality, we assume $f_n \to f$ everywhere. (If the implication holds for everywhere convergence, it must hold for a.e. convergence. Suppose that N is a set such that $\mu(N) = 0$ and $f_n \to f$ outside N, and suppose that A_ϵ are the sets we have obtained for AU convergence from the weaker condition. Then we simply form the sets $N \cup A_\epsilon$ in the definition of AU convergence, and we will automatically have $\mu(N \cup A_\epsilon) \overset{\textit{subadditivity}}{<} \mu(N) + \mu(A_\epsilon) < \epsilon$) and uniform convergence will hold outside of $N \cup A_\epsilon$.)

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

$$E_n(k) := \underbrace{\bigcup_{m=n}^{\infty} \left\{ x : |f_m(x) - f(x)| > \frac{1}{k} \right\}}_{\text{Set of points whose outputs are at least } \frac{1}{k} \text{ away from the target somewhere in the tail starting at } n}$$

Let us fix k, consider this a sequence of sets in n, and check the conditions for continuity from above:

- $E_n(k)$ decreases as n increases? ✓ .
  - Note: $\bigcap_{n=1}^{\infty} E_n(k) = \emptyset$ by hypothesis that $f_n \to f$ everywhere.

- $\mu(E_1(k)) < \infty$? ✓ . By hypothesis, μ is finite.

Thus, by continuity from above,

$$\mu\big(E_n(k)\big) \to 0 \quad \text{as} \quad n \to \infty$$

As a result, given $\epsilon > 0$ and $k \in \mathbb{N}$, we can choose n_k so large that

$$\mu\big(E_{n_k}(k)\big) < \epsilon 2^{-k}$$

Now let $E := \cup_{k=1}^{\infty} E_{n_k}(k)$. By subadditivity, we have

$$\mu(E) \overset{\textit{subadditivity}}{\leq} \sum_{k=1}^{\infty} \mu\big(E_n(k)\big) \overset{\textit{see above}}{<} \sum_{k=1}^{\infty} \epsilon 2^{-k} \overset{\textit{geometric series}}{=} \epsilon$$

Now we show that $f_n \to f$ uniformly on E^c. Applying DeMorgan's law

$$E^c \overset{\textit{def. E, DeMorgan}}{=} \bigcap_{k=1}^{\infty} E_{n_k}(k)^c \overset{\textit{def. } E_n\textit{, DeMorgan}}{=} \underbrace{\bigcap_{k=1}^{\infty} \underbrace{\bigcap_{m=n_k}^{\infty} \Big\{x : |f_m(x) - f(x)| < \frac{1}{k}\Big\}}_{\text{points within a } \frac{1}{k} \text{ distance of target in tail starting at } n_k}}_{\text{...for all } k}$$

Thus, E^c describes precisely the points of uniform convergence. □

Remark 12.4.1. The proof above utilizes a set theoretic definition of uniform convergence. See Remark D.7.1. △

Remark 12.4.2. Combining Egoroff's Theorem with Thm. 12.3.3, we obtain that if μ is finite, then

$$f_n \overset{a.e.}{\to} 0 \overset{\textit{Egoroff's Theorem (Thm. 12.4.1)}}{\Longrightarrow} f_n \overset{AU}{\to} 0 \overset{\textit{Thm. 12.3.3}}{\Longrightarrow} f_n \overset{\mu}{\to} 0$$

This chain of implications explains the two new links that appear in Fig. 13 compared to Fig. 12. △

12.5 Convergence under a dominating function: Implications

12.5.1 Convergence a.e. and convergence in L^p

First we show that in general

$$f_n \overset{a.e.}{\to} f \not\Rightarrow f_n \overset{L^p}{\to} f$$

Let $\Omega = [0, \infty)$ and μ be Lebesgue measure. Let $f_n = n\, 1_{(0,\frac{1}{n}]}$. Then $f_n \to 0$ a.e. But

$$||f_n - 0||_p^p = \int_\Omega \left(n\, 1_{(0,\frac{1}{n}]}\right)^p d\mu \; = n^p \mu((0, 1/n]) = n^p(\frac{1}{n}) = n^{p-1} \not\to 0 \quad \text{as} \quad n \to \infty$$

However, if the sequence has a *dominating function* (in L^p, i.e. its pth power is integrable), then the implication holds.

Theorem 12.5.1. *[Ash et al., 2000, Cor. 1.6.10]. Let $f_1, f_2, \ldots, f, g$ be Borel measurable, and $|f_n| \leq g$ for all n, where $|g|^p$ is μ-integrable* ($p > 0$, fixed). *Then $f \in L^p$ and*

$$f_n \overset{a.e.}{\to} f \implies f_n \overset{L^p}{\to} f$$

Proof. First note that
$$|f_n| \le g \overset{1}{\Longrightarrow} \lim_{n\to\infty} |f_n| \le g \overset{2}{\Longrightarrow} |f| \le g$$
where (1) holds because limits preserve non-strict inequalities and (2) holds because limits and absolute values commute.

From this, by raising both sides of the implication to the p-th power, we immediately obtain
$$|f_n|^p \le |g|^p, \quad |f|^p \le |g|^p$$

At this point, we know that $|f|^p$ is integrable (i.e. $f \in L^p$).

Now we will apply the Dominated Convergence Theorem to show L^p convergence. First, we obtain a dominating function
$$|f_n - f|^p \overset{\text{tr. inequal.}}{\le} (|f_n| + |f|)^p \overset{\text{from above}}{\le} (2|g|)^p \overset{\text{hypoth.}}{<} \infty$$

Then we have
$$\lim_{n\to\infty} \int |f_n - f|^p \, d\mu \overset{\text{LDCT}}{=} \int \lim_{n\to\infty} |f_n - f|^p \, d\mu \overset{1}{=} 0$$
where (1) holds by the hypothesis of a.e. convergence (so $|f_n - f|^p = 0$ a.e.) followed by the a.e. theorems (see Theorem 7.3.1 (a)). □

We can use Theorem 12.5.1 to prove that finite-valued simple functions are dense in L^p.

Theorem 12.5.2. (Finite valued simple functions are dense in L^p.) *Let $f \in L^p$, $(0 < p < \infty)$. If $\epsilon > 0$, there is a simple function $s \in L^p$ such that $||f - s||_p < \epsilon$; s can be chosen to be finite-valued and to satisfy $|s| \le |f|$.*

Proof. First note that Theorem 6.3.7 states that any measurable function is the limit of a sequence of simple functions which it dominates. In particular, we have
$$s_n \to f \quad \text{everywhere}, \qquad |s_n| < f$$
where s_n is a sequence of finite-valued simple functions.

Since we have identified a.e. convergence and a dominating function with the correct integrability condition (note that $|f|^p$ is μ-integrable, since $f \in L^p$ by hypothesis.), we can apply Theorem 12.5.1 to obtain $s_n \overset{L^p}{\to} f$.

The conclusion immediately follows. (For any $\epsilon > 0$, we can choose an n so that $||f - s_n||_p < \epsilon$.)

□

12.6 Counterexamples

The counterexamples below will prove the *non*-existence of all 9 missing edges from the 12 possible edges in Figure 12.

12.6.1 Counterexample 1: No other modes imply L^p convergence in a finite (and therefore in a general) measure space

This counterexample justifies the non-existence of the three crossed-out links in Fig. 15.

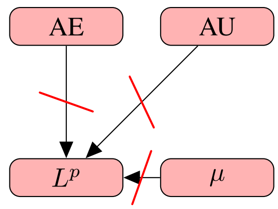


Figure 15: *Three non-existing relationships between modes of convergence in a finite (and therefore general) measure space.* The non-existence of these links is justified by the counterexample in this section.

Let $\Omega = [0,1]$, μ be Lebesgue measure, and $\mathcal{F}$ be the Borel sets. Define

$$f_n(x) = \begin{cases} e^n & \text{if} \quad 0 \leq x \leq \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

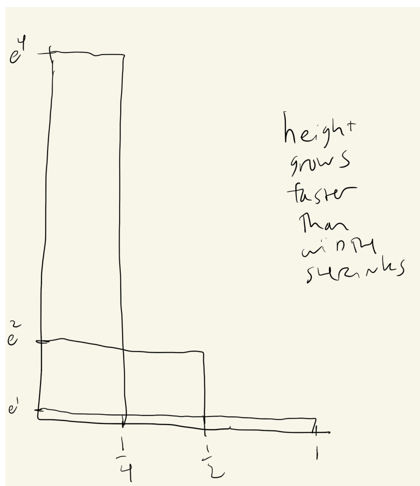


Figure 16

Then clearly $f_n \overset{a.e.}{\to} 0$. Moreover we have

$$f_n \overset{a.e.}{\to} 0 \overset{\text{Egoroff's Theorem (Thm. 12.4.1)}}{\Longrightarrow} f_n \overset{AU}{\to} 0 \overset{\text{Thm. 12.3.3}}{\Longrightarrow} f_n \overset{\mu}{\to} 0$$

(Egoroff's Theorem applies since the measure is finite.)

However, for $0 < p < \infty$,

$$\begin{aligned} ||f_n||_p^p &= \int_0^1 \left| e^n 1_{[0,\frac{1}{n}]} \right|^p d\mu \\ &= \int_0^1 e^{np} 1_{[0,\frac{1}{n}]} \, d\mu \\ &= e^{np} \mu\big[0, \frac{1}{n}\big] && \text{int. simple func.} \\ &= e^{np} \frac{1}{n} \to \infty \end{aligned}$$

And so $f_n \overset{L^p}{\not\to} 0$ for $p \in (0, \infty)$.

Remark 12.6.1. (*Convergence in L^∞.*) Likewise, for $p = \infty$

$$||f_n||_\infty = e^n \to \infty$$

so $f_n \overset{L^p}{\nrightarrow} 0$ for $p = \infty$. △

12.6.2 Counterexample 2: Another counterexample showing that no other modes imply L^p convergence in a general measure space

We provide a second counterexample showing that no other modes imply L^p convergence in a general measure space (see Fig. 15).

Let $\Omega = \mathbb{R}$, μ be Lebesgue measure, and $\mathcal{F}$ be the Borel sets. Define

$$f_n(x) = \begin{cases} \frac{1}{n} & \text{if} \quad 0 \le x \le e^n \\ 0 & \text{otherwise} \end{cases}$$

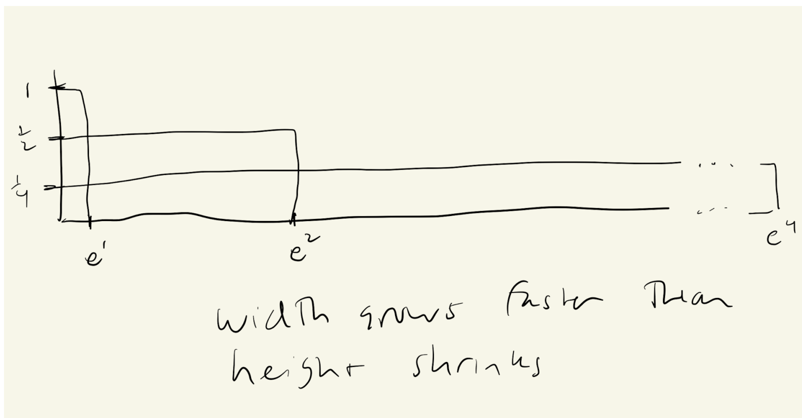


Figure 17

Then $f_n \to 0$ uniformly on $\mathbb{R}$ (Recall Def. D.7.1.) Thus, by Thm. 12.3.3, $f_n \overset{a.e.}{\to} 0$, $f_n \overset{\mu}{\to} 0$.

However, for $0 < p < \infty$,

$$\begin{aligned} ||f_n||_p^p &= \int \left| \frac{1}{n} 1_{[0,e^n]} \right|^p \, d\mu \\ &= (\frac{1}{n})^p \int 1_{[0,e^n]} \, d\mu \\ &= (\frac{1}{n})^p \mu[0, e^n] \\ &= (\frac{1}{n})^p \, e^n \to \infty \end{aligned}$$

Remark 12.6.2. (*Convergence in L^∞.*) Although (f_n) does not converge in L^p for $(0 < p < \infty)$, it does converge in L^∞. This is because uniform convergence a.e. implies convergence in L^∞ ; see Remark 11.6.1. △

12.6.3 Counterexample 3: no other mode of convergence implies AU convergence in a general measure space

Here we provide a counterexample showing that no other mode of convergence implies AU convergence in a general measure space.

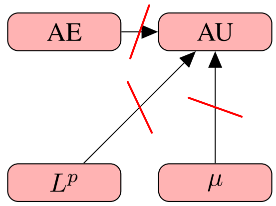


Figure 18: *Three more non-existing relationships between modes of convergence in a general measure space.* The non-existence of these links is justified by the counterexample in this section.

Let $\Omega = [0, \infty)$, μ be Lebesgue measure, and $\mathcal{F}$ be the Borel sets. Define

$$f_n(x) = \begin{cases} 1 & \text{if} \quad n \leq x \leq n + \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

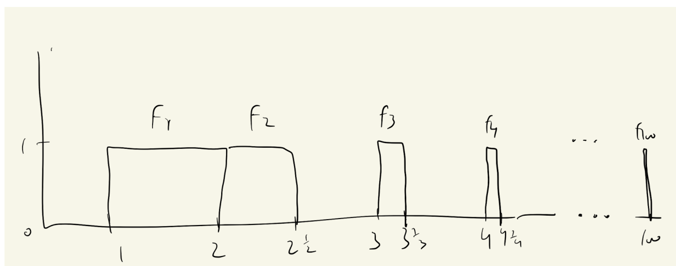


Figure 19

We first show that convergence holds AE, in μ, and in L^p.

- $f_n \xrightarrow{a.e.} 0$? ✓ Clearly $f_n \to 0$ pointwise.
- $f_n \xrightarrow{L^p} 0$? ✓

$$||f_n||_p^p = \int \left| 1_{[n,n+\frac{1}{n}]} \right|^{\cancel{p}} d\mu \; = \mu[n, n+\frac{1}{n}] = \frac{1}{n} \to 0 \quad \text{as} \quad n \to \infty$$

- $f_n \xrightarrow{\mu} 0$? ✓ This is implied by convergence in L^p (Thm. 12.3.2). Alternatively, we could check directly. For all $\epsilon > 0$,

$$\mu\{x : |f_n(x)| \geq \epsilon\} = \frac{1}{n} \to 0 \quad \text{as} \quad n \to \infty$$

But (f_n) does not converge AU (see [Ash et al., 2000, pp.99].

12.6.4 Counterexample 4: Convergence in L^p or in μ does not imply convergence AE or AU

Here we provide a counterexample showing that convergence in L^p (and therefore measure) does not imply convergence AE (and therefore AU).

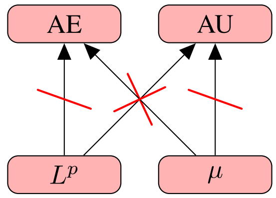


Figure 20: *Four more non-existing relationships between modes of convergence in a finite (and therefore general) measure space.* The non-existence of these links is justified by the counterexample in this section. (Note that the absence of link from L^p to AU was already justified in Fig. 18.)

Let $\Omega = (0, 1]$, μ be Lebesgue measure, and $\mathcal{F}$ be the Borel sets. Define

$$f_{nm}(x) = \begin{cases} 1 & \text{if} \quad \frac{m-1}{n} < x \leq \frac{m}{n}, \qquad m = 1, \ldots, n, \quad n = 1, 2, \ldots \\ 0 & \text{elsewhere} \end{cases}$$

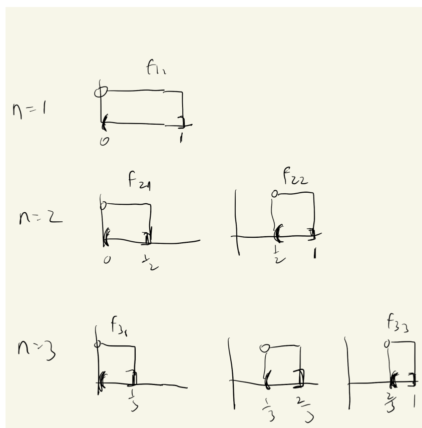


Figure 21

We construct a sequence of functions f_{nm} by ordering the functions first by n and then by m.

Now $f_{nm} \not\to 0$ a.e. (since for any $x \in (0, 1]$, $f_{nm}(x)$ has infinitely many 0's and 1's).. Thus, $f_{nm} \overset{AU}{\not\to} 0$, by Thm. 12.3.3.

However, $f_{nm} \overset{L^p}{\to} 0$ for $0 < p < \infty$, since

$$\begin{aligned} ||f_{nm}||_p^p &= \int \left| f_{nm} \right|^p \, d\mu \\ &= \int 1_{[\frac{m-1}{n} < x \leq \frac{m}{n}]} \, d\mu \\ &= \mu \left[\frac{m-1}{n}, \frac{m}{n} \right] \\ &= \frac{1}{n} \to 0. \end{aligned}$$

Hence, $f_{nm} \overset{\mu}{\to} 0$ by Theorem 12.3.2.

Remark 12.6.3. Although (f_n) does not converge in L^p for $(0 < p < \infty)$, it does converge in L^∞. This is because uniform convergence a.e. implies convergence in L^∞ ; see Remark 11.6.1. $\triangle$

13 § 2.6 Product measures and Fubini's theorem

13.1 Product σ-fields

Throughout this section, let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces.[60]

Definition 13.1.1. A **measurable rectangle** is a subset $A \times B$ of $X \times Y$ where $A \in \mathcal{M}$ and $B \in \mathcal{N}$ are measurable subsets of X and Y, respectively.[61] $\triangle$

Remark 13.1.1. (*Measurable rectangles need not have intervals for "sides".*) The "sides" of a measurable rectangle $A \times B$ are not required to be intervals. For instance, if $\mathbb{R}$ is equipped with the Borel σ-field, then $\mathbb{Q} \times \mathbb{Q}$ is a measurable rectangle in $\mathbb{R} \times \mathbb{R}$. $\triangle$

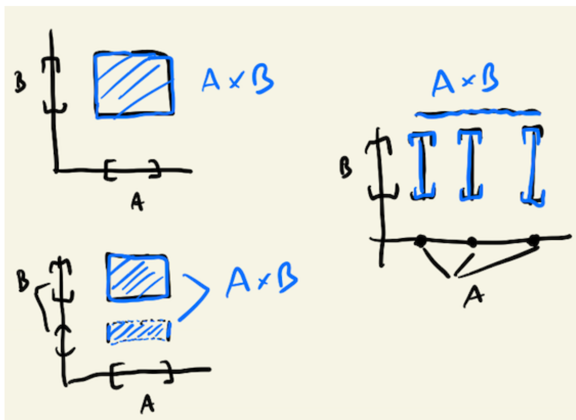


Figure 22: Three examples of measurable rectangles.

Definition 13.1.2. The **product σ-field** $\mathcal{M} \otimes \mathcal{N}$ is the σ-field on $X \times Y$ generated by the collection of all measurable rectangles.

$$\mathcal{M} \otimes \mathcal{N} := \sigma(\{A \times B : A \in \mathcal{M},\, B \in \mathcal{N}\})$$

$\triangle$

Definition 13.1.3. **Sections of sets.** Suppose that $E \subset X \times Y$. For any $x \in X$ and $y \in Y$, we define the *x-section* $E_x \subset Y$ and *y-section* $E^y \subset X$ by:

$$E_x := \{y \in Y : (x, y) \in E\}, \quad E^y := \{x \in X : (x, y) \in E\}$$

$\triangle$

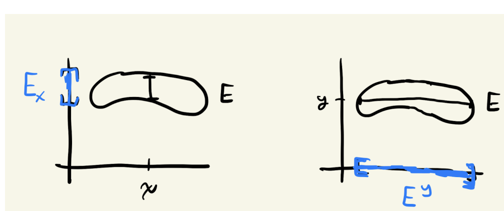


Figure 23: The *x-section* and *y-section* of a set E.

[60]For this section, we follow `https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes_ch5.pdf`. First, the notation for sections seems nicer - and more commonly used - than Ash's. Second, the presentation is more modular (e.g., the proposition that sections of a measurable set are measurable in buried in Ash's general product measure proof.)

[61]See Def. D.6.2 for the definition of the Cartesian product.

Example 13.1.1. (*Sections of rectangles.*) Given a rectangle $A \times B \subset X \times Y$, the sections are given by

$$(A \times B)_x = \begin{cases} B, & x \in A \\ \emptyset, & x \notin A \end{cases} \qquad (A \times B)^y = \begin{cases} A, & y \in B \\ \emptyset, & y \notin B \end{cases} \tag{13.1.1}$$

For intuition, see Fig. 22.

△

Proposition 13.1.1. *(All sections of a measurable set are measurable.) If $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are measurable spaces, and $E \in \mathcal{M} \otimes \mathcal{N}$, then $E_x \in \mathcal{M}$ for every $x \in X$ and $E^y \in \mathcal{N}$ for every $y \in Y$.*

Proof. We apply the "Good sets strategy" (Sec. 3.1.3). Let

$$\mathcal{G} := \{E \subset X \times Y : E_x \in \mathcal{N} \text{ for all } x \in X \quad \text{and} \quad E^y \in \mathcal{M} \text{ for all } y \in Y\}.$$

We have

- $\mathcal{G}$ contains all measurable rectangles, since the sections of $A \times B$ are given by Eq. (13.1.1).
- $\mathcal{G}$ is a σ-field, since, for example, if $E, E_i \subset X \times Y$ and $x \in X$, then

$$\underbrace{(E^c)_x}_{\text{section of complement}} = \underbrace{(E_x)^c}_{\text{complement of section}}, \qquad \underbrace{\left(\bigcup_{i=1}^{\infty} E_i\right)_x}_{\text{section of union}} = \underbrace{\bigcup_{i=1}^{\infty} (E_i)_x}_{\text{union of sections}}. \tag{13.1.2}$$

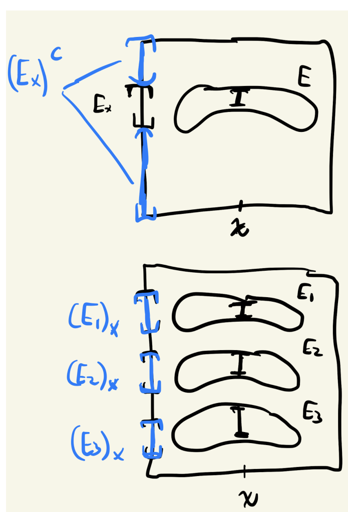


Figure 24: *Complement and union of a section*. Note the complement of the section is the section of the complement, and the union of sections is the section of the union.

□

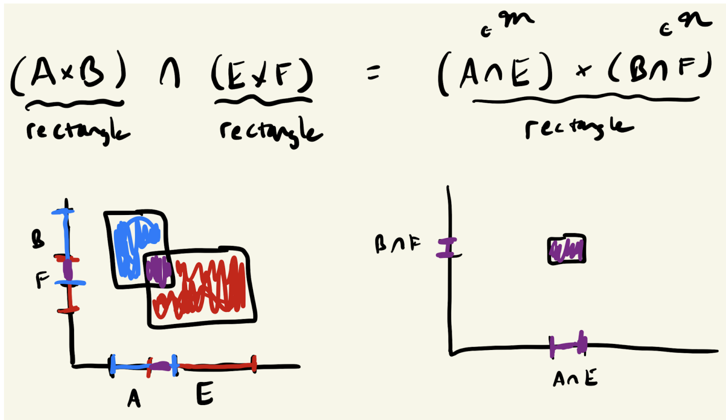
13.2 Product measure

Here we follow [Folland, 1999, pp.64]. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. We have already discussed the product σ-algebra $\mathcal{M} \otimes \mathcal{N}$ on $X \times Y$. Now we construct a measure on $\mathcal{M} \otimes \mathcal{N}$ that is, in an obvious sense, the product of μ and ν.

13.2.1 Preliminaries

Remark 13.2.1. (*Measurable rectangles are an elementary family.*)

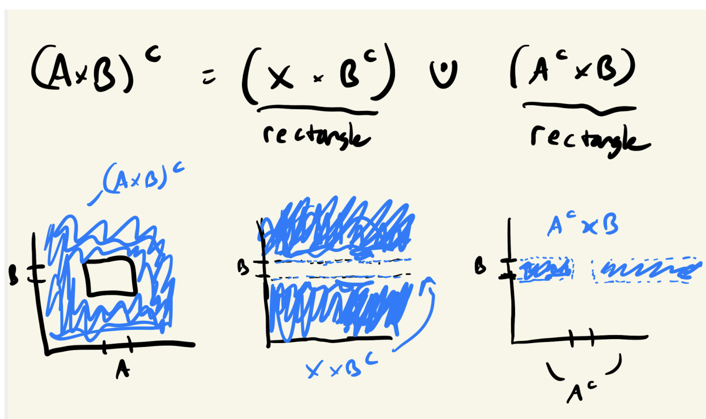
The set of measurable rectangles is an elementary family (Def. 3.1.3). First, we see that rectangles are closed under intersection.



We can formally prove this as follows:

$$\begin{aligned}(A \times B) \cap (E \times F) &= \{(x, y) : x \in A, y \in B\} \cap \{(x, y) : x \in E, y \in F\} \\ &= \{(x, y) : x \in A \cap E, y \in B \cap F\} \\ &= (A \cap E) \times (B \cap F)\end{aligned}$$

Similarly, the complement of a rectangle is a disjoint union of rectangles:

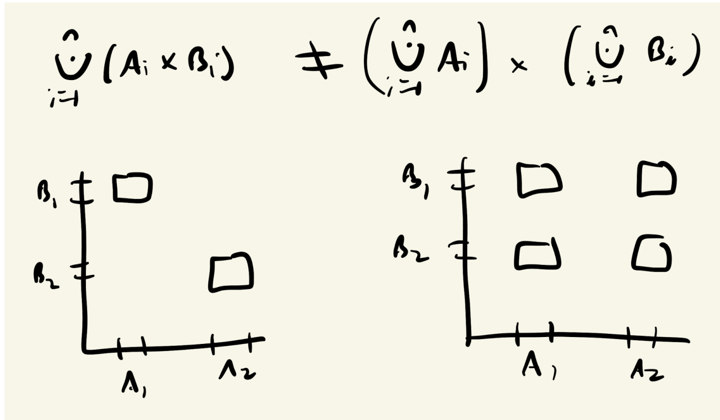


△

Remark 13.2.2. (*The collection of finite disjoint unions of rectangles is a field.*)

This follows immediately from Remark 13.2.1 and Proposition 3.1.1.

Note that this collection now includes sets that aren't rectangles



△

13.2.2 Construction of the product measure

Let $C \in F_0$ be a finite disjoint union of rectangles $A_1 \times B_1, A_2 \times B_2, \ldots A_n \times B_n$. (By Remark Rk. 13.2.2, F_0 is a field.)

Define the set function

$$\pi_0(C) := \sum_{i=1}^{n} \mu(A_i)\nu(B_i)$$

Then π_0 is well-defined, and a premeasure on $\mathcal{F}_0$ (see [Folland, 1999, pp.64] for a brief argument).

By the Carathéodory Extension Theorem, π_0 extends to a measure π on $\mathcal{F} := \sigma(\mathcal{F}_0)$. Moroever, it can be given constructively (see Eq. (4.1.1)) as follows:

$$\pi(E) = \inf\left\{\sum_{i=1}^{\infty} \pi_0(C_j) : C_j \in \mathcal{F}_0, E \subset \cup_{j=1}^{\infty} C_j\right\} \tag{13.2.1a}$$

$$\text{where } \pi_0(C_j) = \sum_{i=1}^{n_j} \mu(A_{ji})\nu(B_{ji})$$

$$= \inf\left\{\sum_{k=1}^{\infty} \mu(A_k)\nu(B_k) : A_k \in \mathcal{M}, B_k \in \mathcal{N},\ E \subset \cup_{k=1}^{\infty}(A_k \times B_k)\right\} \quad \text{by reindexing} \tag{13.2.1b}$$

Eq. (13.2.1a) is the *direct* definition of product measure, given directly by the explicit construction of a measure extended from a pre-measure. Eq. (13.2.1b) can be considered the *computational* definition of product measure; the reindexing simplifies the infimum to be directly over rectangles rather than over disjoint unions of them. For an application of Eq. (13.2.1b), see Ex. 13.5.1.)

The measure π is called the product of μ and ν, and is written $\pi = \mu \times \nu$. Note that by Carathéodory Extension theorem (or direct manipulation of the infimum), $\pi|_{\mathcal{F}_0} = \pi_0$. Note also that (obviously) $\sigma(\mathcal{F}_0) = \mathcal{M} \otimes \mathcal{N}$, the product sigma-field.

Remark 13.2.3. When $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, the product measure is uniquely determined [Folland, 1999, pp.64] △

13.3 Extended product measure theorem

Below we get an alternate view on the product measure as integrated measures of sections (see Cor. 13.3.1). But [Ash et al., 2000] actually works with a more general construction, which we might call an "extended product measure". This extended product measure seems like it will be useful for conditional probabilities.

Theorem 13.3.1. ***(Extended product measure theorem.)*** *[Ash et al., 2000, Thm. 2.6.2] Consider the measure spaces $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu(x, \cdot))$, where*

- *$\nu(x, A)$ is Borel measurable in x for each fixed $A \in A$*
- *$\nu(x, \cdot)$ are uniformly σ-finite* (i.e., $Y = \cup_{n=1}^{\infty} B_n$, where $\nu(x, B) \leq k_n$ for all $x \in X$).

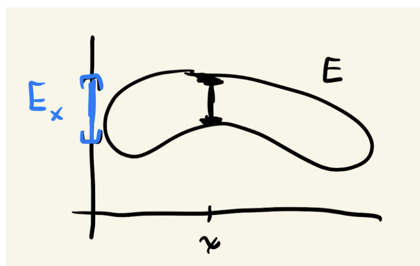
Then $\exists!$ a measure λ on $\mathcal{F} := \mathcal{M} \otimes \mathcal{N}$ such that

$$\lambda(A \times B) = \int_A \nu(x, B)\mu(dx) \qquad \forall A \in \mathcal{M}, B \in \mathcal{N}$$

Namely

$$\lambda(E) = \int_X \nu(x, E_x)\mu(dx), \qquad \forall E \in \mathcal{F} \tag{13.3.1}$$

Remark 13.3.1. Theorem 13.3.1 says that the measure of a set $E \in \mathcal{F}$ is obtained as follows: for each $x \in X$, compute the measure of E_x, the section of E at x, by $\nu(\underbrace{x}_{\text{parameter}}, \underbrace{E_x}_{\text{set to be measured}})$. Then integrate this over all x, weighted by the measure on X.



$\triangle$

Proof. First assume that the $\nu(x, \cdot)$ are finite.

(1.) If $E \in \mathcal{F}$, then $\nu(x, E_x)$ is Borel measurable in $x \in X$. (For a proof, see item (2) of the proof in [Ash et al., 2000]).[62]

(2.) Define

$$\lambda(E) = \int_A \nu(x, E_x)\mu(dx), \qquad \forall E \in F$$

The integral exists by (1.) (since the integral of a non-negative Borel measurable function always exists). Then

a) *λ is a measure on $\mathcal{F}$*. To see this note that

$$\begin{aligned}
\lambda(\cup_{n=1}^{\infty} E_n) &\overset{def}{=} \int_X \nu\Big(x, (\cup_{n=1}^{\infty} E_n)_x\Big)\mu(dx) \\
&\overset{(+)}{=} \int_X \sum_{n=1}^{\infty} \nu\Big(x, (E_n)_x\Big)\mu(dx) \\
&\overset{\text{Ash Cor 1.6.4}}{=} \sum_{n=1}^{\infty} \int_X \nu\Big(x, (E_n)_x\Big)\mu(dx) \\
&\overset{def}{=} \sum_{n=1}^{\infty} \lambda(E_n)
\end{aligned}$$

[62] TODO: Add my proof.

To justify (+), we have

$$
\begin{aligned}
(\cup_{n=1}^{\infty} E_n)_x &= \cup_{n=1}^{\infty} (E_n)_x && \text{sections commute with unions, Eq. (13.1.2)} \\
\implies \nu\Big(x, (\cup_{n=1}^{\infty} E_n)_x\Big) &= \nu\Big(x, \cup_{n=1}^{\infty} (E_n)_x\Big) && \text{substitution} \\
&= \sum_{n=1}^{\infty} \nu\Big(x, (E_n)_x\Big) && \text{countable additivity. each } (E_n)_x \text{ is measurable by Prop. 13.1.1, hence so is the union}
\end{aligned}
$$

b) Moreoever,

$$\lambda(A \times B) = \int_A \nu(x, B)\mu(dx) \qquad \forall A \in \mathcal{M}, B \in \mathcal{N}$$

To see this, note that

$$
\begin{aligned}
\lambda(A \times B) &\overset{\text{def.}}{=} \int_X \nu\Big(x, (A \times B)_x\Big)\mu(dx) \\
&\overset{(*)}{=} \int_X \nu(x, B_x) 1_{x \in A}\, \mu(dx) \\
&= \int_A \nu(x, B_x)\, \mu(dx)
\end{aligned}
$$

where (*) holds since by sections of rectangles (Example 13.1.1), we have $\nu\big(x, (A \times B)_x\big) = \nu(x, B_x) 1_{x \in A}$.

For an extension to the uniformly σ-finite case, and to show uniqueness, see [Ash et al., 2000]. □

Question 13.3.1. I'm not sure that λ in Thm 13.3.1 should be called a product measure. Other references use product measure solely to refer to the more restricted object given below in Cor. 13.3.1. So how shall we refer to λ in Thm 13.3.1 by name? As a joint measure? As an extended product measure? △

Corollary 13.3.1. ***(Classical product measure theorem.)*** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. Then the set function given by*

$$\pi(E) := \int \nu(E_x)\mu(dx) = \int \mu(E^y)\nu(dy)$$

is the unique measure on $\mathcal{M} \otimes \mathcal{N}$ such that

$$\pi(A \times B) = \mu(A)\nu(B)$$

for all $A \in \mathcal{M}, B \in \mathcal{N}$. Moreover, π is σ-finite on $\mathcal{M} \otimes \mathcal{N}$, and is a probability measure if μ and ν are. The measure π is called the product of μ and ν, and is written $\pi = \mu \times \nu$.

Proof. In Theorem 13.3.1, take $\nu(x, \cdot) = \nu$ for all $x \in X$. The second formula for π is obtained by interchanging μ and ν. □

13.4 Classical Fubini-Tonelli theorem

In this section, we follow [Folland, 1999] in restricting our consideration to classical product measures (as defined in Cor. 13.3.1). For more general statements that work with extended product measures (as defined in Theorem. 13.3.1), see [Ash et al., 2000, pp.105].

Definition 13.4.1. **(Sections of functions)** If f is a function on $X \times Y$, we define the **x-section** f_x and **y-section** f^y by

$$f_x(y) = f^y(x) = f(x, y)$$

△

So the section is obtained by fixing one variable and letting the other vary.

Example 13.4.1.

$$(1_E)_x = 1_{E_x} \quad \text{and} \quad (1_E)_y = 1_{E_y}$$

We prove the left hand side only (as the right side follows immediately by interchanging variables). By definition of sections of functions, we have

$$\left(1_E\right)_x = \left(1_{\{(x,y):(x,y)\in E\}}\right)_x \overset{\text{def. sections of functions}}{=} 1_{\{y:(x,y)\in E\}}$$

By definition of sections of sets, we have

$$1_{E_x} = 1_{\{y:(x,y)\in E\}}$$

since by definition, $E_x := \{y \in Y : (x, y) \in E\}$. △

The classical Fubini-Tonelli Theorem tells us when integrating against product measure is the same as iterated integration in either order.

Theorem 13.4.1. ***Classical Fubini-Tonelli Theorem.*** *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces.*

a) (Tonelli.) If f is a non-negative measurable function on $X \times Y$ then

$$g(x) = \int f_x \; d\nu \; , \qquad h(y) = \int f^y \; d\mu$$

are non-negative measurable functions on X and Y, respectively, and

$$\int f \; d(\mu \times \nu) \; = \int \int f(x, y) \; d\nu(y) \;\; d\mu(x) \; = \int \int f(x, y) \; d\mu(x) \;\; d\nu(y) \qquad (13.4.1)$$

b) (Fubini.) If $f \in L^1(\mu \times \nu)$ then

i) $f_x \in L^1(\nu)$ for a.e. $x \in X$, $\quad f^y \in L^1(\mu)$ for a.e. $y \in Y$

ii) $g(x) = \int f_x \; d\nu \; \in L^1(\mu)$, $\quad h(y) = \int f^y \; d\mu \; \in L^1(\nu)$

and Eq. (13.4.1) *holds.*

Proof. Our proof of Fubini-Tonelli mimics the construction of the integral.

We begin with Tonelli's, which covers the first three steps.

1. *Step 1: f is an indicator function*. Here, Tonelli's theorem reduces to the classical product measure theorem [If $f = 1_E$, then Eq. (13.4.1) becomes $\pi(E) := \int \nu(E_x)\mu(dx) = \int \mu(E^y)\nu(dy)$, which holds by Cor. 13.3.1.]

2. *Step 2: f is a non-negative simple function.* Here, Tonelli's theorem follows from step 1 and linearity. (Recall linearity always applies for non-negative simple functions; see Remark 6.3.12.)

3. *Step 3: f is a general non-negative measurable function*. We can find a sequence of simple functions $\{f_n\} \uparrow f$ by Prop. 6.3.4. We now show two things:

   - The inner integrands g, h are measurable. To see this, define corresponding sequences $\{g_n\}, \{h_n\}$ via

     $$g_n(x) = \int (f_n)_x \; d\nu \; , \qquad h_n(y) = \int (f_n)^y \; d\mu$$

     Since $f_n \uparrow f$, by Monotone Convergence Theorem, $g_n \uparrow g, h_n \uparrow h$. Thus, g, h are measurable.

- Tonelli's theorem holds. To see this, we apply Monotone Convergence Theorem again (this time on $g_n \uparrow g$, $h_n \uparrow h$), and we obtain

$$
\begin{aligned}
\int g \, d\mu &\overset{MCT}{=} \lim_n \int g_n \, d\mu \\
&\overset{def\, g_n}{=} \lim_n \int \left(\int (f_n)_x \, d\nu \right) d\mu \\
&\overset{def\ sec.\ of\ functions}{=} \lim_n \int \left(\int f_n(x,y) \, d\nu(y) \right) d\mu \, (x) \\
&\overset{Step\ 2}{=} \lim_n \int f_n \, d(\mu \times \nu) \\
&\overset{MCT}{=} \int f \, d(\mu \times \nu)
\end{aligned}
$$

which is the first equality of Eq. (13.4.1). The second equality is obtained by applying the same logic to h.

Now we do Fubini's part of the theorem. We start by verifying the conditions (i),(ii), and then show that Eq. (13.4.1) holds (which parallels "Step 4" in the construction of the integral).

i)

$$
\begin{aligned}
f \in L^1(\mu \times \nu) &\iff \int f \, d(\mu \times \nu) \; < \infty && \text{by definition} \\
&\iff \int |f| \, d(\mu \times \nu) \; < \infty \\
&\iff \int \int |f| \, d\mu \;\; d\nu \; = \int \int |f| \, d\nu \;\; d\mu \; < \infty && \text{Tonelli} \\
&\iff \int \int f \, d\mu \;\; d\nu \; = \int \int f \, d\nu \;\; d\mu \; < \infty \\
&\implies \underbrace{\int f^y \, d\mu(x)}_{:= h(y)} \; < \infty \text{ a.e. } y, \quad \underbrace{\int f_x \, d\nu(y)}_{:= g(x)} \; < \infty \text{ a.e. } x && \text{Finite integrals have finite integrands a.e.; Thm. 7.3.2} \\
&\iff f^y \in L^1(\mu) \text{ for a.e. } y \in Y, \quad f_x \in L^1(\nu) \text{ for a.e. } x \in X && \text{by definition}
\end{aligned}
$$

ii) TODO

- *Step 4:* TODO

$\square$

When integrating a function f which is not necessarily non-negative, verifying that Fubini's theorem applies requires integrating against product measure. Integrating against product measure can be difficult – product measure is defined in terms of an infimum. Remark 13.4.1 shows us a way out of this quandry.

Remark 13.4.1. (*Using Fubini and Tonelli theorems in tandem in order to avoid integrating against product measure to verify interchanging the order of integration .*) The Fubini and Tonelli theorems are frequently used in tandem. Typically one wishes to reverse the order of integration in a double integral $\int \int f \, d\mu \;\; d\nu$. *First* one verifies that $\int |f| \, d(\mu \times \nu) \; < \infty$ by using Tonelli's theorem to evaluate this integral as an iterated integral; *then* one applies Fubini's theorem to conclude that $\int \int f \, d\mu \;\; d\nu \; = \int \int f \, d\nu \;\; d\mu$. (This remark is taken verbatim from [Folland, 1999, pp.68].)

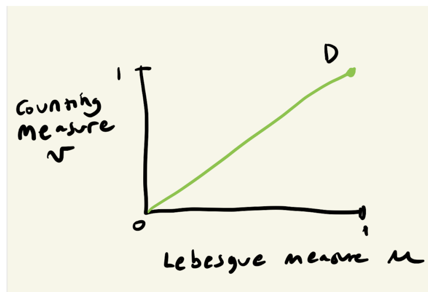
$\triangle$

13.5 Examples

The example below shows that the classical Fubini-Tonelli Theorem (Thm. 13.4.1) can fail when at least one factor in the product measure is not σ-finite. In fact, in the example below, the three kinds of integrals yield *three* distinct values![63]

Example 13.5.1. (*Lebesgue-counting measure of the diagonal of the unit square.*) [Folland, 1999, Exercise 2.46, pp.68] Let $X = Y = [0, 1], \mathcal{M} = \mathcal{N} = \mathcal{B}_{[0,1]}, \mu =$ Lebesgue measure and $\nu =$ counting measure. If $D := \{(x, x) : x \in [0, 1]\}$ is the diagonal in $X \times Y$, then

$$\int\int 1_D \; d\mu \;\; d\nu \neq \int\int 1_D \; d\nu \;\; d\mu \neq \int 1_D \; d(\mu \times \nu)$$



$\triangle$

Proof. a) First we show $\int\int 1_D \; d\mu \;\; d\nu \; = 0$.

$$\begin{aligned}
\int\int 1_D \; d\mu \;\; d\nu \; &= \int_{[0,1]} \int_{\{x:x=y\}} 1 \; d\mu(x) \;\; d\nu(y) \\
&= \int_{[0,1]} \mu(\{x : x = y\}) \; d\nu(y) \\
&= \int_{[0,1]} 0 \; d\nu(y) \;\; = 0
\end{aligned}$$

b) Next, we show $\int\int 1_D \; d\nu \;\; d\mu \; = 1$.

$$\begin{aligned}
\int\int 1_D \; d\nu \;\; d\mu \; &= \int_{[0,1]} \int_{\{y:y=x\}} 1 \; d\nu(y) \;\; d\mu(x) \\
&= \int_{[0,1]} \nu(\{y : y = x\}) \; d\mu(x) \\
&= \int_{[0,1]} 1 \; d\mu(x) \;\; = \mu([0, 1]) = 1
\end{aligned}$$

c) Now we show $\int 1_D \;\; d(\mu \times \nu) \;\; = \infty$. By the definition of integrals of simple functions, we have $\int 1_D \;\; d(\mu \times \nu) \;\; = \mu \times \nu(D)$. So applying the (computational) definition of product measure Eq. (13.2.1b), we want to show

$$\mu \times \nu(D) := \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n)\nu(B_n) : \cup_{n=1}^{\infty}(A_n \times B_n) \supseteq D, \;\; A_n, B_n \in \mathcal{B}(\mathbb{R}) \right\} = \infty \tag{13.5.1}$$

[63] I like this example because it illustrates how integrating against product measure is conceptually *distinct* from iterated integrals, rather than defined in terms of them (as sometimes seems to be the case in calculus).

To do this, we first observe that

$$\cup_{n=1}^{\infty}(A_n \times B_n) \supseteq D \implies \cup_{n=1}^{\infty}(A_n \cap B_n) \supseteq [0,1] \qquad 1.$$

1 holds because

$$\cup_{n=1}^{\infty}(A_n \times B_n) = \{(x,y) : x \in A_n \text{ and } y \in B_n \text{ for some } n\}$$

So if $\cup_{n=1}^{\infty}(A_n \times B_n) \supseteq D := \{(x,x) : x \in [0,1]\}$, then

$$\forall x \in [0,1], \exists n : x \in A_n \text{ and } y \in B_n$$

In other words, the implication holds.

Next we observe that

$$\cup_{n=1}^{\infty}(A_n \times B_n) \supseteq D \implies \exists N : \mu(A_N), \mu(B_N) \geq \epsilon > 0 \qquad 2.$$

2 holds because

$$\begin{aligned}
1 \implies 1 = \mu([0,1]) &\overset{\textit{monotonicity}}{\leq} \mu(\cup_{n=1}^{\infty} A_n \cap B_n) \overset{\textit{subadditivity}}{\leq} \sum_{n=1}^{\infty} \mu(A_n \cap B_n) \\
&\implies \exists N : \mu(A_N \cap B_N) \geq \epsilon > 0 && \text{True by contraposition} \\
&\implies \exists N : \mu(A_N), \mu(B_N) \geq \epsilon > 0 && \text{by monotonicity, since } A_N \cap B_N \subseteq A_N, B_N
\end{aligned}$$

Finally we prove Eq. (13.5.1). For any $\cup_{n=1}^{\infty}(A_n \times B_n) \supseteq D$, we have

$$\begin{aligned}
\mu(B_N) \overset{2}{\geq} \epsilon > 0 &\implies B_N \text{ has (uncountably) infinitely many points} && \text{By contraposition. Otherwise } \mu(B_N) = 0. \\
&\implies \nu(B_N) = \infty && \text{Def. counting measure } \nu. \\
&\implies \mu(A_N)\nu(B_N) \geq \epsilon \cdot \infty = \infty && \mu(A_N) \geq \epsilon \text{ by } 2. \\
&\implies \sum_{n=1}^{\infty} \mu(A_n)\nu(B_n) = \infty && \text{non-negativity of measure} \\
&\implies Eq. \text{ (13.5.1) holds} && \text{Infimum in Eq. (13.5.1) must be lower bound on the set } \{\infty\}
\end{aligned}$$

□

The example below demonstrates a case where we are unable to swap the order of a series. We can read this example through Fubini-Tonelli by seeing each series as an integral against counting measure.

Example 13.5.2. (*A function defined on a bivariate grid of natural numbers for which the Fubini-Tonelli Theorem fails.*) [Folland, 1999, Exercise 2.48, pp.69] Let $X = Y = \mathbb{N}, \mathcal{M} = \mathcal{N} = \mathcal{P}(\mathbb{N}), \mu = \nu =$ counting measure. Define $f(m,n) = 1$ if $m = n$, $f(m,n) = -1$ if $m = n+1$, and $f(m,n) = 0$ otherwise. Then $\int |f| d(\mu \times \nu) = \infty$, and $\int \int f \; d\mu \;\; d\nu$ and $\int \int f \; d\nu \;\; d\mu$ exist and are unequal.

$$\begin{array}{cccccc}
 & \vdots & \vdots & \vdots & \vdots & \\
 & 0 & 0 & 0 & 1 & \dots \\
\uparrow & 0 & 0 & 1 & -1 & \dots \\
n & 0 & 1 & -1 & 0 & \dots \\
 & 1 & -1 & 0 & 0 & \dots \\
 & & m & \rightarrow & &
\end{array}$$

Figure 25: A function defined on a bivariate grid of natural numbers for which the Fubini-Tonelli Theorem fails. When integrating against counting measure, the two iterated integrals have the form of sums. In the words of [Durrett, 2010], "[...] If we sum the columns first, the first one gives us a 1 and the others 0, while if we sum the rows each one gives us a 0." Hence, $\sum_m \sum_n f(m,n) = 1$, but $\sum_n \sum_m f(m,n) = 0$. Image from [Durrett, 2010].

$\triangle$

Proof. a) First we show $\int \int f \, d\mu \ \ d\nu \ = 0$.

$$\int \int f(m,n) \, d\mu(m) \ \ d\nu(n) \ = \int 0 \, d\nu(n)$$
$$= 0 \cdot \nu(\mathbb{N}) = 0 \cdot \infty = 0.$$

b) Next we show $\int \int f \, d\nu \ \ d\mu \ = 1$.

$$\int \int f(m,n) \, d\nu(n) \ \ d\mu(m) \ = \int (1 \cdot 1_{m=1} + 0 \cdot 1_{m>1}) \, d\mu(m)$$
$$\overset{\textit{int. simple functions}}{=} 1 \cdot \mu(m=1) + 0 \cdot \mu(m>1) = 1.$$

c) Finally, we show $\int |f| \, d(\mu \times \nu) \ = \infty$, which is why Fubini-Tonelli does not apply.

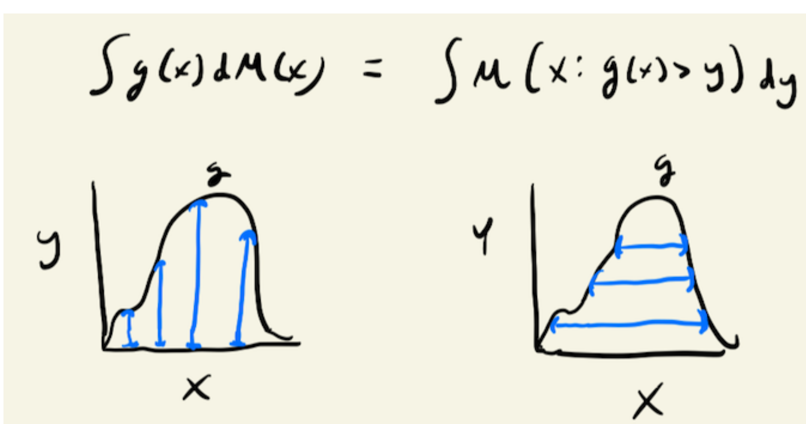
$$\begin{aligned}
\int |f| \, d(\mu \times \nu) \ &= \int 1_{m=n \text{ or } m=n+1} \, d(\mu \times \nu) \\
&= (\mu \times \nu)\{(m,n) : m = n \text{ or } m = n+1\} && \text{integral of indicator} \\
&= \sum_{m=1}^{\infty} (\mu \times \nu)\{(m,m)\} + \sum_{m=2}^{\infty} (\mu \times \nu)\{(m,m-1)\} && \text{countable additivity} \\
&= \sum_{m=1}^{\infty} \mu\{m\}\nu\{m\} + \sum_{m=2}^{\infty} \mu\{m\}\nu\{m-1\} && \mu \times \nu \text{ factorizes across rectangles by construction; see Sec.13.2.2} \\
&= \sum_{m=1}^{\infty} 1 \cdot 1 + \sum_{m=2}^{\infty} 1 \cdot 1 = \infty
\end{aligned}$$

$\square$

Example 13.5.3. (*Area under the curve can be obtained by either horizontal or vertical sections.*) [Durrett, 2010, Exercise 1.7.2] Let $g \geq 0$ be a measurable function on a sigma-finite measure space $(\Omega, \mathcal{F}, \mu)$. Use the classical Fubini-Tonelli Theorem (Thm. 13.4.1) to conclude that

$$\int_{\Omega} g \, d\mu \ = (\mu \times \lambda)\{(x,y) : 0 \leq y < g(x)\} = \int_0^{\infty} \mu(\{x : g(x) > y\}) \, dy$$

for some choice of measure λ.



$\triangle$

Proof. Let λ be Lebesgue measure on $\mathbb{R}$. Then

$$\begin{aligned}
(\mu \times \lambda)\{(x,y) : 0 \leq y < g(x)\} &= \int\int 1_{\{(x,y):0\leq y<g(x)\}}\ d\mu(x)\ \ d\lambda(y) && \text{Tonelli} \\
&= \int 1_{\{y:y\geq 0\}} \int 1_{\{x:g(x)>y\}}\ d\mu(x)\ \ d\lambda(y) && \text{Rewrite indicator; constant multiple} \\
&= \int 1_{\{y:y\geq 0\}} \mu(x : g(x) > y)\ d\lambda(y) && \text{Integral of indicator} \\
&= \int_0^\infty \mu(x : g(x) > y) dy && \lambda \text{ is Lebesgue measure}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
(\mu \times \lambda)\{(x,y) : 0 \leq y < g(x)\} &= \int\int 1_{\{(x,y):0\leq y<g(x)\}}\ d\lambda(y)\ \ d\mu(x) && \text{Tonelli} \\
&= \int \lambda(y : 0 \leq y < g(x))\ d\mu(x) && \text{Integral of indicator} \\
&= \int g(x)\ d\mu(x) && \lambda \text{ is Lebesgue measure}
\end{aligned}$$

□

Remark 13.5.1. In probability theory, the argument of Example 13.5.3 can be used to show that if X is a non-negative random variable, then its expected value equals the integral of its survival function, i.e.

$$\mathbb{E}[X] = \int_0^\infty P(X > x) dx$$

For details, see Remark 16.5.3. △

13.6 The n-fold product measure

13.6.1 Construction

The construction of the n-fold product measure [Folland, 1999, pp.65] is obtained through the virtually the same procedure as the construction of the 2-fold product measure.

Let $(\Omega_j, \mathcal{F}_j, \mu_j)$ be measure spaces for $j = 1, \ldots, n$. Let a rectangle be $A_1 \times \ldots \times A_n$ for $A_j \in \mathcal{F}_j$. Then the collection of finite disjoint union of rectangles is a field, using a similar argument as given in Remark 13.2.2. So following the same procedure as in the construction of the 2-fold product measure (Sec. 13.2.2), we obtain a measure $\mu_1 \times \ldots \times \mu_n$ on $\mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_n$ such that

$$\mu_1 \times \ldots \times \mu_n(A_1 \times \ldots \times A_n) = \prod_{j=1}^n \mu_j(A_j)$$

13.7 Lebesgue measure on $\mathbb{R}^n$

As an example of an n-fold product measure, Lebesgue measure on $\mathbb{R}^n$ is the n-fold product of Lebesgue measure on $\mathbb{R}$ with itself. For convenience, it is sometimes instead defined as the completion of that n-fold product.

Example 13.7.1. (*Lebesgue measure on* $\mathbb{R}^n$.) When $(\Omega_j, \mathcal{F}_j, \mu_j) = (\mathbb{R}, \mathcal{B}(\mathbb{R})$ or $(\mathbb{R})$, Lebesgue measure$)$, the n-fold product measure $\mu_1 \times \ldots \times \mu_n$ on $\mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R})$ (or equivalently $(\mathbb{R}) \otimes \ldots \otimes (\mathbb{R})$) is called *Lebesgue measure on* $\mathbb{R}^n$. △

It is often convenient to define Lebesgue measure as the *completion* of this n-fold product. The domain of the resulting measure is the class of *Lebesgue measurable sets* in $\mathbb{R}^n$, and is denoted ^{n}.

Question 13.7.1. It is interesting that the product of complete measures is not necessarily complete (and so must be completed again). Give an example illustrating why this is necessary. △

13.7.1 Approximation properties

Here we show that we can approximate *any* Lebesgue measurable set in $\mathbb{R}^n$ with simpler sets: open sets, compact sets, of finite collections of disjoint rectangles.

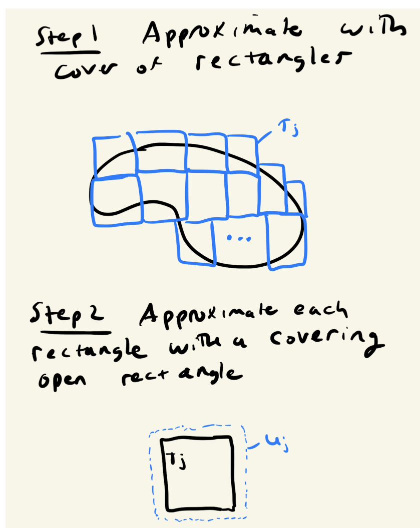
Theorem 13.7.1. *[Folland, 1999, Thm 2.40a,c]. Let m be Lebesgue measure on $\mathbb{R}^n$. Let $E \in {}^n$. Then*

a) $m(E) = \inf\{m(U) : U \supset E, U \text{ open }\} = \sup\{m(K) : K \subset E, K \text{ compact }\}$.

b) If $m(E) < \infty$ for any $\epsilon > 0$, there is a finite collection $\{R_j\}_1^N$ of disjoint rectangles whose sides are intervals such that $m(E \triangle \cup_{j=1}^N R_j) < \epsilon$.

Proof. We prove the first equality of part (a). For the remainder of the proof, see [Folland, 1999, pp.70].

Our strategy is summarized in the Figure below.

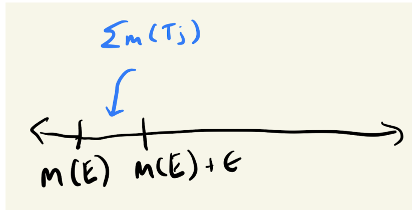


First, we note that by monotonicity, $m(E)$ is a lower bound on the set $\{m(U) : U \supset E, U \text{ open }\}$. So it remains to show that $m(E)$ is the *greatest* lower bound. We proceed in steps.

1. *Each $E \in {}^n$ can be approximated by a cover of countable rectangles.*

   By definition of the product measure (as the *infimum* of measures of countable rectangle covers; see Eq. (13.2.1b); also see Remark A.1.2 for a refresher on infima), we have $\forall \epsilon > 0$, there is a countable family $\{T_j\}_{j=1}^\infty$ of rectangles such that $\{T_j\}_{j=1}^\infty \supset E$ and

$$\underbrace{\sum_{j=1}^{\infty} m(T_j)}_{\text{... 2) and there is a countable union of rectangles with smaller measure}} \le m(E) + \underbrace{\epsilon}_{\text{1) bump a little bit up from infimum}}$$



2. *Each such rectangle can be approximated by a covering open rectangle (by applying the 1-dim case to each side).*

   Now for each rectangle T_j apply the one-dimensional theorem allowing open-set-approximations [Folland, 1999, Thm 1.18] to each side to find a rectangle $U_j \supset T_j$ whose sides are open sets such that
   $$m(U_j) \leq m(T_j) + \epsilon 2^{-j}$$

3. *The lower bound $m(E)$ is the greatest lower bound.*

   First note that the countable union of open rectangle covers $U := \cup_{j=1}^{\infty} U_j$ is an open set. Moreover $U \supset E$. For all $\epsilon > 0$, we have

   $$m(U) \overset{\textit{subadditivity}}{\leq} \sum_{j=1}^{\infty} m(U_j) \overset{\textit{Step 2}}{\leq} \sum_{j=1}^{\infty} m(T_j) + \epsilon \overset{\textit{Step 1}}{\leq} m(E) + 2\epsilon,$$

   which by Remark A.1.2 proves that the lower bound is the greatest lower bound.

□

14 § 2.7 Measures on infinite product spaces

14.1 § 2.7.1-2.7.3: Measures on countably infinite product spaces (i.e., on *sequences*)

In this section, we define a measure on countably infinite product spaces.

The construction is syntactically similar to how we constructed Lebesgue measure. Recall that for Lebesgue measure, we defined a pre-measure on an elementary family (the r.s.c intervals), which gave us a pre-measure on a field (finite disjoint unions of r.s.c intervals), and then we used the Carathéodory extension theorem to obtain a measure on a σ-field.

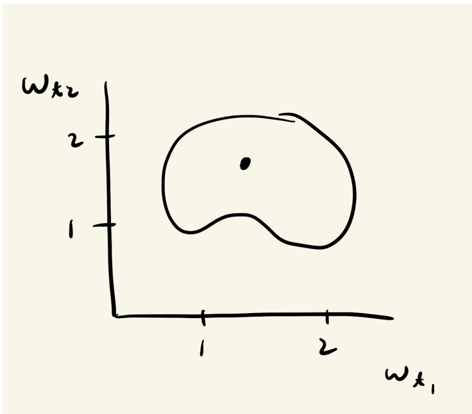
We followed a similar process when constructing the product measure.

Similarly, here, we:

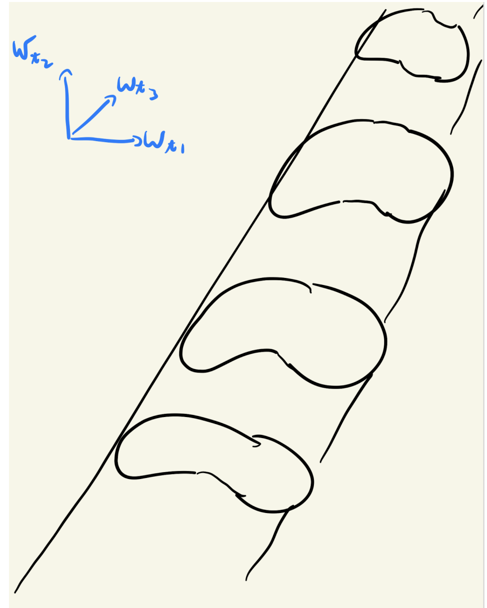
- Define an elementary family that is convenient to work with.
- Extend that elementary family to a field.
- Define a pre-measure on the field.
- Use the Carathéodory extension theorem to extend the pre-measure to a measure on a σ-field.

14.1.1 Cylinders and rectangles

Throughout, for each $j = 1, 2, \ldots$, let $(\Omega_j, \mathcal{F}_j)$ be a measurable space. Let $\Omega := \prod_{j=1}^{\infty} \Omega_j$, the set of all sequences $(\omega_1, \omega_2, \ldots)$ such that $\omega_j \in \Omega_j$.



(a) A base (in 2 coordinates) for a cylinder.



(b) The cylinder projected to the first three coordinates.

Figure 26: A cylinder in a countably infinite product space (i.e., a sequence space).

Definition 14.1.1[64]**.** Given $v := (t_1, \ldots, t_n)$, a finite subset of $\{1, 2, \ldots\}$, and a set $B_v \in \Omega_v := \prod_{i=1}^n \Omega_{t_i}$, we define a **cylinder** as

$$\underbrace{\widehat{B_v}}_{\text{cylinder}} := \{\omega \in \Omega : (\omega_{t_1}, \ldots \omega_{t_n}) \in \underbrace{B_v}_{\text{base}}\} \tag{14.1.1}$$

If $B_v \in \mathcal{F}_v := \prod_{i=1}^n \mathcal{F}_{t_i}$, we call this a **measurable cylinder**. △

Definition 14.1.2. A cylinder whose base is a rectangle with n sides (i.e. $B_v = A_i \times \ldots \times A_n$ with $A_i \in \Omega_{t_i}$ for all i) is called a **rectangle**. If $A_i \in \mathcal{F}_{t_i}$ for each i, the cylinder is called a **measurable rectangle**. △

Example 14.1.1. Let $\Omega_i = \mathbb{R}$ for all i (written $\Omega = \mathbb{R}^{\mathbb{N}}$), and let B_v be the base given in Fig. 26a. A cylinder can be defined for any choice of $v = (t_1, t_2)$ by

$$\widehat{B_v} := \{\omega \in \Omega : (\omega_{t_1}, \omega_{t_2}) \in B_2\} \tag{14.1.2}$$

For instance, $(1.5, 1.5, \omega_3, \omega_4, \ldots) \in \widehat{B}_{(1,2)}$, and $(\omega_1, 1.5, \omega_3, 1.5, \omega_5, \ldots) \in \widehat{B}_{(2,4)}$.

The projection of the cylinder $\widehat{B_v}$ onto $(\omega_{t_1}, \omega_{t_2}, \omega_{t_3})$ is given in Fig. 26b.

△

14.1.2 Measurable cylinders are a field.

Remark 14.1.1. (*A cylinder can always be regarded as having a higher dimensional base.*) For example, given a base $B \subset \mathbb{R}^3$, we can write:

$$\widehat{B_{1:3}} = [\widehat{B_{1:3} \times \Omega_4}]$$

[64] Note that in §2.7.1-2.7.3, [Ash et al., 2000] uses B_n for cylinder and B^n for base. However, I find the superscript/-subscript notation confusing. Therefore, I use a different notation, which actually comes from §2.7.4-2.7.5, the uncountable product section of [Ash et al., 2000]. The notation $B_n(v)$ for a cylinder can be interpreted as the set B_n located at coordinates $v = (t_1, \ldots, t_n)$.

since

$$\begin{aligned}
\widehat{B_{1:3}} &= \{\omega \in \Omega : (\omega_1, \omega_2, \omega_3) \in B_{1:3}\} \\
&= \{\omega \in \Omega : (\omega_1, \omega_2, \omega_3) \in B_{1:3}, \omega_4 \in \Omega_4\} \\
&= \{\omega \in \Omega : (\omega_1, \omega_2, \omega_3, \omega_4) \in B_{1:3} \times \Omega_4\} \\
&= [\widehat{B_{1:3} \times \Omega_4}]
\end{aligned}$$

△

Remark 14.1.2. (*Measurable cylinders are an elementary family.*)

Let $\mathcal{C} \subset \Omega$ be the measurable cylinders. We verify $\mathcal{C}$ satisfies the definition of an elementary family (Def. 3.1.3):

a) $\emptyset \in \mathcal{C}$? ✓

b) if $\widehat{B_v} \in \mathcal{C}$, then $(\widehat{B_v})^c$ is a finite disjoint union of members of $\mathcal{C}$? ✓ In fact, if $\widehat{B_v} \in \mathcal{C}$, then $(\widehat{B_v})^c \in \mathcal{C}$. To see this, observe that given a base $B_v \in \mathcal{F}_v$, we have

$$\begin{aligned}
(\widehat{B_v})^c &= \{\omega \in \Omega : (\omega_{t_1}, \dots \omega_{t_n}) \in B_v\}^c \\
&= \{\omega \in \Omega : (\omega_{t_1}, \dots \omega_{t_n}) \in \underbrace{(B_v)^c}_{\in \mathcal{F}_v, \text{ since } \mathcal{F}_v \text{ is a } \sigma\text{-field}} \}
\end{aligned}$$

c) if $\widehat{B_v}, \widehat{A_w} \in \mathcal{C}$ then $\widehat{B_v} \cap \widehat{A_w} \in \mathcal{C}$? ✓ By Remark 14.1.1, we can assume that $v = w$, so it remains to show that if $\widehat{B_v}, \widehat{A_v} \in \mathcal{C}$ then $\widehat{B_v} \cap \widehat{A_v} \in \mathcal{C}$. This holds because

$$\widehat{B_v} \cap \widehat{A_v} = \{\omega \in \Omega : (\omega_{t_1}, \dots \omega_{t_n}) \in \underbrace{B_v \cap A_v}_{\in \mathcal{F}_v, \text{ since } \mathcal{F}_v \text{ is a } \sigma\text{-field}} \}$$

△

Remark 14.1.3. (*Measurable cylinders are a field.*) By Remark 14.1.2, measurable cylinders are an elementary family. So by Prop. 3.1.1, the collection of finite disjoint unions of measurable cylinders are a field. But finite disjoint unions of measurable cylinders are just measurable cylinders. (This can be seen by a similar argument as in part c) of Remark 14.1.2.) Hence, the measurable cylinders are a field. △

14.1.3 (Extended) product measure theorem over countably infinite factors

Now we provide a version of the (extended) product measure theorem over countably many factors. This extends Theorem 13.3.1, which worked with 2 factors, and [Ash et al., 2000, Thm 2.6.7], which worked with n factors.

We restrict consideration here to *probability* measures. The theorem does not hold for arbitrary measures.

We let $\prod_{i=1}^\infty \mathcal{F}_i$ be the smallest σ-field over measurable cylinders. (This is also the smallest σ-field over measurable rectangles; see [Ash et al., 2000, Sec 2.7, Problem 1].)

Due to Remark 14.1.1, we will assume WLOG in the Theorem that $v = (1, \dots, n)$, and we will write $B_{1:n} (\in \prod_{j=1}^n \Omega_j)$ simply as B_n.

Theorem 14.1.1. *Let $(\Omega_j, \mathcal{F}_j)$, $j = 1, 2, \dots$ be arbitrary measurable spaces. Let $\Omega = \prod_{j=1}^\infty \Omega_j$ and $\mathcal{F} = \prod_{j=1}^\infty \mathcal{F}_j$. Suppose we are given an arbitrary probability measure P_1 on $\mathcal{F}_1$, and for each $j = 1, 2, \dots$ and each $(\omega_1, \dots, \omega_j) \in \Omega_1 \times \cdots \Omega_j$ we are given a probability measure*

$P_{j+1}(\omega_1, \ldots, \omega_j; \cdot)$ on $\mathcal{F}_{j+1}$. Assume that for each fixed $C \in \mathcal{F}_{j+1}$ the function $P_{j+1}(\cdot, \ldots, \cdot; C)$: $(\prod_{i=1}^j \Omega_i, \prod_{i=1}^j \mathcal{F}_i,) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable.[65]

For $B_n \in \prod_{j=1}^n \mathcal{F}_j$, define

$$P_{1:n}(B_n) = \int_{\Omega_1} P_1(\ d\omega_1\ ) \int_{\Omega_2} P_2(\omega_1;\ d\omega_2\ ) \cdots \int_{\Omega_{n-1}} P_{n-1}(\omega_1, \ldots \omega_{n-2};\ d\omega_{n-1}\ )$$
$$\int_{\Omega_n} P_n(\omega_1, \ldots \omega_{n-1};\ d\omega_n\ )\ 1_{(\omega_1, \ldots, \omega_n) \in B_n} \tag{14.1.3}$$

Note that $P_{1:n}$ is a probability measure (on $(\prod_{j=1}^n \Omega_j, \prod_{j=1}^n \mathcal{F}_j)$) by [Ash et al., 2000, Thm. 2.6.7].

Then there is a unique probability measure P on $\mathcal{F}$ such that for all n, P agrees with $P_{1:n}$ on n-dimensional cylinders; that is $P(\underbrace{\widehat{B_n}}_{\text{cylinder}}) = P_{1:n}(\underbrace{B_n}_{\text{base}})$ for all $B_n \in \prod_{j=1}^n \mathcal{F}_j$ and $n = 1, 2, \ldots$.

Proof. 1. *We first show that P is well-defined on measurable cylinders.*

Recall from Remark 14.1.1 that a cylinder can have multiple representations (since we can always increase the dimensionality of the base).

Suppose $\widehat{B_n} = \widehat{C_m}$ for $m < n$.[66] Then we can relate the bases of the cylinders by $B_n = C_m \times \Omega_{m+1} \times \cdots \times \Omega_n$.

We need to show that the probability mass assigned to the cylinder is invariant to representation, i.e. $P(\widehat{B_n}) = P(\widehat{C_m}) \overset{(def\ P)}{\iff} P_{1:n}(B_n) = P_{1:m}(C_m)$.

Observe[67]:

$$\begin{aligned}
P_{1:n}(B_n) &= \int_{\Omega_{1:m}} \left[\int_{\Omega_{m+1:n}} 1_{\omega_{1:m} \in C_m} \cancel{1_{\omega_{m+1:n} \in \Omega_{m+1:n}}} P_{m+1:n}(\omega_{1:m};\ d\omega_{m+1:n}\ ) \right] P_{1:m}(\ d\omega_{1:m}\ ) && \text{def } P_n\text{, Tonelli compression} \\
&= \int_{\Omega_{1:m}} 1_{\omega_{1:m} \in C_m} \underbrace{\left[\int_{\Omega_{m+1:n}} P_{m+1:n}(\omega_{1:m};\ d\omega_{m+1:n}\ ) \right]}_{=1} P_{1:m}(\ d\omega_{1:m}\ ) && \text{constant multiple} \\
&= P_{1:m}(C_m) && \text{Tonelli expansion, def } P_m
\end{aligned}$$

2. *P is finitely additive on the field $\mathcal{F}_0$ of measurable cylinders.*

   Consider a collection of finitely many disjoint cylinders. Represent the bases in a common number of factors, n, using Remark 14.1.1. Now disjoint cylinders have disjoint bases, so apply the finite additivity of $P_{1:n}$ obtain the finite additivity of P.

3. *If we can show that P is countably additive on $\mathcal{F}_0$, then the Carathéodory extension theorem extends P to a probability measure on $\prod_{j=1}^\infty \mathcal{F}_j$ which agrees with $P_{1:n}$ on n-dimensional cylinders.*

---

[65] Unlike [Ash et al., 2000], we use a subscript (e.g. P_n) to differentiate the probability measures over different spaces. We justify this because sometimes these probability measures become independent of certain parameters; e.g. for Markov kernels, $P_n(\omega_1, \ldots \omega_{n-1};\ d\omega_n\ )$ would simplify to $P_n(\omega_{n-1};\ d\omega_n\ )$; without the subscript, one could perhaps mistakenly believe that the quantity represented an evaluation of $P_2(\cdot, \cdot)$.

[66] Recall from the definition of cylinders that

$$\widehat{B_n} = \{\omega \in \Omega : (\omega_1, \ldots, \omega_n) \in B_n\}, \quad B_n \in \prod_{i=1}^n \mathcal{F}_i$$
$$\widehat{C_m} = \{\omega \in \Omega : (\omega_1, \ldots, \omega_n) \in C_m\}, \quad C_m \in \prod_{i=1}^m \mathcal{F}_i$$

[67] Here we use the fact that $P(\omega_1, \ldots, \omega_j; \cdot)$ are probability measures. This theorem does not hold for arbitrary measures

(a) *P is countably additive on $\mathcal{F}_0$.* Since P is finitely additive on $\mathcal{F}_0$, Thm. 3.4.2 tells us that it suffices to show that P is continuous from above at $\emptyset$.

Let $\{\widehat{B_n}\}, n = n_1, n_2, \ldots$ be a sequence of measurable cylinders decreasing to $\emptyset$. We may assume $n_1 < n_2 < \ldots$[68], and in fact nothing is lost if we assume $n_i = i$ for all i.

For each $n > 1$, decompose Eq. (14.1.3) to write

$$P_{1:n}(B_n) = \int_{\Omega_1} g_n(\omega_1) P_1(\ d\omega_1\ )$$

where

$$g_n(\omega_1) = \int_{\Omega_2} P_2(\omega_1;\ d\omega_2\ ) \int_{\Omega_n} P_{n-1}(\omega_1, \ldots \omega_{n-1};\ d\omega_n\ )\, 1_{(\omega_1,\ldots,\omega_n)\in B_n} \quad (14.1.4)$$

Now we have

$$\begin{aligned} & \widehat{B_{n+1}} \subset \widehat{B_n} && \text{decreasing seq. of cylinders} \\ \implies & B_{n+1} \subset B_n \times \Omega_{n+1} && \text{Remark 14.1.1, and cylinder subset} \implies \text{base subset} \\ \implies & 1_{(\omega_1,\ldots,\omega_{n+1})\in B_{n+1}} \leq 1_{(\omega_1,\ldots,\omega_{n+1})\in B_n \times \Omega_{n+1}} && \\ \implies & g_n(\omega_1) \leq g_{n+1}(\omega_1) && \text{monotonicity of } \textstyle\int\text{, and using an argument like Step 1 to align dimensions in def. } g_n\text{, Eq. (14.1.4)} \end{aligned}$$

Since g_n is a decreasing sequence of functions, it has a limit; say h. Thus, by monotone convergence theorem,

$$P(B_n) \to \int h(\omega_1) P_1(\ d\omega_1\ )$$

Now BWOC, assume that $\lim_{n\to\infty} P_{1:n}(B_n) > 0$. Then (by monotonicity) $h(\omega_1') > 0$ for some $\omega_1' \in \Omega_1$. In fact, $\omega_1' \in B_1$. (See [Ash et al., 2000] for a one sentence argument.)

Repeat this inductively (See [Ash et al., 2000]) to obtain points $(\omega_1', \omega_2', \ldots)$ such that for each n, $(\omega_1', \ldots, \omega_n') \in B_n$. Hence $(\omega_1', \omega_2', \ldots) \in \cap_{n=1}^{\infty} \widehat{B_n} = \emptyset$, a contradiction.

4. *P is unique.* This follows immediately from the Carathéodory extension theorem. See [Ash et al., 2000].

□

Remark-for-data-scientists 14.1.1. (*Application to Markov processes.*) Consider the setting of Thm. 14.1.1, where we build up joint probability measures on multivariate or infinite dimensional spaces by chaining together probability measures parametrized by the preceding coordinates. When those probability measures are parametrized by only the immediately preceding coordinate – that is, when $P_n(\omega_1, \ldots \omega_{n-1};\ d\omega_n\ )$ simplifies to $P_n(\omega_{n-1};\ d\omega_n\ )$ for all n – these probability measures are called **Markov kernels** or **transition functions**. Thus, one way to define joint measures is by taking products of Markov kernels.[69] △

14.2 § 2.7.4 - 2.7.5: Measures on uncountably infinite product spaces

Now for t in the arbitrary index set T, let $(\Omega_t, \mathcal{F}_t)$ be a measurable space. Let $\prod_{t\in T} \Omega_t$ be the set of all functions $\omega(t), t \in T$ such that $\omega(t) \in \Omega_t$ for each $t \in T$.

We now consider the problem of constructing probability measures on $\prod_{t\in T} \mathcal{F}_t$. The approach will be as follows. Let $v = (t_1, \ldots, t_n)$ be a finite subset of T, where $t_1 < t_2 < \ldots < t_n$. Assume that for each such v, we are given a probability measure P_v on $\prod_{i=1}^{n} \mathcal{F}_{t_i}$. $P_v(B)$ is to represent

[68] Note that the base growing in dimension from n to $n+1$ means that we *restrict* the extra coordinates relative to Ω_{n+1}.

[69] See https://mathoverflow.net/questions/369747/defining-measures-through-products-of-markov-kernels.

$P(\omega \in \prod_{t\in T} \Omega_t : (\omega(t_1), \dots, \omega(t_n)) \in B)$. We shall require the P_v to be "consistent". To define the consistency needed, we must first define the projection of a probability measure. Note that below, the space $(\prod_{i=1}^n \Omega_{t_i}, \prod_{i=1}^n \mathcal{F}_{t_i})$ will be denoted $(\Omega_v, \mathcal{F}_v)$.

Definition 14.2.1. Let P_v be a probability measure on $\mathcal{F}_v$ and $u \subset v$. The **projection** of P_v on $\mathcal{F}_u$ is the probability measure $\pi_u(P_v)$ on $\mathcal{F}_u$ defined by

$$[\pi_u(P_v)](B) = P_v(\omega \in \Omega_v : y_u \in B), \quad B \in F_u.$$

Similarly, if Q is a probability measure on $\prod_{t\in T} \mathcal{F}_t$, the projection of Q on $\mathcal{F}_v$ is given by

$$[\pi_v(Q)](B) = Q(\omega \in \prod_{t\in T} \Omega_t : \omega_v \in B), \quad B \in F_v.$$

$\triangle$

Now we can provide the Kolmogorov extension theorem. It can be proved when each Ω_t is a complete, separable metric space, with $\mathcal{F}_t$ the class of Borel sets (the σ-field generated by the open sets). However, to avoid serious technical complications, [Ash et al., 2000] takes all Ω_t to be $\mathbb{R}$ and $\mathcal{F}_t = \mathcal{B}(\mathbb{R})$.

Theorem 14.2.1. (Kolmogorov Extension Theorem.) *For each t in the arbitrary index set T, let $\Omega_t = \mathbb{R}$ and $\mathcal{F}_t$ be the Borel sets of $\mathbb{R}$.*

Assume that for each finite nonempty subset v of T, we are given a probability measure P_v on $\mathcal{F}_v$. Assume the P_v are consistent; that is, $\pi_u(P_v) = P_u$ for each nonempty $u \subset v$.

Then there is a unique probability measure P on $\mathcal{F} = \prod_{t\in T} \mathcal{F}_t$ such that $\pi_v(P) = P_v$ for all v.

Proof. See [Ash et al., 2000, pp.18]. $\square$

For illustrations, see [Ash et al., 2000, Fig. 2.7.1, pp.117] or [Matthews, 2017, Fig 1.1, pp.4].

15 § 2.8: Weak convergence of measures

Weak convergence is the starting point for the study of the central limit theorem of probability.

If Ω is a metric space, the class of Borel sets of Ω, denoted by $\mathcal{B}(\Omega)$, is defined as the σ-field generated by the open sets of Ω.

Definition 15.0.1. Let $\mu, \mu_1, \mu_2, \dots$ be finite measures on the Borel sets of a metric space Ω. Then μ_n **converges weakly** to μ, written $\mu_n \xrightarrow{w} \mu$, if any of the conditions of Thm. 15.0.1 hold. $\triangle$

We begin with a lemma that will be useful. It tells us that if a set has a boundary of measure zero, then its measure equals the measure of both its interior and of its closure.

Lemma 15.0.1. *Let $A \subseteq X$ be a subset of set X. Let A° be the interior of A, $\overline{A}$ the closure of A, and ∂A the boundary of A. If $\mu(\partial A) = 0$, then $\mu(A) = \mu(\overline{A}) = \mu(A^\circ)$.*

Proof. We want to show that

$$\mu(A) \overset{\textit{(1)}}{=} \mu(\overline{A}) \overset{\textit{(2)}}{=} \mu(A^\circ).$$

(1). We can write $\overline{A} = A \cup \partial A$ (not necessarily disjointly). So

$$\mu(\overline{A}) \overset{\textit{subadditivity}}{\leq} \mu(A) + \cancelto{0, \textit{(by hypoth.)}}{\mu(\partial A)} = \mu(A)$$

But also

$$\mu(\overline{A}) \overset{\textit{monotonicity}}{\geq} \mu(A).$$

So $\mu(\overline{A}) = \mu(A)$.

(2). Can be argued similarly as in (1), noting that $A = A^\circ \cup \partial A$.

□

Now we provide the various characterizations of weak convergence.[70]

Theorem 15.0.1. (The Portmanteau Theorem.) *Let $\mu, \mu_1, \mu_2, \ldots$ be finite measures on the Borel sets of a metric space Ω. The following conditions are equivalent:*

(a) $\int_\Omega f\ d\mu_n \ \to \int_\Omega f\ d\mu$ *for all bounded continuous $f : \Omega \to \mathbb{R}$.*

(b) $\liminf_{n\to\infty} \int_\Omega f\ d\mu_n \ \geq \int_\Omega f\ d\mu$ *for all bounded lower semicontinuous $f : \Omega \to \mathbb{R}$.*

(b') $\limsup_{n\to\infty} \int_\Omega f\ d\mu_n \ \leq \int_\Omega f\ d\mu$ *for all bounded upper semicontinuous $f : \Omega \to \mathbb{R}$.*

(c) $\int_\Omega f\ d\mu_n \ \to \int_\Omega f\ d\mu$ *for all bounded $f : (\Omega, \mathcal{B}(\Omega)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that f is continuous a.e. $[\mu]$.*

(d) $\liminf_{n\to\infty} \mu_n(A) \geq \mu(A)$ *for every open set $A \subset \Omega$, and $\mu_n(\Omega) \to \mu(\Omega)$.*

(d') $\limsup_{n\to\infty} \mu_n(A) \leq \mu(A)$ *for every closed set $A \subset \Omega$, and $\mu_n(\Omega) \to \mu(\Omega)$.*

(e) $\mu_n(A) \to \mu(A)$ *for every $A \in \mathcal{B}(\Omega)$ such that $\mu(\partial A) = 0$ (∂A denotes the boundary of A).*

Proof. We prove the implications in a cycle.

$\boxed{(a) \implies (b).}$ Let f be a bounded LSC function. By [Ash et al., 2000, Theorem A2.6], there is a sequence of continuous functions $\{g_k\}$ such that

$$\underbrace{g_k}_{\text{continuous}} \uparrow \underbrace{f}_{\text{LSC}}. \qquad (1)$$

Now since each $g_k \leq f$, we have for fixed k that

$$\begin{aligned} \int g_k\ d\mu_n \ &\leq \int f\ d\mu_n \qquad \text{for any measure } \mu_n && \text{monotonicity of integral} \\ \liminf_{n\to\infty} \int g_k\ d\mu_n \ &\leq \liminf_{n\to\infty} \int f\ d\mu_n && \text{order preserved by asymptotics} \\ \int g_k\ d\mu \ &\leq \liminf_{n\to\infty} \int f\ d\mu_n && \text{by (a), } \liminf_{n\to\infty} \int g_k\ d\mu_n = \lim_{n\to\infty} \int g_k\ d\mu_n = \int g_k\ d\mu \end{aligned}$$

Taking the limit as $k \to \infty$ in the equation immediately above, we obtain

$$\liminf_{n\to\infty} \int f\ d\mu_n \ \geq \lim_{k\to\infty} \int g_k\ d\mu \overset{\text{extended MCT}}{=} \int \lim_{k\to\infty} g_k\ d\mu \overset{(1)}{=} \int f\ d\mu$$

where the criteria for the extended MCT are met since μ is finite.[71]

[70] For a bit of intuition, see `https://math.stackexchange.com/questions/3497341/example-of-weak-convergence`.

[71] [Ash et al., 2000] does not use MCT, but rather takes a supremum along with an additional boundedness argument. I'm not sure why he does it this way; it seems more obscure and convoluted to me than just using MCT.

$\boxed{(b) \iff (b').}$ Let f be a bounded LSC function. Then we have

$$\liminf_{n\to\infty} \int f \, d\mu_n \geq \int f \, d\mu \qquad \text{Characterization (b)}$$

$$-\liminf_{n\to\infty} \int f \, d\mu_n \leq -\int f \, d\mu \qquad \text{Multiply both sides by -1}$$

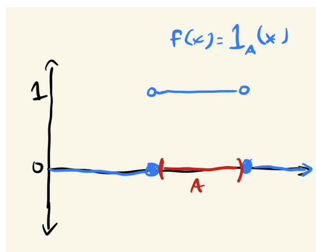
$$\limsup_{n\to\infty} \int -f \, d\mu_n \leq \int -f \, d\mu \qquad \text{Prop. A.4.2, constant multiple prop of } \int$$

Now use that f is LSC if and only if $-f$ is USC (Prop. D.8.2); the negation doesn't affect boundedness.

$\boxed{(b) \implies (c).}$ See [Ash et al., 2000].

$\boxed{(c) \implies (d).}$ Clearly (c) implies (a), which in turn implies (b). If A is open, then $f = 1_A$ is LSC (Def. D.8.1), since

$$\{f^{-1}(y > a)\} = \begin{cases} \emptyset, & \text{if } a > 1 \\ A, & \text{if } 0 < a \leq 1 \\ \Omega, & \text{if } a \leq 0 \end{cases},$$



all of which are open. So by (b) $\liminf_{n\to\infty} \mu_n(A) \geq \mu(A)$. Now $1_\Omega \equiv 1$, so $\mu_n(\Omega) \to \mu(\Omega)$ by (c).

$\boxed{(d) \iff (d').}$ We argue that $(d) \implies (d')$, but the argument goes both ways. Let $B \subseteq \Omega$ be closed. Then $\Omega - B := A$ is open, and any finite measure ν, we have

$$\nu(A) = \nu(\Omega) - \nu(B) \qquad (1)$$

(This equation holds by the piece-and-difference decomposition of Thm. 3.3.1, with the finiteness of ν allowing subtraction.)

So

$$\begin{aligned}
\liminf_{n\to\infty} \mu_n(A) &= \liminf_{n\to\infty} [\mu_n(\Omega) - \mu_n(B)] && \text{by } (1) \\
\mu(A) &\leq \liminf_{n\to\infty} [\mu_n(\Omega) - \mu_n(B)] && \text{by (d)} \\
&= \liminf_{n\to\infty} \mu_n(\Omega) - \limsup_{n\to\infty} \mu_n(B) && \text{Prop. A.2.3 and def. liminf} \\
&= \mu(\Omega) - \limsup_{n\to\infty} \mu_n(B) && \text{Since } \mu_n(\Omega) \to \mu(\Omega) \\
\implies \mu(\Omega) - \mu(B) &\leq \mu(\Omega) - \limsup_{n\to\infty} \mu_n(B) && \text{by } (1) \\
\implies \mu(B) &\geq \limsup_{n\to\infty} \mu_n(B) && \text{Subtract off } \mu(\Omega) \text{ (ok since } \mu \text{ finite); multiply by -1}
\end{aligned}$$

$\boxed{(d) \implies (e).}$ We will show that if (d) hold and $\mu(\partial A) = 0$, then

$$\mu(A) \overset{(1)}{\leq} \liminf_{n\to\infty} \mu_n(A) \overset{\text{Prop. A.4.3}}{\leq} \limsup_{n\to\infty} \mu_n(A) \overset{(2)}{\leq} \mu(A).$$

(1.) Since $A^\circ \subseteq A$, we have

$$\mu_n(A^\circ) \overset{\textit{monotonicity}}{\leq} \mu_n(A) \quad \forall n \qquad \oplus$$

So

$$\liminf_n \mu_n(A) \overset{\oplus}{\geq} \liminf_n \mu_n(A^\circ) \overset{(d)}{\geq} \mu(A^\circ) \overset{\textit{hypothesis, Lemma 15.0.1}}{=} \mu(A)$$

(2.) Since $A \subseteq \overline{A}$, we have

$$\mu_n(A) \overset{\textit{monotonicity}}{\leq} \mu_n(\overline{A}) \quad \forall n \qquad \oplus$$

So

$$\limsup_n \mu_n(A) \overset{\oplus}{\leq} \liminf_n \mu_n(\overline{A}) \overset{(d')}{\leq} \mu(\overline{A}) \overset{\textit{hypothesis, Lemma 15.0.1}}{=} \mu(A)$$

$\boxed{(e) \implies (a).}$ See [Ash et al., 2000].

$\square$

Example-for-data-scientists 15.0.1. For an application of the Portmaneau Theorem, see [Cheng et al., 2020]. $\triangle$

Theorem 15.0.2. (Characterizing weak convergence via distribution functions.) *Let $\mu, \mu_1, \mu_2, \ldots$ be finite measures on $\mathcal{B}(\mathbb{R})$, with corresponding distribution functions $F, F_1, F_2, \ldots$.*[72] *The following are equivalent:*

a) $\mu_n \overset{w}{\to} \mu$

b) $\mu_n(a,b] \to \mu(a,b]$ *at all continuity points a, b of F, where $\mu(a,b] = F(b) - F(a), F(\infty) = \lim_{x\to\infty} F(x), F(-\infty) = \lim_{x\to-\infty} F(x)$.*

If all distribution functions are 0 at $-\infty$, condition (b) is equivalent to the statement that $F_n(x) \to F(x)$ at all points $x \in \mathbb{R}$ at which F is continuous, and $F_n(\infty) \to F(\infty)$.

Remark 15.0.1. Here, we sketch the proof of Thm 15.0.2. For a full proof, see [Ash et al., 2000].

- $\boxed{(a) \implies (b)}$. This follows almost immediately from the characterization given by Thm. 15.0.1 (e). First note that by Example D.9.1, the boundary of $(a,b]$ is given by $\partial(a,b] = \{a\} \cup \{b\}$. Next, we show that $\mu(\partial(a,b]) = 0$ if a, b are continuity points of F. Note that $\mu(\{a\} \cup \{b\}) \overset{\textit{countable additivity}}{=} \mu(\{a\}) + \mu(\{b\})$, so it suffices to show that $\mu(\{c\}) = 0$ at any continuity point c of F.

$$\begin{aligned} \mu(\{c\}) &= F(c) - \lim_{x \uparrow c} F(x) && \text{Properties of Lebesgue-Stieltjes measures; see Prop 5.3.1 (e)} \\ &= F(c) - F(c) && c \text{ is a point of continuity of } F \\ &= 0. \end{aligned}$$

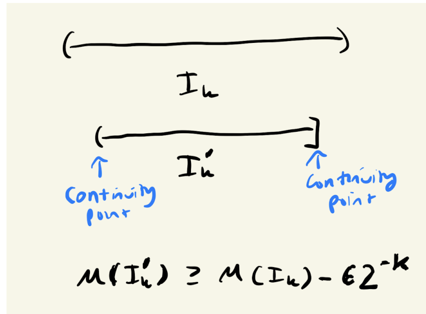
- $\boxed{(b) \implies (a)}$. Weak convergence is proved using the characterization given by Thm. 15.0.1 (d). We note:

  1.) Any open set $A \subset \mathbb{R}$ can be written as a countable union of disjoint open intervals, $A = \cup_{k=1}^\infty I_k$.

[72] Note that any finite measure is a Lebesgue-Stieltjes measure, and hence has a corresponding distribution function.

2.) F can only have at most countably many discontinuities (see [Ash et al., 2000, Sec. 1.5, HW 9]).

2i.) Thus, we can approximate any open interval $I_k = (a_k, b_k)$ arbitrarily well by a right semi-closed subinterval $I'_k = (a'_k, b'_k]$ whose endpoints are continuity points of F.



2ii.) And so $\mu_n(I'_k) \to \mu(I'_k)$ by (b).

3.) Now we prove weak convergence (characterization Thm. 15.0.1 (d)) by

$$\begin{aligned}
\liminf_{n\to\infty} \mu_n(A) &\overset{\text{(1.)}}{=} \liminf_{n\to\infty} \sum_{k=1}^{\infty} \mu_n(I_k) \\
&\geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} \mu_n(I_k) && \text{Fatou's Lemma for series; Rk. 7.5.2} \\
&\overset{\text{(2i.)}}{\approx} \sum_{k=1}^{\infty} \liminf_{n\to\infty} \mu_n(I'_k) \\
&\overset{\text{(2ii.)}}{=} \sum_{k=1}^{\infty} \mu(I'_k) \\
&\overset{\text{(2i.)}}{\approx} \sum_{k=1}^{\infty} \mu(I_k) \\
&\overset{\text{(1.)}}{=} \mu(A)
\end{aligned}$$

△

16 § 4 Basic Concepts of Probability

16.1 Introduction

Definitions 16.1.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. We refer to Ω as the **sample space** and sets in $\mathcal{F}$ as **events**. △

16.2 § 4.6 Random variables

Definition 16.2.1. A **random variable** X on a probability space $(\Omega, \mathcal{F}, P)$ is a Borel measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.[73] X is said to be an **extended random variable** if it is a Borel measurable function from $(\Omega, \mathcal{F})$ to $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. △

Notation 16.2.1. To emphasize the σ-field $\mathcal{F}$ with which the random variable is measurable, some authors (e.g. [Durrett, 2010]) write $X \in \mathcal{F}$. (This notation will be useful when we get to conditional expectation.) △

[73] Ash expresses random variables as $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and he interprets this notation as implying that the measurability condition $X^{-1}(B) \in \mathcal{F}, \quad \forall\, B \in \mathcal{B}(\mathbb{R})$ is satisfied (e.g. see [Ash et al., 2000, pp.176, top paragraph before Sec 4.7]). However, I find this notation to be unclear; to me it suggests that the *direct* images of X are in $\mathcal{B}(\mathbb{R})$ for each set in $\mathcal{F}$.

Example 16.2.1. If $(\Omega, \mathcal{F}, P)$ corresponds to a sequence of 4 Bernoulli trials [Ash et al., 2000, Sec. 4.4] and X is the number of successes, then $X(1011) = 3, X(0100) = 1$, and so on. $\triangle$

Definition 16.2.2. If X is a random variable on $(\Omega, \mathcal{F}, P)$, the **probability measure induced** by X is the probability measure P_X on $\mathcal{B}(\mathbb{R})$ given by

$$P_X(B) := P\{\omega : X(\omega) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}).$$

$\triangle$

Remark 16.2.1. (*Random variables induce pushforward measures.*) The probability measure induced by a random variable, as described in Def. 16.2.2, is also known as a *pushforward measure* (or sometimes *image measure*). Other ways to denote P_X which tend to correspond the "pushforward" terminology are $X_*(P)$, $X\#P$, or $P \circ X^{-1}$. $\triangle$

The numbers $P_X(B),\ B \in \mathcal{B}(\mathbb{R})$ completely characterize the random variable X in the sense that they provide the probabilities of all events involving X. It is useful to know that this information may be captured by a single function from $\mathbb{R}$ to $\mathbb{R}$.

Definition 16.2.3. The **distribution function** of a random variable X is the function F from $\mathbb{R}$ to $[0, 1]$ given by

$$F_X(x) = P\{\omega : X(\omega) \leq x\}, \quad x \in \mathbb{R}.$$

$\triangle$

Remark 16.2.2. (*Correspondence between distribution functions of random variables and probability measures induced by random variables.*) There is a correspondence between distribution functions of random variables F_X and probability measures induced by random variables P_X. In essence, this holds due to the results of Sec. 5, where we found a correspondence between distribution functions F (not necessarily of random variables) and Lebesgue-Stieltjes measures, since P_X is a Lebesgue-Stieltjes measure.

- *Let X be a random variable with induced probability measure P_X. Let F_X be a distribution of a random variable (see Def. 16.2.3). Then F_X is one of the distribution functions F corresponding to the Lebesgue-Stieltjes measure P_X, and we choose the one where $F(\infty) = 1$ and $F(-\infty) = 0$.*

  We have

$$\begin{aligned} F_X(b) - F_X(a) &\overset{\text{def}}{=} P\{\omega : X(\omega) \leq b\} - P\{\omega : X(\omega) \leq a\} \\ &\overset{\text{piece-and-diff}}{=} P\{\omega : a < X(\omega) \leq b\} \\ &\overset{\text{def}}{=} P_X(a, b] \end{aligned}$$

  And so by Thm 5.1.1, F_X is one of the distribution functions F corresponding to the Lebesgue-Stieltjes measure P_X.

  Now note that $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$. We prove the first of these. For $x \in \mathbb{R}$, define $A_x := \{X \leq x\}$. Then $A_x \uparrow A := \{X \in \mathbb{R}\}$. So

$$1 \overset{\text{prob space}}{=} P(A) \overset{\text{cty of measure}}{=} \lim_{x\to\infty} P(A_x) \overset{\text{L-S}}{=} \lim_{x\to\infty} F_X(x).$$

  For this reason, out of all the distribution functions F corresponding to the Lebesgue-Stieltjes measure P_X, we choose the one where $F(\infty) = 1$ and $F(-\infty) = 0$.

- *If $F : \mathbb{R} \to [0, 1]$ is increasing and right-continuous, with $F(\infty) = 1$ and $F(-\infty) = 0$, then F is the distribution function of some random variable.*

  By Theorem 5.2.1, there is a Lebesgue-Stieltjes measure P corresponding to any distribution function F, and $P(\mathbb{R}) = F(\infty) - F(-\infty) \overset{\text{hypoth.}}{=} 1 - 0 = 1$, and so P must further be a probability measure. Now by taking X to be the identity function (see Remark 16.2.3), we have $P_X = P$, and we are done.

$\triangle$

Remark 16.2.3. (*Canonical method for constructing an underlying probability space.*)

A beautiful quote from [Ash et al., 2000, pp.174].

> Very often, the following statement is made: "Let X be a random variable with distribution function F," where F is a given function from $\mathbb{R}$ to $[0, 1]$ that is increasing and right continuous, with $F(\infty) = 1$ and $F(-\infty) = 0$. There is no reference to the underlying probability space $(\Omega, \mathcal{F}, P)$, and actually the nature of the underlying probability space is not important. The distribution function F determines the probability measure P_X, which in turn determines the probability of all events involving X. The only thing we have to check is that there be at least one $(\Omega, \mathcal{F}, P)$ on which a random variable X with distribution function F can be defined. In fact, we can always supply the probability space in a canonical way; take $\Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}(\mathbb{R})$, with P the Lebesgue-Stieltjes measure corresponding to F, and define $X(\omega) = \omega, \omega \in \Omega$, that is, X is the identity map. Since $P_X(B) = P\{\omega : X(\omega) \in B\} = P(B)$, X has induced probability measure P and therefore distribution function F.

$\triangle$

Definition 16.2.4. (*Absolutely continuous random variable.*) A random variable X is said to be **absolutely continuous** if there is a non-negative real-valued Borel measurable function f on $\mathbb{R}$ such that its distribution function F is given by

$$F(x) = \int_{-\infty}^{x} f(t)\, dt\ , \quad x \in \mathbb{R}.$$

We call f the *density* or *density function* of X. $\triangle$

Remark 16.2.4. (*Density functions integrate to 1.*) Because $F(x) \to 1$ as $x \to \infty$ (see Remark 16.2.2 for the brief argument), we have $\int_{-\infty}^{\infty} f(x)\, dx\ = 1$. $\triangle$

Remark 16.2.5. (*The probability induced by an absolutely continuous random variable can be determined by integrating its density function.*) If X is absolutely continuous with density f, it follows that

$$P_X(B) = \int_B f(x)\, dx\ , \quad \forall B \in \mathcal{B}(\mathbb{R}). \tag{16.2.1}$$

This is due to the uniqueness of the L-S measure. In particular, by Cor. 7.1.1, $\mu : \mu(B) = \int_B f(x)\ dx\ , B \in \mathcal{B}(\mathbb{R})$ is a measure. Moreover, this measure satisfies $\mu(a, b] = F(b) - F(a)$ by the Fundamental Theorem of Calculus. So by Thm. 5.2.1, μ is the unique Lebesgue-Stieltjes measure corresponding to F. Hence, $\mu = P_X$. $\triangle$

Remark 16.2.6. (*Justifying the name of absolutely continuous random variables.*) Recall from Sec. 10.1 that absolutely continuity was defined in terms of (signed) measures. So what justifies the name absolutely continuous *random variables*? Recall from Remark 10.1.1 that indefinite integrals give new measures that are absolutely continuous with respect to the original measure. But we saw in Remark 16.2.5 that the probability P_X induced by an absolutely continuous random variable is determined by integrating its density functions (i.e, it can be expressed as an indefinite integral with respect to Lebesgue measure). Thus, $P_X \ll$ Lebesgue measure.[74] $\triangle$

[74] For an alternate view, F_X is an absolutely continuous function, as mentioned by [Ash et al., 2000, p.175]. However, Ash defines absolutely continuous functions in [Ash et al., 2000, Sec. 2.3], and that material is not (yet) covered in these notes.

Remark 16.2.7. (*What a* distribution *is*) The **distribution** of a random variable is a generic term; to say that we know the distribution of a random variable X means that we know how to calculate $P\{X \in B\}$ for all Borel sets B [Ash et al., 2000, pp.185]. Based on the material in this section, we therefore need one of the following pieces of information:

a) The *induced probability measure*,

$$P_X(B) := P\{\omega : X(\omega) \in B\}$$

b) The *distribution function*,

$$F_X(x) := P\{\omega : X(\omega) \le x\}, x \in \mathbb{R}.$$

(The distribution function fully specifies the induced probability measure; see Remark 16.2.2.)

In particular, we have the following special cases:

1. If X is an absolutely continuous random variable, then it suffices to know its density f_X, since by Eq. (16.2.1),

$$P_X(B) = \int_B f(x)\, dx\ , \quad \forall B \in \mathcal{B}(\mathbb{R})$$

2. If X is a discrete random variable (taking on values in a countable set S), then it suffices to know its probability mass function p_X, since then[75]

$$P_X(B) = \sum_{x \in B} p_X(x), \quad \forall B \in \mathbb{P}(S).$$

$\triangle$

16.3 § 4.7 Random objects

We now generalize random variables to random objects, which includes random vectors as a special case.

Definition 16.3.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(\Omega', \mathcal{F}')$ be a measurable space. A **random object** (sometimes called a random element) is a $(\mathcal{F}, \mathcal{F}')$-measurable map from Ω to Ω'. $\triangle$

Examples 16.3.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space.

1. A *random variable* (Sec. 16.2) is a random object where $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

2. A n**-dimensional random vector** is a random object where $(\Omega', \mathcal{F}') = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. That is, an n-dimensional random-vector is a Borel measurable map from Ω to $\mathbb{R}^n$.

3. A **discrete-time stochastic process** is a random object where $(\Omega', \mathcal{F}') = (\mathbb{R}^{\mathbb{N}}, \prod_{n \in \mathbb{N}} \mathcal{B}(\mathbb{R}))$.[76]

$\triangle$

[75] Justify this explicitly. See e.g. [Ash et al., 2000, pp.174].

[76] For more information, see [Zitkovic, 2013b].

16.3.1 Random vectors

Here we focus specifically on random vectors.

Remark 16.3.1. A random vector X may be regarded as an n-tuple $(X_1, \ldots X_n)$ of random variables. This is because X is Borel measurable iff each X_i is Borel measurable (see Prop. 6.2.4). $\triangle$

Definition 16.3.2. (*Absolutely continuous random vector.*) A random vector $X = (X_1, \ldots, X_n)$ is said to be **absolutely continuous** if there is a non-negative real-valued Borel measurable function f on $\mathbb{R}^n$, called the *density* or *density function* of X, such that its distribution function F is given by

$$F(x) = \int_{-(\infty,x]} f(t)\, dt\ , \quad x \in \mathbb{R}^n.$$

$\triangle$

Remark 16.3.2. By Tonelli's Theorem, we can also express the distribution function of an absolutely continuous random vector via

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, dt_1\ \cdots\ dt_n$$

$\triangle$

Remark 16.3.3. (*The probability induced by an absolutely continuous random vector can be determined by integrating its density function.*) If X is absolutely continuous with density f, it follows that

$$P_X(B) = \int_B f(x)\, dx\ , \quad \forall B \in \mathcal{B}(\mathbb{R}^n). \tag{16.3.1}$$

This is due to the uniqueness of the L-S measure. The argument parallels that of Remark 16.2.5. In particular, by Cor. 7.1.1, $\mu : \mu(B) = \int_B f(x)\, dx\ , B \in \mathcal{B}(\mathbb{R}^n)$ is a measure. Moreover, this measure satisfies $\mu(a, b] = \int_{(a,b]} f(x)\ dx\ = \Delta_{(a,b]}F$ (See Sec. 5.6.6, Example 2). So by Thm. 5.2.1, μ is the unique Lebesgue-Stieltjes measure corresponding to F. Hence, $\mu = P_X$. $\triangle$

16.4 § 4.8 Independent random variables

Independence of random events is defined in [Ash et al., 2000, Sec. 4.3]. Here we define independence of random variables.

Definition 16.4.1. Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F}, P)$. Then $X_1, \ldots, X_n$ are said to be **independent** if for all sets $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, we have

$$P_X[X_1 \in B_1, \ldots, X_n \in B_n] = P_{X_1}[X_1 \in B_1] \cdots P_{X_n}[X_n \in B_n]. \tag{16.4.1}$$

$\triangle$

Remark 16.4.1. (*Independence as induced product measure: A collection of independent RV's has an induced probability measure which is the product of the individual induced probability measures.*) Note that independent random variables are defined according to a constraint that the (induced) probability measure must satisfy *only for measurable rectangles* (sets of the form $B = B_1 \times \cdots \times B_n$); see the left hand side of Eq. (16.4.1). However, this definition automatically determines the probability measure on *all* sets (see the classical product measure theorem as given in Cor. 13.3.1, or in [Ash et al., 2000, 2.6.8(b)] for a corresponding statement over n factors). Hence, independence of $X_1, \ldots, X_n$ may be expressed by saying that if $X = (X_1, \ldots, X_n)$ is a collection of independent random variables, then P_X is the product of the $P_{X_i}, i = 1, \ldots, n$. We write $P_X = \prod_{i=1}^n P_{X_i}$. $\triangle$

Proposition 16.4.1. ***(Functions of independent random variables are independent.)*** *Let $X_1, \dots, X_n$ be independent random variables on $(\Omega, \mathcal{F}, P)$. Let $g_i : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable functions for $i = 1, \dots, n$. Then $g_1(X_1), \dots, g_n(X_n)$ are independent random variables.*

Proof. Let $X = (X_1, \dots, X_n)$ be the n-dimensional random vector with independent components. Let $g \circ X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ be defined by $g \circ X = (g_1 \circ X_1, \dots g_n \circ X_n)$. Then for any $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$,

$$
\begin{aligned}
&P_{g(X)}[g_1(X_1) \in B_1, \dots g_n(X_n) \in B_n] \\
&\quad = P_\Omega[\omega \in \Omega : (g_1 \circ X_1)(\omega) \in B_1, \dots, (g_n \circ X_n)(\omega) \in B_n] && \text{Def. prob measure induced by } g(X) \\
&\quad = P_\Omega[\omega \in \Omega : X_1(\omega) \in g_1^{-1}(B_1), \dots, X_n(\omega) \in g_n^{-1}(B_n)] && \text{Def. inverse image} \\
&\quad = P_X[X_1 \in \underbrace{g_1^{-1}(B_1)}_{\in \mathcal{B}(\mathbb{R}) \text{ since } g \text{ measurable}}, \dots, X_n \in \underbrace{g_n^{-1}(B_n)}_{\in \mathcal{B}(\mathbb{R}) \text{ since } g \text{ measurable}}] && \text{Def. prob measure induced by } X \\
&\quad = P_{X_1}[X_1 \in g_1^{-1}(B_1)] \cdots P_{X_n}[X_n \in g_n^{-1}(B_n)] && \text{Independence of X} \\
&\quad = P_{g_1(X_1)}[g_1(X_1) \in B_1] \cdots P_{g_n(X_n)}[g_n(X_n) \in B_n] && \text{Def. inverse image, def. prob measure induced by g(X)}
\end{aligned}
$$

□

Remark 16.4.2. (*Functions of independent random* objects *are independent.*) The conclusion of Prop. 16.4.1 applies more generally even if each X_i is a *random object*. Recall from Def. 16.3.1 that random objects are relaxations of random variables that allow X_i to map to any measurable space (not just the reals equipped with the Borel sets). In this context, we have $(\Omega, \mathcal{F}, P) \xrightarrow{X_i} (\Omega_i, \mathcal{F}_i) \xrightarrow{g_i} (\Omega_i', \mathcal{F}_i')$. △

We now turn to independence of random variables and σ-fields, which will we important when we discuss conditional expectations with respect to a σ-field.

Definition 16.4.2. (*Independence of a random variable and a σ-field.*) We say that a random variable Y on $(\Omega, \mathcal{F})$ and a σ-field $\mathcal{G} \subset \mathcal{F}$ are independent (denoted $Y \perp\!\!\!\perp \mathcal{G}$) if

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad G \in \mathcal{G} : \tag{16.4.2}$$

$$P\big(\{Y \in B\} \cap G\big) = P\big(\{Y \in B\}\big)\ P(G) \tag{16.4.3}$$

△

Example 16.4.1. (*Any random variable is independent of the trivial σ-field.*) Let Y be a random variable on $(\Omega, \mathcal{F})$ and define the sub σ-field $\mathcal{G}$ as the trivial σ-field, that is $\mathcal{G} \triangleq \{\emptyset, \Omega\}$. Then Eq. (16.4.3) holds. For all $B \in \mathcal{B}(\mathbb{R})$,

$$
\begin{aligned}
\underbrace{P(\{Y \in B\} \cap \emptyset)}_{= P(\emptyset)} &= P(\{Y \in B\})\ P(\emptyset) && \text{since } P(\emptyset)\text{=0} \\
\underbrace{P(\{Y \in B\} \cap \Omega)}_{= P(\{Y \in B\})} &= P(\{Y \in B\})\ P(\Omega) && \text{since } P(\Omega)\text{=1}
\end{aligned}
$$

△

Proposition 16.4.2. *(A random variable is independent from an indicator for any set belonging to a sigma field that is independent from that random variable.) If $Y \perp\!\!\!\perp \mathcal{G}$, then $Y \perp\!\!\!\perp 1_G$ for all $G \in \mathcal{G}$.*

Proof. By Def. 16.4.1, we need to show that for all sets $B_1, B_2 \in \mathcal{B}(\mathbb{R})$, we have

$$P\{Y \in B_1, 1_G \in B_2\} = P\{Y \in B_1\} P\{1_G \in B_2\} \tag{$*$}$$

But

$$\{1_G \in B\} = \begin{cases} \Omega, & \text{if } 0 \in B, 1 \in B \\ G, & \text{if } 0 \notin B, 1 \in B \\ G^c, & \text{if } 0 \in B, 1 \notin B \\ \emptyset, & \text{if } 0 \notin B, 1 \notin B \end{cases}$$

So (*) holds by the definition of independence of a random variable and a σ-field (Def. 16.4.2). $\square$

16.5 § 4.10 Expectation

Definition 16.5.1. If X is a random variable on $(\Omega, \mathcal{F}, P)$, then the **expectation** of X is defined by

$$E(X) = \int_\Omega X \, dP, \tag{16.5.1}$$

provided the integral exists. $\triangle$

Theorem 16.5.1. ***The Law of the Unconscious Statistician (LOTUS).*** *Let X be a random variable on $(\Omega, \mathcal{F}, P)$. Let g be a Borel measurable function from $\mathbb{R}$ to $\mathbb{R}$. If $Y = g \circ X$, then*

$$E(Y) = \int_\mathbb{R} g \, dP_X \tag{16.5.2}$$

in the sense that if either of the two sides exist, so does the other, and the two sides are equal.

Proof. We use the basic technique of starting with indicators and proceeding to more complicated functions.

- $\boxed{\text{Step 1: Indicator functions.}}$ Let g be an indicator function, $g = 1_B, B \in \mathcal{B}(\mathbb{R})$. Then

$$\begin{aligned} E(Y) &= E(1_B \circ X) && \text{def. } Y \\ &= P(\{\omega : X(\omega) \in B\}) && \text{Integral of indicator} \\ &= P_X(B) && \text{def. } P_X \\ &= \int_R 1_B \, dP_X && \text{integral of indicator} \\ &= \int_R g \, dP_X && \text{def. } g \end{aligned}$$

  so if one side exists, so does the other, and the two sides are equal.

- $\boxed{\text{Step 2: Non-negative simple functions.}}$ Let g be a non-negative simple function, $g(x) = \sum_{j=1}^n x_j 1_{B_j}(x)$, with B_j disjoint sets in $\mathcal{B}(\mathbb{R})$.
  Then

$$g \circ X = \left(\sum_{j=1}^n x_j 1_{B_j}(x)\right) \circ X = \sum_{j=1}^n x_j \left(1_{B_j}(x) \circ X\right) \qquad (a)$$

So

$$
\begin{aligned}
E(g \circ X) &= \sum_{j=1}^{n} x_j E(1_{B_j}(x) \circ X) && \text{by (a), linearity of integral} \\
&= \sum_{j=1}^{n} x_j \int_{\mathbb{R}} 1_{B_j}(x)\, dP_X && \text{Step 1} \\
&= \int_{\mathbb{R}} \sum_{j=1}^{n} x_j 1_{B_j}(x)\, dP_X && \text{linearity of integral} \\
&= \int_{R} g\, dP_X && \text{def. } g
\end{aligned}
$$

where linearity of the integral holds due to non-negativity (see Remark 7.2.2). Thus, if one side exists, so does the other, and the two sides are equal.

- $\boxed{\text{Step 3: Non-negative Borel measurable functions.}}$ Let g be a non-negative Borel measurable function. Let $g_1, g_2, \ldots$ be non-negative simple functions with $g_n \uparrow g$. We have

$$\lim_{n\to\infty} E[g_n \circ X] \overset{MCT}{=} E[\lim_{n\to\infty}(g_n \circ X)] \overset{*}{=} E[g \circ X]$$

where in justifying the MCT and in (*) we used that if $g_n \uparrow g$, then $(g_n \circ f) \uparrow (g \circ f)$.[77]

At the same time, though, we have

$$\lim_{n\to\infty} E[g_n \circ X] \overset{Step\ 2}{=} \lim_{n\to\infty} \int_{\mathbb{R}} g_n\, dP_X \overset{MCT}{=} \int_{\mathbb{R}} g\, dP_X$$

Thus, $E[g \circ X] = \int_{\mathbb{R}} g\, dP_X$; if one side exists, so does the other, and the two sides are equal.

- $\boxed{\text{Step 4: Arbitrary Borel measurable functions.}}$ Let $g = g^+ - g^-$ be an arbitrary Borel measurable function and $Y = g \circ X$. Then

$$
\begin{aligned}
E[Y] &= E[Y^+] - E[Y^-] && \text{def. integral of arbitrary Borel measurable function} \\
&= E[g^+ \circ X] - E[g^- \circ X] && \text{positive, negative parts are determined by outer layer of composition} \\
&= \int_{\mathbb{R}} g^+\, dP_X - \int_{\mathbb{R}} g^-\, dP_X && \text{by Step 3} \\
&= \int_{\mathbb{R}} g\, dP_X && \text{def. integral of arbitrary Borel measurable function}
\end{aligned}
$$

If $E(Y)$ exists and, say, $E(Y^-)$ is finite, then $\int_{\mathbb{R}} g^-\, dP_X$ is finite, and hence $\int_{\mathbb{R}} g\, dP_X$ exists; by the same reasoning, the existence of $\int_{\mathbb{R}} g\, dP_X$ implies the existence of $E(Y)$.

□

Remark 16.5.1. (*Motivating LOTUS.*) In many situations, it may be inconvenient to compute $E(X)$ by integrating over Ω. LOTUS allows us to write the integral with respect to the induced probability measure P_X, therefore integrating over $\mathbb{R}$ (or $\mathbb{R}^n$ for random vectors). In other words, LOTUS converts an integral over $(\Omega, \mathcal{F}, P)$ into an integral over $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.[78] △

[77] Justification: $\lim_{n\to\infty}(g_n \circ f)(x) = \lim_{n\to\infty} g_n(f(x)) = g(f(x))$.

[78] For an example of where this conversion is useful, consider that independence of random variables is *defined* in terms of P_X, not P (see Def. 16.4.1). So we use LOTUS when proving that the expectation of the product of independent RV's is equal to the product of the expectations.

Remark 16.5.2. (*Writing an expectation as an integral against the -induced- probability measure.*) Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ be a random variable. Although the definition of expectation (Def. 16.5.1) expresses it as an integral against a probability measure P defined on (events within) outcome space Ω , we may express it instead as an integral against the induced probability measure P_X defined on (Borel subsets of) the reals. Consider the Law of the Unconscious Statistician (Theorem 16.5.1), take g to be the identity, and then one sees that

$$\int_\Omega X(\omega)\, dP(\omega) \;=\int_\mathbb{R} x\, dP_X(x)$$

△

Examples 16.5.1. (*Applying LOTUS to absolutely continuous or discrete random variables.*)

Let the hypotheses of LOTUS hold. Then:

a) $\boxed{\text{Absolutely continuous random variable.}}$ If X is a random variable with density f, then

$$\int g(x)\, dP_X \;=\int g(x)f(x)\, dx$$

(with integration over $\mathbb{R}$) in the sense that if either integral exists, then so does the other, and the two are equal.

b) $\boxed{\text{Discrete random variable.}}$ If X is a discrete random variable with probability mass function p, then

$$\int g(x)\, dP_X \;=\sum_x g(x)p(x)$$

△

Proof. The proof hinges on change of variables with Radon-Nikodym derivatives.[79] For any $B \in \mathcal{B}(\mathbb{R})$, we have

Absolutely continuous case	**Discrete case**	**Justification**
$P_X(B) = \int_B f(x)\, dx$ $i.e.\ \frac{dP_X}{d\mu} = f$ (μ is Lebesgue measure)	$P_X(B) = \sum_x p(x)$ $i.e.\ \frac{dP_X}{d\lambda} = p$ (λ is counting measure)	see Remark 16.2.7 and Radon-Nikodym derivs.
$\boxed{E[Y] = \int g\, dP_X}$		LOTUS
$\implies E[Y] = \int g\, f\, d\mu$ $= \int g(x)f(x)\, dx$	$\implies E[Y] = \int g\, p\, d\lambda$ $\overset{\text{Ex. 6.3.3}}{=} \sum_x g(x)p(x)$	calculus w/ Radon-Nikodym derivatives (Prop. 10.3.1)

□

Question 16.5.1. (*An alternate notation for LOTUS.*) [Ash et al., 2000, pp.188] writes Eq. (16.5.2) as

$$E(Y) = \int_\mathbb{R} g(x)dF(x)(= \int_\mathbb{R} g\, dP_X),$$

[79] To my surprise, [Ash et al., 2000, Sec. 4.10.3(c),(d)] justifies these claims by suggesting that one work from indicators to simple functions to nonnegative Borel measurable functions to arbitrary Borel measurable functions. Applying calculus with Radon-Nikodym derivatives seems much more expedient and insightful.

where F is the distribution function of X. He states that $\int_{\mathbb{R}} g(x)dF(x)$ *means* $\int_{\mathbb{R}} g\ dP_X$; it is *not* a Riemann-Stieltjes integral.[80] This raises two questions

1. Why introduce the new, indirect notation at all?

2. What does it mean to say that this is *not* a Riemann-Stieltjes integral?

△

Remark 16.5.3. (*Using LOTUS to see that the expected value equals the integral of the survival function.*) In an application of Fubini-Tonelli (see Example 13.5.3), we observed that the areas under the curve could be obtained via integrating either vertical or horizontal cross-sections.[81]

$$\int_{\Omega} g\ d\mu\ = \int_0^{\infty} \mu(\{x : g(x) > y\})\ dy \tag{16.5.3}$$

In Remark 13.5.1, we stated that as a special case of Eq. (16.5.3), we can obtain that the expected value of a random variable X equals the integral of its survival function; i.e.

$$\mathbb{E}[X] := \int_{\Omega} X(\omega)\ dP(\omega)\ = \int_0^{\infty} P(\omega : X(\omega) > y)dy \tag{16.5.4}$$

We can perhaps see this most clearly using LOTUS, which allows us to rewrite Eq. (16.5.4) in terms of the induced probability measure P_X:

$$\mathbb{E}[X] := \int_{\mathbb{R}} x\ dP_X(x)\ = \int_0^{\infty} P_X(x : x > y)dy \tag{16.5.5}$$

In particular, the integrals are both over the real line.

Now Eq. (16.5.5) is obtained from Eq. (16.5.3) by simply taking the measure space to be $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ and the function of interest to be the identity, $g(x) = x$.

So Eq. (16.5.4) is just a special case of the fact (from Remark 13.5.1) that areas can be obtained either by integrating vertical or horizontal cross sections. For an illustration, see Fig. 27.

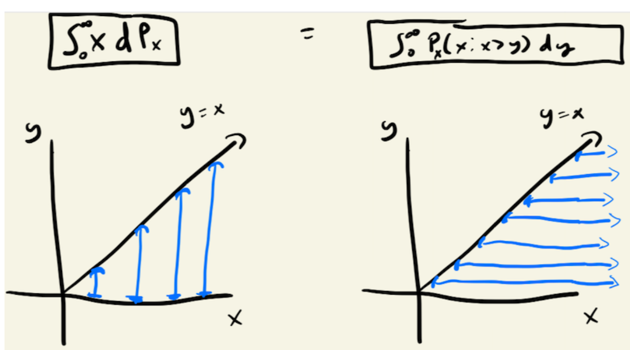


Figure 27: *The expected value of a non-negative random variable equals the integral of its survival function.* Note from the left that the expected value can be seen as the area under $y = x$ if the x-axis is measured with P_X.

△

[80] Which, interestingly, contradicts the current statement of LOTUS on Wikipedia, as of 08/05/2022.

[81] For intuition, see the figure in Example 13.5.3.

Remark 16.5.4. (*LOTUS also applies for random vectors and random objects*) LOTUS (Theorem 16.5.1) still holds if

a) X is a random vector, i.e. $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

b) X is a random object, i.e $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$.

The proof is exactly as in Theorem 16.5.1, with $\mathbb{R}^n$ replacing $\mathbb{R}$ for (a) and $(\Omega', \mathcal{F}')$ replacing $\mathbb{R}, \mathcal{B}(\mathbb{R})$ for (b). $\triangle$

16.5.1 Moments

We now investigate moments, which are special cases of expectations from which we can build means, variances, and so forth.

Definition 16.5.2. Let X be a random variable. If $k > 0$, then

- $E(X^k)$ is called the k-th **moment** of X
- $E(|X|^k)$ is called the k-th **absolute moment** of X

$\triangle$

Proposition 16.5.1. ***A moment is finite iff the absolute moment is finite.*** *Let X be a random variable and $k > 0$. Then $E(X^k)$ is finite iff $E(|X|^k)$ is finite.*

Proof. This is a special case of Cor. 7.2.1 (a). $\square$

Proposition 16.5.2. ***Finiteness of moments implies finiteness of lower moments.*** *If $k > 0$ and $E(X^k)$ is finite, then $E(X^j)$ is finite for $0 < j < k$.*

Proof.

$$\begin{aligned}
E[|X|^j] &= \int_\Omega |X|^j \, dP && \text{def. expectation} \\
&= \int_{\{|X|^j<1\}} |X|^j \, dP + \int_{\{|X|^j \geq 1\}} |X|^j \, dP && \text{countable additivity of measure (recall Cor. 7.1.1) :-)} \\
&\leq \underbrace{P(|X|^j < 1)}_{\text{(A)}} + \underbrace{\int_\Omega |X|^k \, dP}_{\text{(B)}} && \text{see below} \\
&< \infty. && \text{since probabilities are less than one (left) and by assumption (right)}
\end{aligned}$$

Some Details: To justify that the inequality holds for term (A), note

$$\int_{\{|X|^j<1\}} |X|^j \, dP \overset{\textit{monotonicity of integral}}{\leq} \int_{\{|X|^j<1\}} 1 \, dP \overset{\textit{integral of indicator}}{=} P(|X|^j < 1)$$

To justify that the inequality holds for term (B), note

$$\int_{\{|X|^j \geq 1\}} |X|^j \, dP \overset{*}{\leq} \int_{\{|X|^j<1\}} |X|^k \, dP \overset{\textit{monotonicity}}{=} \int_\Omega |X|^k \, dP$$

where (*) holds since $|X|^k \geq |X|^j$ on the set $\{|X|^j \geq 1\}$, because $|X|^k = (|X|^j)^{k/j}$ and the exponent > 1 by assumption and the base ≥ 1 on the given set.

$\square$

Theorem 16.5.2. *Let $X_1, \ldots, X_n$ be independent random variables on $(\Omega, \mathcal{F}, P)$. If all the X_i are nonnegative or if $E(X_i)$ is finite for all i, then $E(X_1 \cdots X_n)$ exists and equals $E(X_1)E(X_2)\cdots E(X_n)$.*

Proof.

$$
\begin{aligned}
E(X_1 \cdots X_n) &= \int_{\mathbb{R}^n} x_1 \cdots x_n \, dP_X && \text{LOTUS for random vectors (Remark 16.5.4)} \\
&= \int_{\mathbb{R}^n} x_1 \cdots x_n \;\; d\big(\prod_{i=1}^{n} P_{X_i}(x_i)\big) && \text{Independence as induced product measure (Remark 16.4.1)} \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} x_1 \cdots x_n \;\; dP_{X_1}(x_1) \;\cdots\; dP_{X_n}(x_n) && \text{Fubini's theorem} \\
&= \int_{\mathbb{R}} x_1 \;\; dP_{X_1}(x_1) \;\cdots \int_{\mathbb{R}} x_n \;\; dP_{X_n}(x_n) && \text{Constant multiple property (Cor. 6.3.6 (a))} \\
&= E(X_1) \cdots E(X_n) && \text{LOTUS (Theorem 16.5.1)}
\end{aligned}
$$

To see why Fubini's theorem applies when $E(X_i)$ is finite for all i, recall first that $E(X_i)$ is finite iff $E(|X_i|)$ is finite (by Prop 16.5.1), in which case we can justify Fubini using the strategy of Remark 13.4.1. □

Definition 16.5.3. Let X be a random variable. Then

- The **mean** of X is $E[X]$.
- The **variance** of X is given by $\sigma_X^2 := E[(X - E[X])^2]$.
- The **standard deviation** of X is given by $\sigma_X := +\sqrt{\sigma_X^2}$.

△

Definition 16.5.4. Let X, Y be random variables with finite expectation, and assume $E[XY]$ is also finite. Then

- The **covariance** of X and Y is given by

$$\mathrm{Cov}(X, Y) := E[(X - E[X])(Y - E(Y)]$$

- If the variances σ_X^2, σ_Y^2 are finite, the **correlation coefficient** between X and Y are defined by

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

△

Proposition 16.5.3. *The correlation coefficient between X and Y satisfies*

$$-1 \leq \rho(X, Y) \leq 1$$

when it exists.

Proof. Apply the Cauchy-Schwarz inequality to $X - E(X)$ and $Y - E(Y)$.

Recall [Ash et al., 2000, pp.85] that the Cauchy-Schwarz inequality (in the case of real-valued functions) says that if $f, g \in L^2$ and $fg \in L^1$, then

$$\int_{\Omega} fg \; d\mu \;\leq \Big(\int_{\Omega} |f|^2 \; d\mu \; \int_{\Omega} |g|^2 \; d\mu \Big)^{1/2}.$$

When $\mu = P$ is a probability measure, and using the stated substitutions, this becomes

$$\mathrm{Cov}(X,Y) \le \Big(\mathrm{Var}(X)\mathrm{Var}(Y)\Big)^{1/2}$$

□

16.6 § 4.11 Infinite Sequences of Random Variables

[Ash et al., 2000, Thm. 4.11.1] gives the existence of a single probability space on which one can define an infinite sequence of *independent* random variables. Now we move to construct a probability space for a *Markov chain*.

Theorem 16.6.1. (Markov chains exist.) *Let S be a finite or countably infinite set. Let $A = (a_{ij})$ be a stochastic matrix, and let $\pi_i, i \in S$ be a probability distribution. Then there is a sequence of random variables $X_0, X_1, \ldots$, all defined on the same probability space and taking values in S such that*

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \pi_{i_0} a_{i_0 i_1} \cdots a_{i_{n-1} i_n} \tag{16.6.1}$$

for all $i_0, i_1, \ldots, i_n \in S$ and all $n = 0, 1, \ldots$.

Proof. We prove existence as an application of Thm. 14.1.1. Let $\Omega = S^\infty, \mathcal{F} = \mathcal{P}(S)^\infty$.[82]

Define a probability measure on $\mathcal{F}_0 = \mathcal{P}(S)$ by

$$P_0(B) = \sum_{j\in B} \pi_j \quad \text{for all} \quad B \in \mathcal{P}(S)$$

For each $n = 0, 1, 2, \ldots$ and each $(i_0, i_1, \ldots, i_n) \in S^{n+1}$, we define a probability measure $P_{n+1}(i_0, i_1, \ldots, i_n; B)$ on $\mathcal{F}_{n+1} = \mathcal{P}(S)$ by

$$P_{n+1}(i_0, i_1, \ldots, i_n; B) = \sum_{j\in B} a_{(i_n, j)} \quad \text{for all} \quad B \in \mathcal{P}(S).$$

Now for each $B \in \mathcal{P}(S)$, the function $P_{n+1}(\cdot, \cdot, \ldots, \cdot; B) : (S^{n+1}, \mathcal{P}(S)^{n+1}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is automatically measurable.[83]

Now for $\omega \in \Omega$ and $n = 0, 1, \ldots$, define the random variable X_n by

$$X_n(\omega) = \omega_n.$$

Then we can use Eq. (14.1.3) to evaluate $P_{0:n}$, the extended product measure on $\prod_{i=0}^n \mathcal{F}_i$, at the singleton set $B_n := (\omega : X_0(\omega) = i_0, \ldots, X_n(\omega) = i_n)$ to obtain

$$\begin{aligned} P_{0:n}(X_0(\omega) = i_0, \ldots, X_n(\omega) = i_n) &\overset{\text{Eq. (14.1.3)}}{=} \\ &\int_S P_0(\,d\omega_0\,) \int_S P_1(\omega_0;\, d\omega_1\,) \cdots \int_S P_{n-1}(\omega_0, \omega_1, \ldots, \omega_{n-2};\, d\omega_{n-1}\,) \\ &\cdot \underbrace{\int_S \mathbb{I}[(\omega_1, \ldots, \omega_n) = (i_1, \ldots i_n)]\; P_n(\omega_0, \omega_1, \ldots, \omega_{n-1};\, d\omega_n\,)}_{\text{(A)}} \end{aligned}$$

[82] We use $\mathcal{P}(S)$ to denote the power set of S.

[83] The product of power sets $\mathcal{P}(S)^{n+1}$ contains all subsets of S^{n+1} and is a sigma field (Example 3.1.2). Thus, the definition of measurable functions (Def. 6.2.2) is automatically satisfied.

Now by Example 6.3.3, we can reduce

$$\underbrace{\int_S \mathbb{I}[(\omega_1,\ldots,\omega_n)=(i_1,\ldots i_n)]\; P_n(\omega_0,\omega_1,\ldots,\omega_{n-1};\; d\omega_n\;)}_{\text{(A)}} = \mathbb{I}[(\omega_1,\ldots,\omega_n)=(i_1,\ldots i_n)]\; a_{\omega_{n-1},i_n}.$$

[Specifically, recall from Example 6.3.3 that if f is a function on a countably infinite set and μ is the counting measure on that set, then the Lebesgue integral is just a sum: $\int f\; d\mu \;=\sum_{n=1}^{\infty} f(n)$. Similarly, if P is a discrete probability measure with probability mass function p, then $\int f\; dP \;= \int fp\; d\mu \;= \sum_{n=1}^{\infty} f(n)p(n)$. See also [Ash et al., 2000, 4.10.3(d)].]

Continuing this process gives

$$P_{0:n}(X_0=i_0, X_1=i_1,\ldots,X_n=i_n)=\pi_{i_0}a_{i_0 i_1}\cdots a_{i_{n-1}i_n}$$

And by Thm. 14.1.1, there is a probability measure P on $\mathcal{F}$ such that for all n, P agrees with $P_{0:n}$ on all n-dimensional cylinders. ☐

Remark 16.6.1. The sequence of random variables (X_n) satisfying Eq. (16.6.1) in Thm. 16.6.1 is called a **Markov chain** corresponding to **transition matrix** A and **initial distribution** π. The set S is called the **state space** of the chain. The values a_{ij} are called the **transition probabilities**. △

Theorem 16.6.2. ***Weak law of large numbers.*** *Let $X_1, X_2,\ldots$ be independent random variables (not necessarily with the same distribution), each with finite mean and variance. Assume the variances to be uniformly bounded by $M<\infty$. Let $S_n = X_1+\ldots X_n$. Then $(S_n-\mathbb{E}[S_n])/n$ converges in probability to 0. That is, given $\epsilon>0$,*

$$P\left(\left|\frac{S_n-E[S_n]}{n}\right|\geq\epsilon\right)\to 0 \quad as \quad n\to\infty$$

Proof.

$$\begin{aligned}
P\left(\left|\frac{S_n-E[S_n]}{n}\right|\geq\epsilon\right) &\leq \frac{\mathbb{E}(S_n-\mathbb{E}S_n)^2}{n^2\epsilon^2} && \text{Chebyshev's inequality}\\
&= \frac{1}{n^2\epsilon^2}\mathrm{Var}S_n && \text{Def. variance}\\
&= \frac{1}{n^2\epsilon^2}\sum_{i=1}^{n}\mathrm{Var}S_i && \text{See 4.10.11 of [Ash et al., 2000]}\\
&\leq \frac{M}{n\epsilon^2} && \text{Hypothesis} \qquad \to 0
\end{aligned}$$

☐

Remark 16.6.2. (*Concentration of sample mean.*) In the special case where $\mathbb{E}[X_i]=m$ for all i, Thm. 16.6.2 says that $\frac{S_n}{n}\to m$ in probability. That is, for large n, the arithmetic average of n independent random variables each with finite expectation m (and with the variances uniformly bounded) is quite likely to be very close to m. △

Remark 16.6.3. (*Weak vs strong law of large numbers.*) See [Ash et al., 2000, pp.199] for a discussion of WLLN vs SLLN. Essentially, the WLLN is a statement about what happens for repeated experiments with fixed n. SLLN is a statement about what happens for a single experiment as n increases. △

17 § 5 Conditional Probability and Expectation

17.1 Introduction

Let us start with a simple motivating example.

Consider the measurable space $(\Omega, \mathcal{F}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ with measurable functions $X(x, y) = x$ and $Y(x, y) = y$. Instead of directly specifying the joint distribution of X and Y, we specify the marginal distribution P_X and a set of probability measures $\{P(x, \cdot)\}$ defined on $\mathcal{B}(\mathbb{R})$ for all $x \in \mathbb{R}$. As we will see, $\{P(x, \cdot)\}$ are the conditional probabilities.

By the extended product measure theorem, the probability of any event of the form $\{(X, Y) \in C\}$ is determined:

$$P(C) = \int_{-\infty}^{\infty} P(x, C(x))\, dP_X(x) \tag{17.1.1}$$

where $C(x)$ the section of C at x.

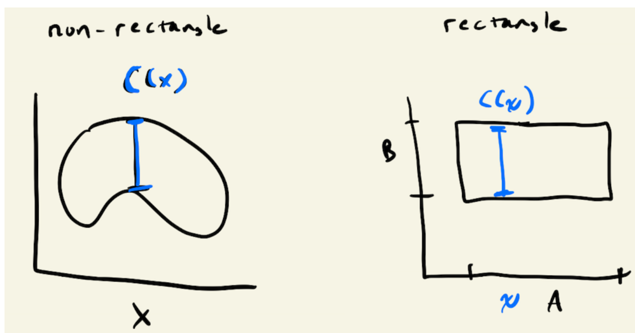


Figure 28: *Cross section of a measurable rectangle.* Let $(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \times \mathcal{B})$ be a measurable space. Let $C = A \times B$ be a measurable rectangle, where $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

In the special case where $C = A \times B$ is a measurable rectangle, the cross section $C(x)$ has a special form (see Fig. 28): Then the cross section of C at point $x \in \mathcal{X}$ has the special form

$$C(x) = \begin{cases} B, & x \in A \\ \emptyset, & x \notin A \end{cases}$$

This special form implies that $P(x, C(x)) = P(x, B)\mathbb{I}[x \in A]$. Thus, Eq. (17.1.1) becomes

$$P(A \times B) = \int_B P(x, B)\, dP_X(x) \tag{17.1.2}$$

Thus, we see that we can specify a joint distribution P on $\mathcal{F} = \mathcal{B}(\mathbb{R}^2)$ by specifying a marginal distribution P_X and a set of measures $\{P(x, \cdot), x \in \mathbb{R}\}$. Intuitively, this set of measures gives the conditional probabilities $P(Y \in B \mid X = x)$.

17.2 § 5.3.1-5.3.2 General Concept of Conditional Probability

In Sec. 17.1, we illustrated how the extended product measure theorem plays out in the context of random variables. Namely, we can construct a joint distribution from a marginal distribution P_X and a set of distributions parameterized by $X = x$. Intuitively, this set of distributions gives the conditional probabilities.

In this section, we work *the other way*. Here, given the marginal distribution P_X and the joint distribution, we will determine the conditional distribution.

Theorem 17.2.1. ***Conditional probability measure exists and is essentially unique.*** *Let $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a random object on $(\Omega, \mathcal{F}, P)$. Let B be a fixed set in $\mathcal{F}$. Then there is a real-valued Borel measurable function g on $(\Omega', \mathcal{F}')$ such that for each $A \in \mathcal{F}'$,*

$$P(\{X \in A\} \cap B) = \int_A g(x)\, dP_X(x) \tag{17.2.1}$$

Furthermore, if h is another such function, then $g = h$ a.e. $[P_X]$. We define the ***conditional probability measure*** *by $P(B \mid X = x) = g(x)$; it is essentially unique* for a given B.

Proof. Let $\lambda(A) := P(\{X \in A\} \cap B)$ for all $A \in \mathcal{F}'$.[84] [85]

Then λ is

- a finite measure on $\mathcal{F}'$. [The countable additivity of λ is inherited from P. Namely:

$$\lambda(\cup_{i=1}^{\infty} A_i) := P[\{X \in \cup_{i=1}^{\infty} A_i\} \cap B] = P[(\cup_{i=1}^{\infty}\{X \in A_i\}) \cap B] = P[\cup_{i=1}^{\infty}(\{X \in A_i\} \cap B)]$$
$$\overset{\text{ctble additivity of } P}{=} \sum_{i=1}^{\infty} P[\{X \in A_i\} \cap B] = \sum_{i=1}^{\infty} \lambda(A_i).]$$

- absolutely continuous with respect to P_X. (That is, $P_X(A) = 0 \implies \lambda(A) = 0$. This follows since $P(\{X \in A\} \cap B) = P\{\omega : X(\omega) \in A \text{ and } \omega \in B\} \overset{\text{monotonicity}}{\leq} P\{\omega : X(\omega) \in A\} := P_X(A)$.)

The result then follows from the Radon-Nikodym theorem. □

Notation 17.2.1. In Theorem 17.2.1, given a random object $X : (\Omega, \mathcal{F}, P) \to (\Omega', \mathcal{F}')$ and some set $A \in \mathcal{F}'$, we use the notation $\{X \in A\}$.

The interpretation is:

$$\{X \in A\} := X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

△

Remark 17.2.1. If we choose $A = \Omega'$ in Theorem 17.2.1, then Eq. (17.2.1) becomes

$$P(B) = \int_{\Omega'} g(x)\, dP_X(x), \tag{17.2.2}$$

which is the **law of total probability**. △

Corollary 17.2.1. *Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable on $(\Omega, \mathcal{F}, P)$. Let B be a fixed set in $\mathcal{F}$.*

a) If X is discrete and $P(X = x_i) > 0$ for some i, then

$$P(B \mid X = x_i) = \frac{P(B \cap \{X = x_i\})}{P\{X = x_i\}} \tag{17.2.3}$$

b) If X, Y has has joint density $f_{X,Y}$, then

$$P(B \mid X = x_i) = \int_{B(x)} \frac{f_{X,Y}(x, y)}{f_X(x)}\; dy\;, \tag{17.2.4}$$

where $B(x)$ is the cross section of B at x and $f_X(x) := \int_{\infty}^{\infty} f(x, y)\, dy$ is the density of X.

[84] Here is an interesting case of a measure (besides Lebesgue) which is not a probability measure but which we care about.

[85] To make the two probabililty measures in Eq. (17.2.1) more directly comparable, note that

$$P_X(A) := P\{\omega : X(\omega) \in A\}$$
$$P(\{X \in A\} \cap B) = P\{\omega : X(\omega) \in A \text{ and } \omega \in B\}$$

Remark 17.2.2. Some points of interest of Cor. 17.2.1:

- In (a), we see that if the event $\{X = x\}$ has positive probability, then our general definition of conditional probability in Eq. (17.2.1) agrees with the simpler definition of conditional probability [Ash et al., 2000, Sec.4.5] that $P(B \mid A) = P(A \cap B)/P(A)$ provided $P(A) > 0$.
- In (b), we see how our general definition of conditional probability can handle events $\{X = x\}$ of zero probability.

$\triangle$

Proof. a) We write $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{P}(S), P_X)$, where $S := \{x_1, x_2, \ldots\}$ is the set of countably many values that X can take on. Since X is discrete, we can write its induced probability measure in terms of a probability mass function p_X; namely, for any $B \in \mathcal{P}(S)$, we have $P_X(B) = \sum_{x \in B} p_X(x)$.

Now

$$P(B \cap \underbrace{\{X = x_i\}}_{:= A}) \overset{\textit{Eq. (17.2.1)}}{=} \int_{\{X = x_i\}} g(x)\ dP_X(x) \tag{17.2.5}$$

$$\overset{\textit{Ex. 16.5.1}}{=} g(x_i) p_X(x_i) \tag{17.2.6}$$

So

$$\frac{P(B \cap \{X = x_i\})}{P\{X = x_i\}} = \frac{g(x_i)\cancel{p_X(x_i)}}{\cancel{p_X(x_i)}}$$

and $g(x_i) = P(B \mid X = x_i)$ by definition.

b) See [Ash et al., 2000, pp.206-208].

$\square$

17.3 § 5.3.3-5.3.5 Conditional Expectation

Theorem 17.3.1. ***The conditional expectation exists and is essentially unique.***
Let

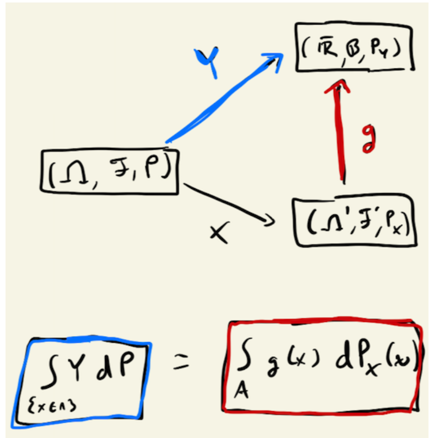
$$\begin{aligned} &(\Omega, \mathcal{F}, P) \ \textit{ be a probability space} \\ &X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}') \quad \textit{(a random object)} \\ &Y : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}) \quad \textit{(an extended random variable)} \\ \mathbb{E}[Y] = &\int_\Omega Y\ dP \quad \textit{exist} \end{aligned}$$

Then

$$\exists g : (\Omega', \mathcal{F}') \to (\overline{\mathbb{R}}, \mathcal{B}) \textit{ such that } \forall\, A \in \mathcal{F}',$$
$$\int_{\{X \in A\}} Y\ dP = \int_A g(x)\ dP_X(x)\ .$$

Furthermore, if h is another such function, then $g = h$ a.e. $[P_X]$. We define $\mathbb{E}[Y \mid X = x]$ as $g(x)$; it is essentially unique for a given Y.

Remark 17.3.1. Thm. 17.3.1 is summarized in the diagram below



By the interpretation of $g(x)$, we can write the conclusion as

$$\mathbb{E}[Y1_{\{X\in A\}}] = \int_A \mathbb{E}[Y1_{\{X\in A\}} \mid X = x] \; dP_X(x)$$

If $A = \Omega'$, this reduces to

$$\mathbb{E}[Y] = \int_{\Omega'} \mathbb{E}[Y \mid X = x] \; dP_X(x) \; .$$

$\triangle$

Proof. Let $\lambda(A) := \int_{\{X\in A\}} Y \; dP$. Then

- λ is a countably additive set function on $\mathcal{F}'$ by Thm. 7.1.1.
- λ is also absolutely continuous w.r.t P_X.

We need to show that $P_X = 0 \implies \lambda = 0$.

$$\begin{aligned}
&\text{Suppose } P_X(A) := P\{\omega : X(\omega) \in A\} = 0 \\
&\implies 1_{\{\omega : X(\omega)\in A\}} = 0 \quad a.e.[P] \\
&\implies Y \; 1_{\{\omega : X(\omega)\in A\}} = 0 \quad a.e.[P] \\
&\overset{Thm.\ 7.3.1(a)}{\implies} \int Y \; 1_{\{\omega : X(\omega)\in A\}} \; dP \; = 0
\end{aligned}$$

Thus, g exists (uniquely a.e. P_X) by the Radon-Nikodym Theorem. $\square$

Question 17.3.1. Can we better motivate the equality defining conditional expectation in Thm. 17.3.1?[86] For example, why do we not define $\mathbb{E}[Y \mid X = x]$ as the function $g(x)$ satisfying

$$\mathbb{E}[Y] = \int_{\Omega'} g(x) \; dP_X(x)$$

rather than

$$\mathbb{E}[Y1_{\{X\in A\}}] = \int_A g(x) \; dP_X(x) \qquad ?$$

$\triangle$

[86] I find the motivation provided by [Ash et al., 2000, pp.209, bottom] to be confusing.

Corollary 17.3.1. ***Conditional expectation includes conditional probability as a special case.*** *If X is a random object on $(\Omega, \mathcal{F}, P)$ and $B \in \mathcal{F}$, then*

$$\mathbb{E}[1_B \mid X = x] = P(B \mid X = x) \quad a.e.[P_X].$$

Proof. In Theorem 17.3.1, set $Y = 1_B$; the defining equation for conditional expectation becomes

$$P(\{X \in A\} \cap B) = \int_A \mathbb{E}[1_B \mid X = x] \, dP_X(x) \; .$$

The result now follows from Theorem 17.2.1. $\square$

Examples 17.3.1. (Conditional expectation when the random variables are discrete or absolutely continuous.) The conditional expectation $\mathbb{E}[Y \mid X = x]$ is defined (in Theorem 17.3.1) only *implicitly* (as an integrand). In these examples, we show how to get $\mathbb{E}[Y \mid X = x]$ on the LHS for special cases. Recall from Theorem 17.3.1 that we assume that $\mathbb{E}[Y]$ exists for all examples.

a) If X takes on countably many values $x_1, x_2, \ldots$ (assume WLOG that $P(X = x_i) > 0$ for all i), then

$$\mathbb{E}[Y \mid X = x] = \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} Y \, dP \tag{17.3.1}$$

If Y is also discrete, then

$$\mathbb{E}[Y \mid X = x] = \sum_j y_j P(Y = y_j \mid X = x_i) \tag{17.3.2}$$

b) Let X, Y be random variables with joint density f. Let $h = h(y|x)$ be the conditional density of Y given X.[87] Then

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y h(y \mid x) \, dy \tag{17.3.3}$$

$\triangle$

Proof. a) First we prove Eq. (17.3.1). Let X be a discrete random variable taking on values in $S = \{x_1, x_2, \ldots\}$. Let $\mathcal{P}(S)$ be the power set of S. For any $A \in \mathcal{P}(S)$, we have

$$\begin{aligned}
\int_{\{X \in A\}} Y \, dP &= \sum_{x_i \in A} \int_{\{X = x_i\}} Y \, dP && \text{Countable additivity of integral (Thm. 7.1.1)} \\
&= \sum_{x_i \in A} P(X = x_i) \underbrace{\frac{1}{P(X = x_i)} \int_{\{X = x_i\}} Y \, dP}_{:=\mathrm{g(x)}} && \text{Multiply and divide by 1} \\
&= \int_A g(x) \, dP_X(x) && \text{LOTUS expectations for discrete RVs (Ex. 16.5.1)}
\end{aligned}$$

Thus $g(x) := \frac{1}{P(X = x_i)} \int_{\{X = x_i\}} Y \; dP$ satisfies the definition of $\mathbb{E}[Y \mid X = x]$ given in Theorem 17.3.1.

[87]The conditional density is defined from the joint density by $h(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y) \, dy}$. See [Ash et al., 2000, Sec. 5.3.2].

Now we prove Eq. (17.3.2). Here, we additionally assume that Y is discrete. We have

$$\begin{aligned}
\mathbb{E}[Y \mid X = x] &= \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} Y \, dP && \text{Eq. (17.3.1)} \\
&= \frac{1}{P\{X = x_i\}} \sum_j y_j P(\{X = x_i\} \cap \{Y = y_j\}) && \text{Integral of simple function} \\
&= \sum_j y_j \frac{P(\{X = x_i\} \cap \{Y = y_j\})}{P\{X = x_i\}} && \text{Constant multiple} \\
&= \sum_j y_j P(Y = y_j \mid X = x_i) && \text{Def. conditional probability for sets}
\end{aligned}$$

b) For all $A \in \mathcal{F}'$ (the target sigma-field of X), we have[88]

$$\begin{aligned}
\int_{\{X \in A\}} Y \, dP &= \iint_{\{(x,y): x \in A\}} y f(x,y) \, dx \;\; dy && \text{LOTUS Expectations for abs. cts RVs (Ex. 16.5.1)} \\
&= \iint_{\{(x,y): x \in A\}} y f_1(x) \underbrace{\frac{f(x,y)}{f_1(x)}}_{:= h(y \mid x)} \, dx \;\; dy && \text{Mult. and divide by 1; here } f_1(x) := \int_{-\infty}^{\infty} f(x,y) \, dy \\
&= \int_{x \in A} f_1(x) \left[y \, h(y \mid x) \, dy \right] dx && \text{Constant multiple, Fubini} \\
&= \int_A \left[y \, h(y \mid x) \, dy \right] dP_X(x) && \text{LOTUS expectations for abs. cts RV's (Ex. 16.5.1)}
\end{aligned}$$

□

Remark 17.3.2. Some remarks on Example 17.3.1 (a):

1. We can motivate Eq. (17.3.1) as follows. First, recall from the section with conditional *probability*, that Eq. (17.2.3) gave the following result when the conditioning variable is discrete. For all $B \in \mathcal{F}$,
$$P(B \mid X = x_i) = \frac{P(B \cap \{X = x_i\})}{P\{X = x_i\}}.$$
We would therefore expect that the expectation of the indicator function 1_B would be given by
$$\mathbb{E}[1_B \mid X = x_i] = \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} 1_B \, dP \, .$$
Proceeding from indicators to simple functions to non-negative Borel measurable functions to arbitrary Borel measurable functions, we expect
$$\mathbb{E}[Y \mid X = x_i] = \frac{1}{P\{X = x_i\}} \int_{\{X = x_i\}} Y \, dP \, .$$

2. Let $B \in \mathcal{F}$ and assume $P(B) > 0$. If $\mathbb{E}[Y]$ exists, we can define the **conditional expectation of Y given B** as follows. Set $X = 1_B$, and set $E[Y \mid X = 1]$. Then by a special case of

[88] In using Fubini, I am following [Ash et al., 2000, pp.212]. But all we know is that $\mathbb{E}[Y]$ *exists*. We don't know that $\mathbb{E}|Y| < \infty$, which is I think what is required to use Fubini. How to reconcile? Note, perhaps relatedly, that we might want to clarify the statement and the application of LOTUS for abs continuous random *vectors*. It really gives us an integral with respect to the vector $\boldsymbol{z} := (x, y)$, whereas the way I've written it up seems to have already reduced the product Lebesgue measure into an iterated integral.

Example 17.3.1 (a), we obtain

$$\mathbb{E}[Y \mid B] = \frac{1}{P(B)} \int_B Y \; dP \; = \frac{\mathbb{E}[Y 1_B]}{P(B)} \tag{17.3.4}$$

We can interpret this quantity as the average value of Y over B [Durrett, 2010, pp.224].

△

17.4 § 5.4 Conditional expectation given a σ-field

We now define $\mathbb{E}[Y \mid \mathcal{G}]$ the conditional expectation of a random variable Y with respect to a σ-field $\mathcal{G}$. The quantity $\mathbb{E}[Y \mid \mathcal{G}]$ can be interpreted as the average value of $Y(\omega)$, given that we know, for each $G \in \mathcal{G}$, whether or not $\omega \in G$. The fundamental conceptual notion here is that the coarseness vs. fineness of a σ-field determines the amount of information we have about a random variable (see Remark 17.4.3). In particular, the σ-field $\mathcal{G} \subset \mathcal{F}$ providing information about some random variable Y on $(\Omega, \mathcal{F})$ is typically induced by some other random variable X on $(\Omega, \mathcal{F})$.

Theorem 17.4.1. ***(Conditional expectation given a σ-field exists and is essentially unique.)*** *Let*

$$\begin{aligned} Y: &\quad (\Omega, \mathcal{F}, P) \to \big(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\big) \\ \mathbb{E}[Y] &\quad \textit{exist.} \end{aligned}$$

Then for all σ-fields $\mathcal{G} \subseteq \mathcal{F}$,

$$\begin{aligned} \exists \quad \textit{function} \quad & \mathbb{E}[Y \mid \mathcal{G}] : (\Omega, \mathcal{G}) \to \big(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\big) : \\ & \int_G Y \; dP \; = \int_G \mathbb{E}[Y \mid \mathcal{G}] \; dP \qquad \forall G \in \mathcal{G}. \end{aligned}$$

Moreover, any two such functions coincide a.e. $[P]$.

Proof. Define $\lambda(G) = \int_G Y \; dP$ for all $G \in \mathcal{G}$. Then

- λ is a countably additive set function <u>on $\mathcal{G}$</u>. [89] (Indefinite integrals are countably additive set functions. See Thm 7.1.1.)
- $\lambda \ll P$. (Indefinite integrals are absolutely continuous. See Remark 10.1.1.)

The result follows by Radon-Nikodym. □

Remark 17.4.1. (*The <u>two</u> conditions for existence of conditional expectation with respect to a σ-field.*) One might wonder why Thm. 17.4.1 doesn't always yield that $\mathbb{E}[Y \mid \mathcal{G}] = Y$, since after all it is always true that

$$\int_G Y \; dP \; = \int_G Y \; dP \; .$$

Note that the theorem actually gives *two* conditions for a function h on $(\Omega, \mathcal{G})$ to be the conditional expectation given a σ-field $\mathcal{G}$:

1. [*"Measurability".*] h is measurable with respect to $\mathcal{G}$.

   (That is, we must have $h^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}(\overline{\mathbb{R}})$).

[89] Note that if we had defined λ on $\mathcal{F}$ instead of $\mathcal{G}(\subseteq \mathcal{F})$, the definition of $\mathbb{E}[Y \mid \mathcal{F}]$ just gives Y back; see the example below.

2. [*"Average matching".*] $\int_G Y \, dP = \int_G h \, dP$ for all $G \in \mathcal{G}$.

(We might call this this "integral condition", but we have decided to call this condition average matching, since $\frac{\int_G Y \, dP}{P(G)}$ is the average value of Y over G – see Eq. (17.3.4) – and the condition above implies that $\frac{\int_G Y \, dP}{P(G)} = \frac{\int_G h \, dP}{P(G)}$ for all $G \in \mathcal{G}$.)

Although condition (2) always holds for $h = Y$, condition (1) does not generally hold for a σ-field $\mathcal{G} \subset \mathcal{F}$. That is, Y will not in general be measurable w.r.t a σ-field $\mathcal{G} \subset \mathcal{F}$.[90]

Overall, we see that the two conditions of the definition present a tradeoff that must be balanced. A flexible choice of h (like $h = Y$) makes it easier to meet the average-matching condition, but harder to meet the measurability condition w.r.t $\mathcal{G}$. An inflexible choice of h (like $h = \mathbb{E}[Y]$) makes it easy to meet the average-matching condition, but harder to meet the integral condition. The conditional expectation is the sweet spot; it does *measurable* average matching. △

Example 17.4.1. (*Conditional expectation with respect to a simple σ-field on the unit square.*)

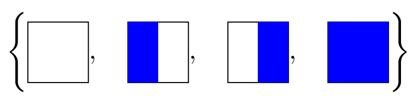
Consider the probability space given by

$$\begin{aligned} \Omega &= [0,1]^2 && \text{(the unit square)} \\ \mathcal{F} &= \mathcal{B}([0,1]^2) \\ P &= \texttt{Uniform} \end{aligned}$$

and the random variable Y on $(\Omega, \mathcal{F}, P)$ given by

3	1
5	3

Now consider the sub σ-field $\mathcal{G} \subset \mathcal{F}$ given by

$\mathcal{G} =$ 

We want to think about the conditional expectation $\mathbb{E}[Y \mid \mathcal{G}]$ given a σ-field. In particular, we want to use this example to illuminate the two conditions for the conditional expectation, as discussed in Remark 17.4.1.

We ask, can $\mathbb{E}[Y \mid \mathcal{G}]$ be ...

9	20

? ✗ This violates the *average matching* condition of conditional expectation. (Remark 17.4.1, condition 2)

3	1
5	3

? ✗ This violates the the *measurability* condition of conditional expectation.[91] (Remark 17.4.1, condition 1)

4	2

? ✓ This perfectly meets both conditions.

[90]For instance, let $(\Omega, \mathcal{F}) = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, and let Y be the identity function. Define $\mathcal{G} \subsetneq \mathcal{F}$ by $\mathcal{G} = \{\emptyset, A, A^c, \overline{\mathbb{R}}\}$ for $A \subsetneq \overline{\mathbb{R}}$. Let B be any Borel set that is not in $\mathcal{G}$. Then $Y^{-1}(B) = B \notin \mathcal{G}$ and so condition (1) fails. For a concrete illustration, see Example 17.4.1.

△

Remark 17.4.2. (*Seeing like a σ-field.*)[92] Example 17.4.1 illustrates two fundamental points about conditional expectations:

1. The true primitive for (the conditioning set of) a conditional expectation is a σ-field, *not* a random variable. No additional random variable is mentioned in Example 17.4.1! Indeed, as we will see in Sec. 17.4.1, the conditional expectation with respect to a random variable $E[Y \mid X]$ is a special case of the conditional expectation with respect to a σ-field $E[Y \mid \mathcal{G}]$. In case of $E[Y \mid X]$, the additional random variable X simply plays the intermediary role of *inducing* a particular kind of σ-field over the sample space Ω.

2. The notion of measurability is **not** needed simply to avoid odd pathological sets (like Lebesgue non-measurable sets); it is in fact fundamental to the notion of conditional expectation, even when dealing with very simple collections of sets.

△

Remark 17.4.3. (*Interpreting the conditional expectation with respect to a σ-field.*) Since an event corresponds to a question that has a yes or no answer, $\mathbb{E}[Y \mid \mathcal{G}]$ can be interpreted as the average value of $Y(\omega)$, given that we know, for each $G \in \mathcal{G}$, whether or not $\omega \in G$ [Ash et al., 2000, pp.218].

- If $\mathcal{G}$ is "coarse" (i.e. contains few sets and/or sets that are large), then we have weak information about Y. In particular, if $\mathcal{G}$ contains few sets, there are not many integral equalities to verify. If the sets $\{G \in \mathcal{G}\}$ are never small, then we don't need to have pinpoint accuracy to get the integrals to match.

- If $\mathcal{G}$ is "fine" (i.e. contains many sets and/or sets that are small), then we have strong information about Y.

△

Examples 17.4.1. Now we give some examples of conditional expectation with respect to a σ-field.

a) *Perfect information.*[93] $\mathbb{E}[Y \mid \mathcal{F}] = Y$. (That is, if we know Y, then our "best guess" is Y itself.)

b) *No information.*[94] If Y is independent of a σ-field $\mathcal{G} \subsetneq \mathcal{F}$ (see Def. 16.4.2), then $\mathbb{E}[Y \mid \mathcal{G}] = \mathbb{E}[Y]$. (That is, if we don't know anything about Y, then our best guess is the mean $\mathbb{E}[Y]$.)

△

Proof. For each example, we check that the two conditions of Remark 17.4.1 are satisfied.

a) Condition (1) is satisfied automatically, since Y is a random variable, and therefore measurable by definition.

   Condition (2) is satisfied since $\int_F Y \, dP = \int_F Y \, dP$ for all $F \in \mathcal{F}$.

[91] Condition 1 of conditional expectation - namely that $\mathbb{E}[Y \mid \mathcal{G}]$ must be $\mathcal{G}$-measurable - is violated for example because $\{1\} \in \mathcal{B}$, but the inverse image of $\{1\}$ is the top right square, which is not a member of $\mathcal{G}$.

[92] I plan to make a slide deck called "Seeing like a σ-field", a reference to James Scott's book *Seeing like a state*.

[93] See [Durrett, 2010, pp.223].

[94] See [Durrett, 2010, pp.223].

b) Condition (1) is satisfied. $h(\omega) = \mathbb{E}[Y] := c$ is constant in ω. So for all $B \in \mathcal{B}(\overline{\mathbb{R}})$,

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

Now $\{\emptyset, \Omega\}$ are contained in *any* σ-field, so $h(\omega) = \mathbb{E}[Y]$ is measurable with respect to *any* $\mathcal{G} \subset \mathcal{F}$.

Condition (2) is satisfied since $\forall G \in \mathcal{G}$,

$$\begin{aligned} \int_G Y\, dP &\overset{\textit{notation}}{=} \mathbb{E}[Y 1_G] \overset{\textit{independence}}{=} \mathbb{E}[Y]\mathbb{E}[1_G] \overset{\textit{notation}}{=} \mathbb{E}[Y] \int_\Omega 1_G\, dP \\ &\overset{\textit{constant multiple}}{=} \int_\Omega \mathbb{E}[Y] 1_G\, dP \\ &\overset{\textit{notation}}{=} \int_G \mathbb{E}[Y]\, dP \end{aligned}$$

In particular, the equality involving independence holds by Prop. 16.4.2 and Thm. 16.5.2.

□

Remark 17.4.4. For an example where “intermediate" information between the two extremes (of perfect or no information), see Sec. 17.4.1, where partial information about Y is provided by another random variable X. △

Remark 17.4.5. Here we provide further insight on conditional expectations with respect to σ-fields in the *perfect information* and *no information* cases. In particular, we show that the solutions for each doesn't work for the other case.

That is, if Y is a random variable on $(\Omega, \mathcal{F})$, we show that in general

1. $\mathbb{E}[Y \mid \mathcal{G}] \neq Y$ for σ-field $\mathcal{G} \subset \mathcal{F}$.
2. $\mathbb{E}[Y \mid \mathcal{F}] \neq \mathbb{E}[Y]$.

For (1), see Remark 17.4.1. In particular, condition (1) fails: Y will not in general be $\mathcal{G}$-measurable in general.

For (2), condition (2) fails in general. For instance, let $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We will show that, in general,

$$\int_B Y\, dP \neq \int_B \mathbb{E}[Y]\, dP \qquad (*)$$

First we analyze the RHS. Note that $\int_B \mathbb{E}[Y]\, dP = \mathbb{E}[Y]P(B)$ by the constant multiple property.

Now we analyze the LHS. Take $B = \{x\}, x \in \mathbb{R}$. By integrals for simple functions, we obtain $\int_B Y\, dP = Y(x)P\{x\}$.

Combining the LHS and RHS, (*) becomes

$$\begin{aligned} (*) &\implies Y(x)P\{x\} = \mathbb{E}[Y]P\{x\} \qquad \forall x \in \mathbb{R} \\ &\implies Y(x) = \mathbb{E}[Y] \qquad \forall x \in \mathbb{R}. \end{aligned}$$

which holds only if Y is constant.

△

17.4.1 Conditional expectation given a σ-field generated by another random variable

Given a random variable Y on $(\Omega, \mathcal{F})$, one σ-field $\mathcal{G} \subset \mathcal{F}$ of *particular* interest is the one generated by some *other* random variable X on $(\Omega, \mathcal{F})$.

Definition 17.4.1. Let $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a random object. Then the **σ-field generated by** X, denoted $\sigma(X)$, is given by:

$$\sigma(X) = X^{-1}(\mathcal{F}') = \{X^{-1}(B) : B \in \mathcal{F}'\} \tag{17.4.1}$$

$\triangle$

Remark 17.4.6. (*The σ-field generated by a random vector.*) In particular, if $X = (X_1, \ldots, X_n)$ is a random vector, $\sigma(X)$ consists of all sets $\{X^{-1}(B), B \in \mathcal{B}(\mathbb{R}^n)\}$. $\triangle$

Remark 17.4.7. (*Verifying that the σ-field generated by a random object is a σ-field.*) How do we know that Definition 17.4.1 is a valid definition – that is, how do we know that the collection of sets given by Eqn. 17.4.1 is indeed a σ-field? To answer, recall from Remark 6.2.2 that set operations and inverse images commute. $\triangle$

Example 17.4.2. (*The σ-field induced by a random vector that takes on finitely many values.*) Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_X)$ be a random vector that takes on finitely many values $\{s_1, \ldots, s_k\}$. Then

$$\underbrace{\sigma(X)}_{\sigma\text{-field generated by } X} = \sigma(\{\Omega_1, \ldots, \Omega_k\}) = \big\{\text{unions of } \{\Omega_1, \ldots, \Omega_k, \emptyset\}\big\}$$

where $\Omega_k \triangleq X^{-1}(\{s_k\})$ is an element of the partition $\Omega = \Omega_1 \uplus \cdots \uplus \Omega_k$.

In other words, the σ-field induced by a random variable that takes on finitely many values is given by unions of the empty set and inverse images of singletons.

To prove this, note that since X takes on finitely many values $\{s_1, \ldots, s_k\}$, the inverse image of any Borel set is determined solely by which subset of $\{s_1, \ldots, s_k\}$ it contains. $\triangle$

Proposition 17.4.1. ***(Conditional expectation with respect to a σ-field induced by another random variable.)*** *Let $Y : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ be an extended random variable and $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a random object. Define the σ-field $\mathcal{G} \subset \mathcal{F}$ by $\mathcal{G} = \sigma(X)$. Then*

$$h := \mathbb{E}[Y \mid \mathcal{G}] : (\Omega, \mathcal{G}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$$

is given by

$$h(\omega) = g(X(\omega))$$

where

$$g(x) = \mathbb{E}[Y \mid X = x]$$

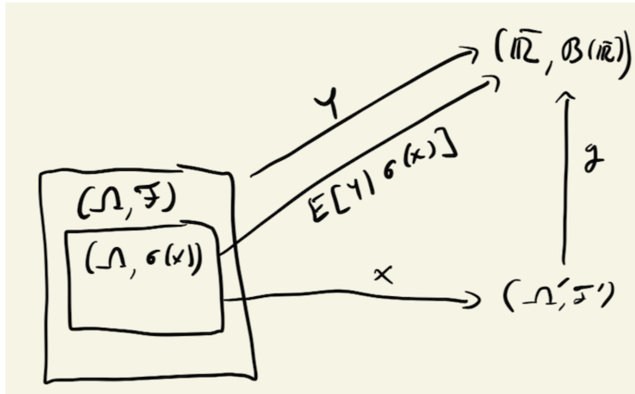


Figure 29: *The conditional expectation of Y with respect to $\sigma(X)$, the σ-field induced by the random variable X.* Note that $\sigma(X) \subset \mathcal{F}$. The two functions on the diagonals must integrate to the same value over sets in $\sigma(X)$, but the conditional expectation must also be measurable with respect to $\sigma(X)$.

Proof. We verify the two conditions for conditional expectation with respect to a σ-field.

1. X is measurable with respect to $\sigma(X)$ by definition, and g is measurable with respect to $\mathcal{F}'$ by Thm. 17.3.1. So $h = g \circ X$ is measurable w.r.t $\sigma(X)$ since the composition of two measurable functions is measurable (Prop. 6.2.3).

2. $$\begin{aligned} \int_{\{X \in A\}} h(\omega)\, dP(\omega) &= \int_{\{X \in A\}} g(X(\omega))\, dP(\omega) && \text{def. } h \\ &= \int_A g(x)\, dP_X(x) && \text{LOTUS} \\ &= \int_{\{X \in A\}} Y\, dP && \text{Assumption that } g \text{ satisfies Thm. 17.3.1.} \end{aligned}$$

□

Notation 17.4.1. (*The conditional expectation of a random variable given another random variable is defined in terms of the conditional expectation of a random variable given a σ-field.*) In Prop. 17.4.1, we take the conditional expectation with respect to a particular choice of σ-field $\mathcal{G} = \sigma(X)$. In this case, we notate

$$\mathbb{E}[Y \mid \mathcal{G}] = \mathbb{E}[Y \mid \sigma(X)] \quad \text{by} \quad \mathbb{E}[Y \mid X].$$

Note that $\mathbb{E}[Y \mid X]$ and $\mathbb{E}[Y \mid X = x]$ both give the conditional expectation, but over different domains :

$$\begin{gathered} \mathbb{E}[Y \mid \mathcal{G}] : (\Omega, \mathcal{G}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})) \\ \mathbb{E}[Y \mid X = x] : (\Omega', \mathcal{F}') \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})) \end{gathered}$$

△

Remark 17.4.8. (*Perfect information, revisited.*) Now that we know what it means to condition on a σ-field *generated by another random variable*, let us revisit the scenario of perfect information. In Example 17.4.1, we noted that if Y is a random variable on $(\Omega, \mathcal{F}, P)$, then

$$\mathbb{E}[Y \mid \mathcal{F}] = Y. \tag{17.4.2}$$

By the same argument, we can also write:

$$\mathbb{E}[Y \mid \sigma(Y)] = Y \tag{17.4.3}$$

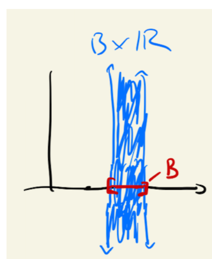
Eq. (17.4.3) is stronger than Eq. (17.4.2); all knowledge about Y is encoded in $\sigma(Y)$; any other set in $\mathcal{F}$ provides no value.[95]

Using Notation 17.4.1, Eq. (17.4.3) would be expressed in self-suggestive manner: $\mathbb{E}[Y \mid Y] = Y$. △

Example 17.4.3.

$$\begin{aligned}
&\text{Let } X, Y \text{ be random variables with joint density } f\\
&X : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to (\underbrace{\mathbb{R}}_{=\Omega'}, \underbrace{\mathcal{B}(\mathbb{R})}_{=\mathcal{F}'})\\
&Y : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))\\
&X(x,y) = x\\
&Y(x,y) = y\\
&P(B) = \iint_B f(x,y)\, dx \;\; dy \quad \forall B \in \mathcal{B}(\mathbb{R}^2)
\end{aligned}$$

Then the σ-field induced by X, i.e. $\sigma(X) = X^{-1}(\mathcal{F}')$, consists of the sets $\{B \times \mathbb{R},\ B \in \mathcal{B}(\mathbb{R})\}$.



By Prop. 17.4.1, the conditional expectation $h = \mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \sigma(X)]$ with respect to the sigma field $\sigma(X)$ is given by

$$h(\omega) = g(X(\omega)), \qquad \text{where } g(x) = \mathbb{E}[Y \mid X = x].^{96}$$

So $\mathbb{E}[Y \mid X]$ is constant on vertical strips of $\Omega = \mathbb{R}^2$, and the only sets $G \in \sigma(X)$ that we would want to interrogate to obtain knowledge about Y are sets of the form $\{x \times \mathbb{R}, x \in \mathbb{R}\}$.[97] △

17.5 § 5.5 Properties of and perspectives on conditional expectation

17.5.1 Overview

Here in Sec. 17.5 we provide some properties of conditional expectation. Let Y be an extended random variable on $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}[Y]$ assumed to exist, $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be a random

[95] Similarly, if some other σ-field $\mathcal{F}' \supset \mathcal{F}$, then $\mathbb{E}[Y \mid \mathcal{F}'] = Y$. That is, we already know everything there is to know about Y from $\mathcal{F}$, no knowledge can be added (or lost) by moving to the larger σ-field $\mathcal{F}'$. This fact can be proven immediately by verifying the two conditions in the definition of conditional expectation given a σ-field. Or, the argument can be made more compact by noting simply that if Y is a random variable on $(\Omega, \mathcal{F}, P)$, then it is also a random variable on $(\Omega, \mathcal{F}', P)$.

[96] Moreover, by Example 17.3.1, we have

$$g(x) = \mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y h(y \mid x)\, dy$$

where $h(y \mid x) = \frac{f(x,y)}{f_1(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\, dy}$ is the conditional density of y given x.

[97] But when is this *not* true? Is this just a true statement in general, or is there something about having a joint density that matters here?

object, and $\mathcal{G}$ a sub-σ-field of $\mathcal{F}$. Ash et al. [2000] presents properties of conditional expectation in pairs, so that we can work with either $\mathbb{E}[Y \mid \mathcal{G}]$ (which has $\mathbb{E}[Y \mid X]$ as a special case) or $\mathbb{E}[Y \mid X = x]$. To represent the distinction between $\mathbb{E}[Y \mid X]$ and $\mathbb{E}[Y \mid X = x]$, we give a small variant of Fig. 29 below in Fig. 30. We can consider $\mathbb{E}[Y \mid X = x]$ as our "best guess" of Y given that $X = x$, and $\mathbb{E}[Y \mid X]$ as our "best guess" of Y given that all we know about observed outcome ω is whether or not ω belongs to each set in $\sigma(X)$.

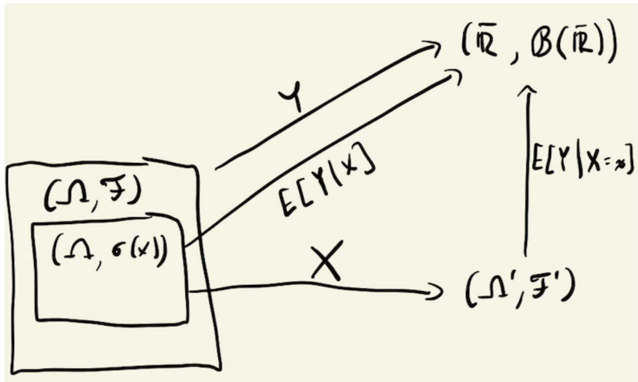


Figure 30: *Two types of conditional expectation given a random object.* Given a random variable Y and random object X defined on the same probability space, we can consider the conditional expectation to be either $\mathbb{E}[Y \mid X]$ or $\mathbb{E}[Y \mid X = x]$. These functions are defined on different spaces: $\mathbb{E}[Y \mid X]$ is defined on Ω and $\mathbb{E}[Y \mid X = x]$ is defined on Ω'. They share that they both must be measurable, and both must match integrals of Y: (1) Comparing the two diagonal lines, we need $\int_A \mathbb{E}[Y \mid X]\ dP \ = \int_A Y\ dP$ for all $A \in \sigma(X)$. (2) Comparing the vertical line to the top diagonal, we need $\int_B \mathbb{E}[Y \mid X = x]\ dP_X \ = \int_{X^{-1}(B)} Y\ dP$ for any $B \in \mathcal{F}'$.

Since the functions $\mathbb{E}[Y \mid \mathcal{G}]$ and $\mathbb{E}[Y \mid X = x]$ were defined implicitly as integrands, what justifies the expectation notation? Some thoughts:

1. Theorem 17.5.2 shows us how can view conditional expectations as averages when we apply it to "atoms" of a probability space.
2. Conditional expectations "behave" like expectations. In particular, we have
   - Monotonicity [Ash et al., 2000, Thm. 5.5.1b].
   - Linearity [Ash et al., 2000, Thm. 5.5.2].
   - The triangle inequality for integrals [Ash et al., 2000, Thm. 5.5.1c].
   - Monotone convergence theorem, and the fact that non-negative series can be integrated term-by-term [Ash et al., 2000, Thm. 5.5.3].
   - Dominated convergence theorem [Ash et al., 2000, Thm. 5.5.5].
   - Extended monotone convergence theorem, and Fatou's lemma [Ash et al., 2000, Thm. 5.5.6].

DISCUSS: Are there other reasons?

Although we noted above that conditional expectations "behave" like expectations, they also have properties that have no analogue for "ordinary" expectation:

- The law of iterated expectation [Ash et al., 2000, Thm. 5.5.4].
- Iterated expectation with nested σ-fields [Ash et al., 2000, Thm. 5.5.10].
- "Taking out what is known" [Ash et al., 2000, Thm. 5.5.11].

Notation. In the results to follow, $Y, Y_1, Y_2, \ldots$ are extended random variables on $(\Omega, \mathcal{F}, P)$ with all expectations assumed to exist; $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ is a random object; and $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$. The phrase "a.e." with no measure specified will always mean a.e. $[P]$. If $Z : (\Omega, \mathcal{G}) \to (\overline{\mathbb{R}}, \mathcal{B})$, we say that Z is $\mathcal{G}$-measurable, and if $g : (\Omega', \mathcal{F}') \to (\overline{\mathbb{R}}, \mathcal{B})$, we say that g is $\mathcal{F}'$-measurable.

17.5.2 Properties

Remark 17.5.1. Below we provide some properties of conditional expectation. These properties are stated in terms of $\mathbb{E}[Y|X]$. Ash et al. [2000, Sec. 5.5] instead gives a "pairing" structure to provide these properties in terms of both $\mathbb{E}[Y|X]$ and $\mathbb{E}[Y|X = x]$, which are different objects on different spaces. For a similar (partial) construction on our end, see Theorem E.0.1. However, I don't think that it is necessary to use the pairing structure; other authors don't, and there seems to be a generic way to obtain theorems about $E[Y|X = x]$ given a corresponding theorem about $E[Y \mid X]$; see [Ash et al., 2000, Sec. 5.5, Problem 3]. △

Theorem 17.5.1. *The following properties hold for conditional expectation given a σ-field*

(a) (Constant property.) If Y is a constant k a.e., then $\mathbb{E}[Y \mid \mathcal{G}] = k$ a.e

(b) (Monotonicity.) If $Y_1 \leq Y_2$ a.e., then $\mathbb{E}[Y_1 \mid \mathcal{G}] \leq \mathbb{E}[Y_2 \mid \mathcal{G}]$ a.e.

(c) (Linearity.) If $a, b \in \mathbb{R}$ and $a\mathbb{E}[Y_1] + b\mathbb{E}[Y_2]$ is well-defined (not of the form $\infty - \infty$), then

$$\mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}] = a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}] \quad a.e..$$

(d) (Triangle inequality.)

$$\left|\mathbb{E}[Y \mid \mathcal{G}]\right| \leq \mathbb{E}\left[|Y| \mid \mathcal{G}\right] \quad a.e.$$

(e) (Monotone convergence theorem.) If $Y_n \geq 0$ for all n and $Y_n \uparrow Y$ a.e., then

$$\mathbb{E}[Y_n \mid \mathcal{G}] \uparrow \mathbb{E}[Y \mid G] \quad a.e.$$

(f) (Non-negative series can be integrated term by term.) If $Y_n \geq 0$, then

$$\mathbb{E}\left[\sum_{n=1}^{\infty} Y_n \mid \mathcal{G}\right] = \sum_{n=1}^{\infty} \mathbb{E}[Y_n \mid G] \quad a.e.$$

(g) (Countable additivity of conditional probability.) In particular, if $B_1, B_2, \ldots$ are disjoint sets in $\mathcal{F}$, then

$$P\left(\bigcup_{n=1}^{\infty} B_n \mid \mathcal{G}\right) \uparrow \sum_{n=1}^{\infty} P(B_n \mid G) \quad a.e.$$

(h) (Dominated convergence theorem.) If $|Y_n| \leq Z$ for all n, with $\mathbb{E}[Z]$ finite and $Y_n \to Y$ a.e., then

$$\mathbb{E}[Y_n \mid \mathcal{G}] \to \mathbb{E}[Y \mid \mathcal{G}] \quad a.e.$$

(i) (Extended monotone convergence theorem.) Assume $Y_n \geq Z$ for all n, where $\mathbb{E}[Z] > -\infty$. If $Y_n \uparrow Y$ a.e., then

$$\mathbb{E}[Y_n \mid \mathcal{G}] \uparrow \mathbb{E}[Y \mid \mathcal{G}] \quad a.e.$$

(j) (Fatou's lemma.) Assume $Y_n \geq Z$ for all n, where $\mathbb{E}[Z] > -\infty$. Then

$$\liminf_{n\to\infty} \mathbb{E}[Y_n \mid \mathcal{G}] \geq \mathbb{E}[\liminf_{n\to\infty} Y_n \mid \mathcal{G}] \quad a.e$$

(k) (The law of total expectation.)

$$\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}]] = \mathbb{E}[Y]$$

(l) *(Tower property.) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then*

   (I) $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[Y \mid \mathcal{G}_1]$ *a.e.*

   (II) $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \mathbb{E}[Y \mid \mathcal{G}_1]$ *a.e.*

(m) *("Taking out what is known.") If $Z \in \mathcal{G}$ and $\mathbb{E}|Y|, \mathbb{E}|YZ| < \infty$, then*

$$\mathbb{E}[YZ \mid \mathcal{G}] = Z\,\mathbb{E}[Y \mid \mathcal{G}] \quad \textit{a.e.}$$

Remark 17.5.2. Some verbal description of these properties:

(k) If we take the expectation of a conditional expectation, the result is the same as if we were to take the expectation directly. TODO: Add intuition.

(l) This property handles a scenario where we do successive conditioning with respect to two sigma-fields, one of which is coarser than (that is, a subset of) the other. The observation is that *the smaller σ-field always wins.*

(m) This result shows that for conditional expectation with respect to $\mathcal{G}$, random variables $X \in \mathcal{G}$ (i.e. $\mathcal{G}$-measurable random variables) are like constants. They can be brought outside the integral.

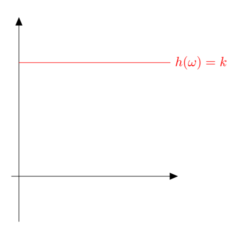
$\triangle$

Proof. (a) Let $h = \mathbb{E}[Y \mid \mathcal{G}] : \Omega \to \mathbb{R}$ be defined a.e. by $h(\omega) = k$. We verify the two conditions for the existence (and essential uniqueness) of conditional expectation.

1. $h^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}(\overline{\mathbb{R}})$)? ✓ . We have

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

and $\Omega, \emptyset \in \mathcal{G}$.



2. $\int_G Y\,dP = \int_G h\,dP$ for all $G \in \mathcal{G}$? ✓ . By hypothesis on Y and our guess on h, the condition becomes $\int_G k\,dP = \int_G k\,dP$, which obviously holds.

(b) If $Y_1 \le Y_2$ a.e., then

$$\begin{aligned}
& & Y_1 1_G &\le Y_2 1_G & &\text{a.e. } \forall G \in \mathcal{G} & &\text{obvious} \\
&\implies & \int_G Y_1\,dP &\le \int_G Y_2\,dP & &\forall G \in \mathcal{G} & &\text{by monotonicity} \\
&\implies & \int_G \mathbb{E}[Y_1 \mid \mathcal{G}]\,dP &\le \int_G \mathbb{E}[Y_2 \mid \mathcal{G}]\,dP & &\forall G \in \mathcal{G} & &\text{by def. cond'l expect.} \\
&\implies & \mathbb{E}[Y_1 \mid \mathcal{G}] &\le \mathbb{E}[Y_2 \mid \mathcal{G}] & &\text{a.e.} & &\text{by monotonicity converse; Thm. 7.3.3}
\end{aligned}$$

(c) By the definition of conditional expectation,

$$\int_G (aY_1 + bY_2)\,dP = \int_G \mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}]\,dP \qquad \forall G \in \mathcal{G} \qquad (1)$$

On the other hand, $\forall\, G \in \mathcal{G}$, we have

$$\begin{aligned}
\int_G (aY_1 + bY_2)\, dP \ &= a\int_G Y_1\, dP \ + b\int_G Y_2\, dP && \text{linearity} \\
&= a\int_G \mathbb{E}[Y_1 \mid \mathcal{G}]\, dP \ + b\int_G \mathbb{E}[Y_2 \mid \mathcal{G}]\, dP && \text{def. cond'l expect.} \\
&= \int_G \big(a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}]\big)\, dP && \text{linearity} \quad (2)
\end{aligned}$$

Combining (1) and (2), we obtain

$$\int_G \mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}]\, dP \ = \int_G \big(a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}]\big)\, dP \qquad \forall G \in \mathcal{G}$$

Since the integrands are equal for all measurable sets, Cor. 7.3.1 yields the conclusion

$$\mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}]\, dP \ = a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}] \qquad a.e.$$

(d) [98]

$$\begin{array}{rcccccl}
 & -|Y| & \leq & Y & \leq & Y & \text{obvious} \\
\Longrightarrow & \underbrace{\mathbb{E}[-|Y| \mid \mathcal{G}]}_{= -\mathbb{E}[|Y| \mid \mathcal{G}] \text{ by part (c)}} & \leq & \mathbb{E}[Y \mid \mathcal{G}] & \leq & \mathbb{E}[|Y| \mid \mathcal{G}] & \text{monotonicity (part b)} \\
\Longrightarrow & \big|\mathbb{E}[Y \mid \mathcal{G}]\big| & \leq & \mathbb{E}[|Y| \mid \mathcal{G}] & & & \text{def. absolute value}
\end{array}$$

(e) We have $Y_n \geq 0\, \forall\, n$ and $Y_n \uparrow Y$ a.e. By monotonicity of conditional expectation [part (b)], $\{\mathbb{E}[Y_n \mid \mathcal{G}]\}_n$ is an increasing sequence of non-negative functions. Since also measurable functions are closed under limits, we find

$$\mathbb{E}[Y_n \mid \mathcal{G}] \uparrow h \qquad (1)$$

where h is $\mathcal{G}$-measurable and non-negative.

Now applying MCT to the hypothesis ($Y_n \uparrow Y$ a.e.), we see

$$\int_G \mathbb{E}[Y_n \mid \mathcal{G}]\, dP \ \overset{\textit{def. cond. expect.}}{=} \int_G Y_n\, dP \ \overset{\textit{MCT}}{\uparrow} \int_G Y\, dP \qquad \forall G \in \mathcal{G}.$$

But applying MCT to (1), we also see

$$\int_G \mathbb{E}[Y_n \mid \mathcal{G}]\, dP \ \overset{\textit{MCT}}{\uparrow} \int_G h\, dP \qquad \forall G \in \mathcal{G}.$$

So by uniqueness of limits, we have

$$\int_G h\, dP \ = \int_G Y\, dP \qquad \forall G \in \mathcal{G} \qquad (2)$$

Result (2) and the $\mathcal{G}$-measurability of h reveal that $h = \mathbb{E}[Y \mid \mathcal{G}]$ by the definition of conditional expectation. Hence, we can rewrite (1) as

$$\mathbb{E}[Y_n \mid \mathcal{G}] \uparrow \mathbb{E}[Y \mid \mathcal{G}].$$

and we are done.

(f) Let all $Y_n \geq 0$. We have

$$
\begin{aligned}
\mathbb{E}[\sum_{k=1}^{n} Y_k \mid \mathcal{G}] &= \sum_{k=1}^{n} \mathbb{E}[Y_k \mid \mathcal{G}] \quad a.e. && \text{by linearity [part (c)]} \\
\implies \lim_{n\to\infty} \mathbb{E}[\sum_{k=1}^{n} Y_k \mid \mathcal{G}] &= \sum_{k=1}^{\infty} \mathbb{E}[Y_k \mid \mathcal{G}] \quad a.e. && \text{Take } n \to \infty \\
\stackrel{1}{\implies} \mathbb{E}[\sum_{k=1}^{\infty} Y_k \mid \mathcal{G}] &= \sum_{k=1}^{\infty} \mathbb{E}[Y_k \mid \mathcal{G}]
\end{aligned}
$$

where (1) holds by applying MCT for conditional expectation [part (e)] to the sequence of partial sums, since $(S_n := \sum_{k=1}^{n} Y_k) \uparrow (S := \sum_{k=1}^{\infty} Y_k)$ and each $S_n \geq 0$.

(g) Let $B_1, B_2, \ldots$ be disjoint sets in $\mathcal{F}$. We will apply the result that non-negative series can be integrated term by term [part (f)] in the special case where $Y_n \triangleq 1_{B_n}$. Note that in this case

$$
\sum_{n=1}^{\infty} Y_n \stackrel{def.}{=} \sum_{n=1}^{\infty} 1_{B_n} = 1_{\cup_{n=1}^{\infty} B_n}
$$

The left and right hand sides of part (f) become expectations of indicators, and therefore probabilities. So part (f) becomes

$$
P\Big(\bigcup_{n=1}^{\infty} B_n \mid \mathcal{G}\Big) \uparrow \sum_{n=1}^{\infty} P(B_n \mid G) \quad a.e.
$$

and we are done.

(h) See [Ash et al., 2000, Thm. 5.5.5].

(i) See [Ash et al., 2000, Thm. 5.5.6a].

(j) See [Ash et al., 2000, Thm. 5.5.6b].

(k) By the definition of conditional expectation for $\mathbb{E}[Y \mid \mathcal{G}]$, we have

$$
\int_{\Omega} \mathbb{E}[Y \mid \mathcal{G}] \, dP = \int_{\Omega} Y \, dP \, .
$$

(l) (I) We want to prove that if $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[Y \mid \mathcal{G}_1]$ a.e. To do so, we N.T.S. that the two conditions defining conditional expectation hold.

1. $\mathbb{E}[Y \mid \mathcal{G}_1]$ is $\mathcal{G}_1$ measurable? ✓ . Automatic.
2. $\int_{G_1} \mathbb{E}[Y \mid \mathcal{G}_2] \, dP = \int_{G_1} \mathbb{E}[Y \mid \mathcal{G}_1] \, dP \quad \forall \, G_1 \in \mathcal{G}_1$? ✓ . By definitions of conditional expectation for $\mathbb{E}[Y \mid \mathcal{G}_1]$ and $\mathbb{E}[Y \mid \mathcal{G}_2]$, respectively, we have

$$
\begin{aligned}
\int_{G_1} \mathbb{E}[Y \mid \mathcal{G}_1] \, dP &= \int_{G_1} Y \, dP \quad \forall \, G_1 \in \mathcal{G}_1 && (1) \\
\int_{G_2} \mathbb{E}[Y \mid \mathcal{G}_2] \, dP &= \int_{G_2} Y \, dP \quad \forall \, G_2 \in \mathcal{G}_2 && (2)
\end{aligned}
$$

Now since $\mathcal{G}_1 \subset \mathcal{G}_2$, equation (2) also applies to all $G_1 \in \mathcal{G}_1$, so we have

$$
\int_{G_1} \mathbb{E}[Y \mid \mathcal{G}_1] \, dP \stackrel{(1)}{=} \int_{G_1} Y \, dP \stackrel{(2)}{=} \int_{G_1} \mathbb{E}[Y \mid \mathcal{G}_1] \, dP \quad \forall \, G_1 \in \mathcal{G}_1.
$$

(II) We want to prove that if $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \mathbb{E}[Y \mid \mathcal{G}_1]$ a.e. To do so, we N.T.S. that the two conditions defining conditional expectation hold.

1. $\mathbb{E}[Y \mid \mathcal{G}_1]$ is $\mathcal{G}_2$ measurable? ✓. True since $\mathcal{G}_1 \subset \mathcal{G}_2$.
2. $\int_{G_2} \mathbb{E}[Y \mid \mathcal{G}_1]\, dP = \int_{G_2} \mathbb{E}[Y \mid \mathcal{G}_1]\, dP \quad \forall\, G_2 \in \mathcal{G}_2$? ✓. Obvious.

(m) We N.T.S. that the two conditions defining conditional expectation $\mathbb{E}[YZ \mid \mathcal{G}]$ hold.

1. $Z\,\mathbb{E}[Y \mid \mathcal{G}]$ is $\mathcal{G}$ measurable? ✓. True since each term is $\mathcal{G}$ measurable individually, and the product of measurable functions is measurable.
2. We need to show

$$\int_G YZ\, dP = \int_G Z\,\mathbb{E}[Y \mid \mathcal{G}]\, dP \quad \forall\, G \in \mathcal{G} \qquad 1$$

We proceed by mimicking the construction of the integral.

(I) $\boxed{\text{Indicator functions.}}$ Let $Z = 1_{\widetilde{G}}$ be an indicator function for $\widetilde{G} \in G$. Then

$$\underbrace{\int_G YZ\, dP}_{\text{LHS of } 1} \overset{\text{def. } Z}{=} \int_G Y 1_{\widetilde{G}}\, dP = \int_{G \cap \widetilde{G}} Y\, dP$$

$$\underbrace{\int_G Z\,\mathbb{E}[Y \mid \mathcal{G}]\, dP}_{\text{RHS of } 1} \overset{\text{def. } Z}{=} \int_G 1_{\widetilde{G}}\,\mathbb{E}[Y \mid \mathcal{G}]\, dP = \int_{G \cap \widetilde{G}} \mathbb{E}[Y \mid \mathcal{G}]\, dP$$

and the right hand sides of the two equations above are equal a.e. by the definition of $\mathbb{E}[Y \mid \mathcal{G}]$ (note that $G \cap \widetilde{G} \in \mathcal{G}$). Thus 1 holds.

(II) $\boxed{\text{Simple functions.}}$ Let Z be simple, i.e $Z = \sum_{i=1}^r a_i 1_{G_i}$ for $G_i \in \mathcal{G}$. Then again

$$\underbrace{\int_G YZ\, dP}_{\text{LHS of } 1} \overset{\text{def. } Z}{=} \int_G Y\Big(\sum_{i=1}^r a_i 1_{G_i}\Big)\, dP \overset{\text{linearity}}{=} \sum_{i=1}^r a_i \int_{G \cap G_i} Y\, dP$$

$$\underbrace{\int_G Z\,\mathbb{E}[Y \mid \mathcal{G}]\, dP}_{\text{RHS of } 1} \overset{\text{def. } Z}{=} \int_G \Big(\sum_{i=1}^r a_i 1_{G_i}\Big)\mathbb{E}[Y \mid \mathcal{G}]\, dP \overset{\text{linearity}}{=} \sum_{i=1}^r a_i \int_{G \cap G_i} \mathbb{E}[Y \mid \mathcal{G}]\, dP$$

and again the right hand sides of the two equations above are equal a.e. by the definition of $\mathbb{E}[Y \mid \mathcal{G}]$ (note that $G \cap \widetilde{G} \in \mathcal{G}$). Thus 1 holds.

(III) $\boxed{\text{Arbitrary measurable functions.}}$ By Prop. 6.3.7, there exists a sequence of $\mathcal{G}$-measurable simple functions (S_n) such that

$$|S_n| \leq Z \qquad \text{and} \qquad S_n \to Z$$

where the limit is pointwise. Now by step (b), we have

$$\mathbb{E}[YS_n \mid \mathcal{G}] = S_n\,\mathbb{E}[Y \mid \mathcal{G}] \quad \text{a.e.}\,.$$

But the LHS $\to \mathbb{E}[YZ \mid \mathcal{G}]$ by the dominated convergence theorem for conditional expectation [Ash et al., 2000, Thm. 5.5.5] (noting that since $S_n \to Z$, then $YS_n \to YZ$), and the RHS $\to Z\mathbb{E}[Y \mid \mathcal{G}]$ by limit properties.

□

[98] [Ash et al., 2000, pp.220-221] proves (d) before (c). I am not sure how this is possible.

17.5.3 Conditional expectation as averages over atoms

Definition 17.5.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then A is an **atom** of $\mathcal{F}$ relative to μ if

1. $A \in \mathcal{F}$, $\mu(A) > 0$
2. For all $B \in \mathcal{F}, B \subset A$, either $\mu(B) = 0$ or $\mu(B - A) = 0$.

$\triangle$

Example 17.5.1. (*A measurable partition is composed of atoms.*)[99] Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $A_1, A_2, \ldots$ be disjoint sets in $\mathcal{F}$ whose union is Ω. Let $\mu(A_n) > 0$ for all n. Let $\mathcal{G}$ be the minimal σ-field over $\{A_n\}$ (i.e. $\mathcal{G}$ is the collection of all unions formed from the A_n).[100] Then $A_1, A_2, \ldots$ are atoms of $\mathcal{G}$ relative to μ. $\triangle$

Proof. We verify the two conditions in Def. 17.5.1.

1. This condition holds by hypothesis.
2. The σ-field $\mathcal{G}$ is the collection of all unions formed from the A_n. Thus, each A_n has no proper subset. There is nothing to verify.

$\square$

Lemma 17.5.1. (Measurable functions are almost everywhere constant on atoms.) *If $f : (\Omega, \mathcal{G}) \to \overline{\mathbb{R}}, \mathcal{B})$, μ is a measure on $\mathcal{G}$, and A is an atom of $\mathcal{G}$ relative to μ, then f is a.e. constant on A.*

Proof. See [Ash et al., 2000, Lemma 5.5.8, pp. 225]. $\square$

Now we show that conditional expectation is an "averaging" or "smoothing" operation; if A is an atom of $\mathcal{G}$, then $\mathbb{E}[Y \mid \mathcal{G}] = k$ a.e. on A, where k is the average value of Y on A.

Theorem 17.5.2. (Conditional expectation gives averages over atoms.) *Let A be an atom of $\mathcal{G}$ relative to P. Then*

$$\mathbb{E}[Y \mid \mathcal{G}] = \frac{1}{P(A)} \int_A Y \, dP = \frac{\mathbb{E}[Y 1_A]}{P(A)} \qquad \text{a.e. on } A$$

Proof. The function $f \triangleq \mathbb{E}[Y \mid \mathcal{G}]$ on Ω is measurable with respect to $\mathcal{G}$ by definition of conditional expectation. Thus, if A is an atom of $\mathcal{G}$ relative to P, Lemma 17.5.1 tells us that $f 1_A = k$ a.e. for some constant k. Now we appeal to the definition of conditional expectation to obtain

$$\begin{aligned} \int_A Y \, dP &= \int_A f \, dP && \text{def. conditional expect.} \\ &= \int_A k \, dP && \text{see above} \\ &= k P(A) && \text{integral of indicator} \end{aligned}$$

Dividing gives

$$k = \frac{1}{P(A)} \int_A Y \, dP$$

$\square$

[99] More precisely, we could say that a measurable partition (whose cells all have positive probability) is composed of atoms (of the minimal σ-field containing them).

[100] TODO: Add a small proof.

Example 17.5.2. (*A simple example of conditional expectation as averaging over atoms.*) [Zitkovic, 2013a]. Let $(\Omega, \mathcal{F}, P)$ be a probability space where $\Omega = \{a, b, c, d, e, f\}$, $\mathcal{F} = 2^\Omega$ and P is uniform. Let X, Y and Z be random variables given by

$$X \sim \begin{pmatrix} a & b & c & d & e & f \\ 1 & 3 & 3 & 5 & 5 & 7 \end{pmatrix},$$
$$Y \sim \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 1 & 1 & 7 & 7 \end{pmatrix}, \quad \text{and} \quad Z \sim \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 2 & 2 \end{pmatrix}.$$

Then

$$\sigma(Y) = \{Y^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\} = \sigma\big(\{a, b\}, \{c, d\}, \{e, f\}\big)$$

by Example 17.4.2.

Hence, by Example 17.5.1, the atoms of $\sigma(Y)$ are given by $A_1 \triangleq \{a, b\}, A_2 \triangleq \{c, d\}, A_3 \triangleq \{e, f\}$. So by Theorem 17.5.2,

$$\mathbb{E}[X \mid \sigma(Y)] = \frac{1}{P(A_n)} \int_{A_n} X \, dP \qquad \text{a.e. on } A_n.$$

We can compute this function explicitly:

- *Denominator.* Since P is uniform, we compute the denominator as $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$.
- *Numerator.*

$$\int_{A_1} X \, dP \overset{\text{(simple function)}}{=} X(\{a\})P\{a\} + X(\{b\})P\{b\} = 1 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} = \frac{4}{6},$$

  and similarly for A_2, A_3.

Thus, all together, we have

$$\mathbb{E}[X \mid \sigma(Y)] = \begin{cases} 2 & \text{on } \{a, b\} \\ 4 & \text{on } \{c, d\} \\ 6 & \text{on } \{e, f\} \end{cases}$$

A similar computation gives $\mathbb{E}[Z \mid \sigma(Y)]$. See Fig. 31.

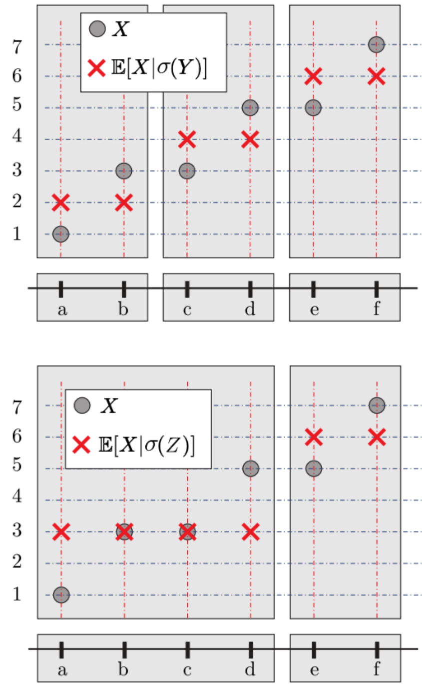


Figure 31: *A simple example of conditional expectation as averaging over atoms.* Image credit: [Zitkovic, 2013a].

△

TODO: Can we decompose any probability space into atoms? I think no.

Considering conditional expectation as averages over atoms is useful for various exercises. For instance, below we show via a counterexample that the tower property does not hold when σ-fields are not nested.

Exercise 17.5.1. [Durrett, 2010, Exercise 5.1.6] Give an example on $\Omega = \{a, b, c\}$ in which

$$\mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_2] \mid \mathcal{G}_1] \neq \mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}_1] \mid \mathcal{G}_2].$$

△

Solution. Similarly to Example 17.5.2, let $(\Omega, \mathcal{F}, P)$ be a probability space where $\Omega = \{a, b, c\}$, $\mathcal{F} = 2^\Omega$ and P is uniform. Let X be a random variable defined by

$$X \sim \begin{pmatrix} a & b & c \\ 2 & 4 & 6 \end{pmatrix}$$

Define

$$\mathcal{G}_1 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$$
$$\mathcal{G}_2 = \{\emptyset, \{a, b\}, \{c\}, \Omega\}$$

Then we have

$$\mathbb{E}[X \mid \mathcal{G}_1] = \begin{cases} 2 & \text{on } \{a\} \\ 5 & \text{on } \{b, c\} \end{cases}$$

$$\mathbb{E}[X \mid \mathcal{G}_2] = \begin{cases} 4 & \text{on } \{b\} \\ 4 & \text{on } \{a, c\} \end{cases}$$

by Theorem 17.5.2, since $\{a\}, \{b, c\}$ are atoms of $\mathcal{G}_1$ w.r.t P and $\{a, b\}, \{c\}$ are atoms of $\mathcal{G}_2$ w.r.t P.

We now have constructed three random variables on Ω:

Ω	a	b	c
X	2	4	6
$\mathbb{E}[X \mid \mathcal{G}_1]$	2	5	5
$\mathbb{E}[X \mid \mathcal{G}_2]$	4	4	4

Applying the same logic again (i.e. taking averages over atoms, per Theorem 17.5.2), we obtain

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \begin{cases} 3.5 & \text{on } \{a, b\} \\ 5 & \text{on } \{c\} \end{cases}$$

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \begin{cases} 4 & \text{on } \{a\} \\ 4 & \text{on } \{b, c\} \end{cases}$$

Clearly $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] \neq \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1]$. □

17.5.4 Conditional expectation as projection

Theorem 17.5.3. *[Durrett, 2010, Thm. 5.1.8] Suppose $\mathbb{E}[X^2] < \infty$. Then $\mathbb{E}[X \mid \mathcal{G}]$ is the variable $Y \in \mathcal{G}$ that minimizes the "mean squared error" $\mathbb{E}(X - Y)^2$.*

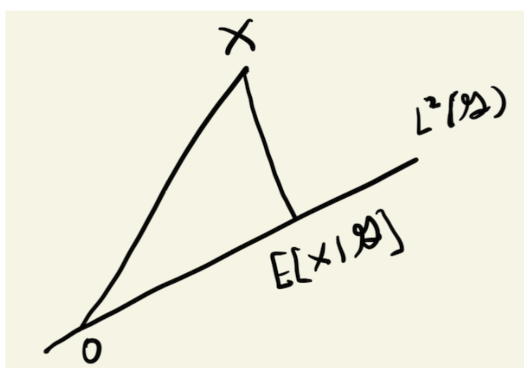


Figure 32: Conditional expectation as projection in L^2.

Remark 17.5.3. [Durrett, 2010, pp.229] This result gives a "geometric interpretation" of $\mathbb{E}[X \mid \mathcal{G}]$ (see Figure 32). In particular, $L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space, and $L^2(\Omega, \mathcal{G}, P)$ is a closed subspace.[101] Since $Y = \mathbb{E}[X \mid \mathcal{G}]$ minimizes the (squared) distance function $E(X - Y)^2$, it is the projection of X onto $L^2(\Omega, \mathcal{G}, P)$. That is, it is the point in the subspace closest to X. △

[101] TODO: Give a proof of this.

Proof. Let $Y \in L^2(\Omega, \mathcal{G}, P)$ and define the "residual" as $Z \triangleq E[X \mid \mathcal{G}] - Y$. We will show that the mean squared error is optimized when $Z = 0$.

$$
\begin{aligned}
\mathbb{E}(X-Y)^2 &= \mathbb{E}(X - E[X \mid \mathcal{G}] + Z)^2 && \text{def. Z}\\
&= \mathbb{E}(X - E[X \mid \mathcal{G}])^2 + \mathbb{E}[Z^2] + 2\mathbb{E}[(X - E[X \mid \mathcal{G}])Z] && \text{FOIL}\\
&\overset{1}{=} \mathbb{E}(X - E[X \mid \mathcal{G}])^2 + \mathbb{E}[Z^2] + 2\,\cancelto{0}{\mathbb{E}[(X - E[X \mid \mathcal{G}])Z]}
\end{aligned}
$$

which is clearly minimized when $Z = 0$.

It remains to verify Equation (1). Now $Z \in L^2(\Omega, \mathcal{G}, P)$ since it is a subspace, so by the "taking out what is known" property of conditional expectation (Thm. 17.5.1), we have

$$
\underset{\substack{\| \\ \mathbb{E}[ZX]}}{\mathbb{E}[ZX \mid \mathcal{G}]} = Z\,\mathbb{E}[X \mid \mathcal{G}].
$$

The vertical equality holds by the law of total expectation.

Applying linearity, we obtain

$$
\mathbb{E}\big[Z(X - \mathbb{E}[X \mid \mathcal{G}])\big] = 0,
$$

which justifies Equation (1). $\square$

17.5.5 Jensen's Inequality[102]

In geometry, a **line of support** L of a curve C in the plane is a line that contains at least one point of C and causes C to lie completely in one of the two closed half planes defined by L. Theorem 17.5.4 shows that given any real-valued convex function defined on an open interval, we can construct a sequence of lines of support such that the convex function evaluates to the supremum over the lines of support.

Theorem 17.5.4. ***Line of Support Theorem.*** *Let $g : I \to \mathbb{R}$, where I is an open interval of reals, bounded or unbounded. Assume g is convex, that is,*

$$
g\big(\alpha x + (1-\alpha)y\big) \le \alpha g(x) + (1-\alpha)g(y)
$$

for all $x, y \in I$ and all $\alpha \in [0,1]$. Then there are sequences $\{a_n\}, \{b_n\}$ of real numbers such that for all $x \in I$,

$$
g(x) = \sup_n \, (a_n x + b_n)
$$

Proof. See [Ash et al., 2000, Theorem 6.3.4]. $\square$

[102] This material is covered in [Ash et al., 2000, Thm 6.3.4, 6.3.5].

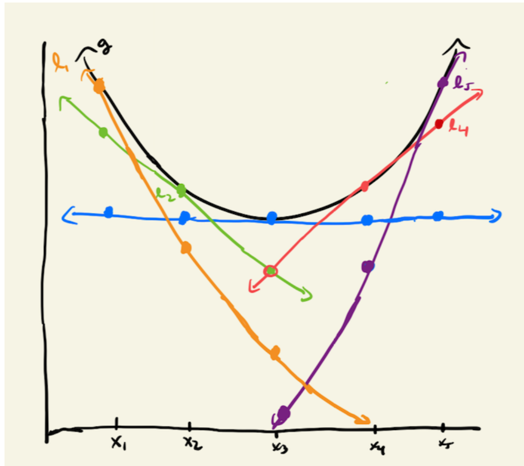


Figure 33: *Cartoon sketch of line of support theorem.* Shown are a convex function g and five elements of a sequence $\{\ell_n\}$ of lines of support.

Question 17.5.1. How is a *countable* number of lines sufficient to exactly match the *uncountable* number of points on the curve? (A study of the proof should reveal the answer.) $\triangle$

Remark 17.5.4. (*The line of support theorem does not hold for non-open domains.*) For a counter-example, let $I = [0, 1]$, and define $g : I \to \mathbb{R}$ by

$$g(x) = \begin{cases} 0, & x \in [0, 1) \\ 1, & x = 1. \end{cases}$$

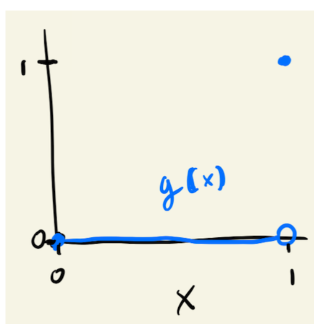


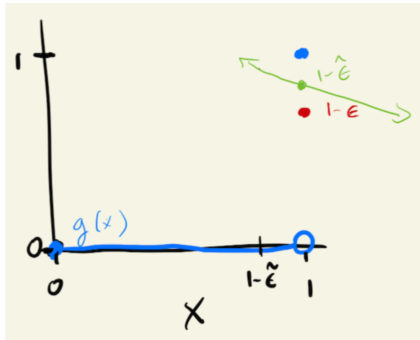
Figure 34: The function g is convex (for any pair of points in the domain, the secant line lies above or on the graph of the function). However, g is defined on the interval $[0, 1]$, which is not open. Therefore, the line of support theorem not apply.

The line of support theorem would imply that there exists a sequence of lines ℓ_n (with slopes m_n and intercepts b_n) such that

$$g(x) = \sup_n (m_n x + b_n) \qquad (1)$$

We show that (1) cannot hold.

1. By the alternate characterization of the supremum (Prop A.1.1), we can find some line in the sequence $\{\ell_n\}$ whose value at $x = 1$ is arbitrarily close to 1. That is, for any $\epsilon > 0$, there is some $\widehat{\epsilon}$ satisfying $0 < \widehat{\epsilon} < \epsilon$ and some natural number n such that $\ell_n(1) = 1 - \widehat{\epsilon}$.



2. But then when $x = 1 - \widetilde{\epsilon}$, the line takes value $\widetilde{y} = (1 - \widehat{\epsilon}) - m\widetilde{\epsilon}$ (where we have defined $m \triangleq m_n$). At this point, the line is above 0 (violating the upper bound condition of the supremum) whenever

$$\widetilde{y} = (1 - \widehat{\epsilon}) - m\widetilde{\epsilon} > 0 \implies \widehat{\epsilon} + m\widetilde{\epsilon} < 1,$$

which is always possible since $\widehat{\epsilon}$ and $\widetilde{\epsilon}$ can be made arbitrarily small.

△

Theorem 17.5.5. *Let g be a convex function from I to $\mathbb{R}$, where I is an open interval of reals, bounded or unbounded. Let X be random variable on $(\Omega, \mathcal{F}, P)$, with $X(\omega) \in I$ for all ω. Assume $E[X]$ to be finite.*[103] *If $\mathcal{H}$ is a sub σ-field of $\mathcal{F}$, then*

$$\mathbb{E}[g(X) \mid \mathcal{H}] \geq g(\mathbb{E}[X \mid \mathcal{H}]) \quad a.e. \tag{17.5.1}$$

In particular, $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$.

Remark 17.5.5. (*Remembering the direction of Jensen's inequality (via a discrete distribution over two outcomes).*) To recall the direction in which the inequality goes [Durrett, 2010, pp.28], take $P(X = x) = \lambda, P(X = y) = 1 - \lambda$. Then (see Figure 35)

$$\begin{aligned}
\mathbb{E}[g(X)] &= \lambda\, g(x) + (1-\lambda)\, g(y) && \text{LOTUS} \\
&\geq g\big(\lambda x + (1-\lambda) y\big) && \text{def. convexity} \\
&= g\big(\mathbb{E}[X]\big). && \text{def. expectation}
\end{aligned}$$

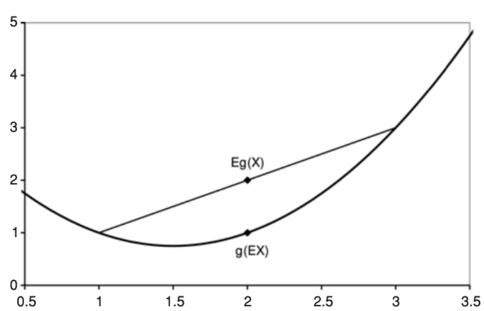


Figure 35: Jensen's inequality for $g(x) = x^2 - 3x + 3, P(X = 1) = P(X = 3) = 1/2$. Image credit: [Durrett, 2010, pp.28].

△

[103] Why do we make this assumption? See Question 17.5.2.

Proof. 1. *We show that the expression $g(\mathbb{E}[X \mid \mathcal{H}])$ is well-defined.* First, note that $\mathbb{E}[g(X \mid \mathcal{H})] \in I$ a.e. For if, say, a is real and $X > a$, then $\mathbb{E}[g(X \mid \mathcal{H})] > a$ a.e. because

$$\begin{aligned}
0 &\geq \int_{\{\mathbb{E}[X \mid \mathcal{H}] \leq a\}} \mathbb{E}[X - a \mid \mathcal{H}]\, dP && \text{linearity of C.E., set of integration} \\
&\overset{1}{=} \int_{\{\mathbb{E}[X \mid \mathcal{H}] \leq a\}} (X - a)\, dP && \text{def of C.E.} \\
&\geq 0 && \text{we assumed } X > a
\end{aligned}$$

Note: To apply the def. of CE in (1), we must justify that the set of integration satisfies $\{\mathbb{E}[X \mid \mathcal{H}] \leq a\} \in \mathcal{H}$, but this follows immediately by the measurability condition $\mathbb{E}[X \mid \mathcal{H}] : (\Omega, \mathcal{H}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Thus, $X = a$ a.e. on $\{\mathbb{E}[X \mid \mathcal{H}] \leq a\}$. (by Theorem 7.3.2 b).

But we assumed $X > a$, so $P\{\mathbb{E}[X \mid \mathcal{H}] \leq a\} = 0$.

Thus, $g(\mathbb{E}[X \mid \mathcal{H}])$ is well-defined.

2. *Now we show that Jensen's inequality (Eq.* (17.5.1)*) holds.* We have

$$\begin{aligned}
& g(y) = \sup_n \, a_n y + b_n, \quad \forall y \in I && \text{Line of Support Theorem (Theorem 17.5.4)} \\
\implies & g(y) \geq a_n y + b_n && \text{sup is UB} \\
\implies & g(X) \geq a_n X + b_n && \text{hypothesis } (X \in I) \\
\implies & \mathbb{E}[g(X) \mid \mathcal{H}] \geq a_n \mathbb{E}[X \mid \mathcal{H}] + b_n \quad a.e. && \text{monotonicity of C.E., linearity} \\
\implies & \mathbb{E}[g(X) \mid \mathcal{H}] \geq \underbrace{\sup_n \{a_n \mathbb{E}[X \mid \mathcal{H}] + b_n\}}_{= g(\mathbb{E}[X \mid \mathcal{H}]) \text{ by LoS Thm.}} \quad a.e. && \text{take sup over } n
\end{aligned}$$

3. *Now we show that the statement holds with unconditional expectations as a special case.* If we set $\mathcal{H} = \{\emptyset, \Omega\}$ to be the trivial σ-field, then any random variable[104] is independent of $\mathcal{H}$ (see Example 16.4.1), and so the conditionals disappear by the “no information" rule of conditional expectation (see Example 17.4.1).

□

Remark 17.5.6. (*On representing convex functions with linear functions.*) In the proof of Jensen's inequality, we see that the Line of Support Theorem (Theorem 17.5.4) allows us to take an arbitrary convex function and represent it in terms of linear functions (and therefore apply properties that hold under linearity). △

Remark 17.5.7. (*Theorems about conditional expectations automatically apply to unconditional expectations.*) Step 3 in the proof of Jensen's inequality gives a general strategy through which statements about conditional expectations will also hold for unconditional expectations. △

Question 17.5.2. Does Jensen's Inequality (Theorem 17.5.5) fail for infinite expectations? If so, why? △

18 § 6 Strong laws of large numbers and martingales

18.1 § 6.1 Introduction

We begin by defining a regular matrix summability method (Def. 18.1.1). This is a matrix transformation of a convergence sequence which preserves the limit (Thm. 18.1.1).

[104] In particular, here we consider the random variables X and $g(X)$.

Definition 18.1.1. Let $\boldsymbol{A} \triangleq \{a_{rc}\}$ be an infinite matrix of real numbers. Then $\boldsymbol{A}$ is a matrix summability method if

1. $\lim_{r\to\infty} a_{rc} = 0, \quad \forall c \in \mathbb{N}$ (Each column sequence converges to 0.)
2. $\lim_{r\to\infty} \sum_{c=1}^{\infty} a_{rc} = 1, \quad \forall c \in \mathbb{N}$ (The row sums converge to 1.)
3. $\sum_{c=1}^{\infty} |a_{rc}| \leq B < \infty, \quad \forall r \in \mathbb{N}$ (The absolute row sums are bounded.)

△

Example 18.1.1. An example is **Cesaro summation**, a matrix summability method with

$$a_{rc} = \begin{cases} \frac{1}{r}, & r \leq c \\ 0, & r > c \end{cases} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \dots \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \dots \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \dots \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

△

Theorem 18.1.1. ***Silverman-Toeplitz Theorem.*** *Let $\boldsymbol{A} \triangleq \{a_{rc}\}$ be a regular matrix summability method (Def. 18.1.1). Let $\{x_c\}$ be real-valued sequence such that $x_c \to x$. Define*

$$y_r \triangleq \sum_{c=1}^{\infty} a_{rc} x_c \tag{18.1.1}$$

Then $y_r \to x$.

Remark 18.1.1. For the sequence $\{y_r\}$ in Eq. (18.1.1) to exist, it is sufficient for $\{x_c\}$ to be bounded. Suppose $\{x_c\}$ is bounded by some number $M < \infty$. Then $\sum_{c=1}^{\infty} |a_{rc} x_j| \leq BM < \infty$ (where B was defined in Def. 18.1.1). That is, the series doesn't take the form of $\infty - \infty$. (To verify existence, consider the series as an integral against counting measure, as in Example 6.3.3, and then refer to the condition of the existence of the integral, as in Remark 6.3.13.) In particular, $\{y_r\}$ exists whenever $\{x_c\}$ converges, since every convergent sequence is bounded. △

Proof. We argue as follows.

a) Conditions 1, 3, $x_c \to 0$ implies $y_r \to 0$.

b) Conditions 1, 2, 3, $x_c \to x$ implies $y_r \to x$.

where the conditions refer to those of Def. 18.1.1.

Now we verify each step.

a) For any $\epsilon > 0$, we have

$$|y_r| = \left| \sum_{c=1}^{\infty} a_{rc} x_c \right| \overset{\text{tr. ineq.}}{\leq} \sum_{c=1}^{\infty} |a_{rc} x_c| \overset{1}{=} \underbrace{\sum_{c=1}^{C} |a_{rc}||x_c|}_{\leq \frac{C\,\epsilon \sum_{c=1}^{C} |x_c|}{2C \sum_{c=1}^{C} |x_c|} = \frac{\epsilon}{2}} + \underbrace{\sum_{c=C+1}^{\infty} |a_{rc}||x_c|}_{\leq B \frac{\epsilon}{2B} = \frac{\epsilon}{2}}$$

Where in 1, we

- choose C so that $|x_c| \leq \epsilon/2B$ for $c \geq C$ (possible since x_c converges)

- choose R so that $|a_{rc}| \leq \epsilon/2(C\sum_{c=1}^{C}|x_c|)$ for $r \geq R$ (possible by condition 1 on $\boldsymbol{A}$).

(The triangle inequality holds by considering the series as an integral against counting measure; see Example 6.3.3.)

b)

$$\text{Set} \quad \widetilde{x_c} = x_c - x$$

$$\text{Then by (a)} \quad \widetilde{y_r} \triangleq \sum_{c=1}^{\infty} a_{rc}(x_c - x) \to 0$$

Now

$$\sum_{c=1}^{\infty} a_{rc}(x_c - x) \stackrel{1}{=} \sum_{c=1}^{\infty} a_{rc}x_c - \overbrace{\sum_{c=1}^{\infty} a_{rc}}^{1} \; x$$

where (1) holds by linearity (e.g. consider these as integrals against counting measure, as in Example 6.3.3. Then linearity holds so long as the RHS does not have the form $\infty - \infty$, which is avoided since the subtrahend equals $x \in \mathbb{R}$.) Thus

$$\widetilde{y_r} \to 0 \implies \underbrace{\sum_{c=1}^{\infty} a_{rc}x_c}_{\triangleq y_r} - x \to 0 \implies \underbrace{\sum_{c=1}^{\infty} a_{rc}x_c}_{\triangleq y_r} \to x$$

$\square$

Earlier, we presented the First Borel-Cantelli Lemma (Lemma 3.5.1). In a probabilistic context, the Lemma is stated as

$$\textit{If } A_1, A_2, \ldots \textit{ are events s.t. } \sum_n P(A_n) < \infty, \textit{ then } P(\limsup_n An) = 0.$$

According to [Ash et al., 2000], the second Borel-Cantelli Lemma gives a partial converse. TODO: How is this a converse?

Lemma 18.1.1. ***Second Borel-Cantelli Lemma.*** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $A_1, A_2, \ldots$ are independent events in $\mathcal{F}$ such that $P(A_n) = \infty$, then $P(\limsup_n A_n) = 1$.*

Proof. We begin the argument similarly as in the first Borel-Cantelli Lemma (Lemma 3.5.1).

Recall from Definition 2.1.1 that

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

where we have also introduced some notation for convenience.

Thus

$$\begin{aligned}
P\Big(\underbrace{\limsup_{n\to\infty} A_n}_{:= B}\Big) &\stackrel{1}{=} P\Big(\bigcup_{k=n}^{\infty} \underbrace{\bigcap_{n=1}^{\infty} A_k}_{:= B_n}\Big) && \text{def set limsup} \\
&\stackrel{2}{=} \lim_{n\to\infty} P\Big(\underbrace{\bigcup_{k=n}^{\infty} A_k}_{:= B_n}\Big) && \text{cty from above} \\
&\stackrel{3}{=} \lim_{n\to\infty} \lim_{m\to\infty} P\Big(\bigcup_{k=n}^{m} A_k\Big) && \text{cty from below}
\end{aligned} \tag{18.1.2}$$

(The continuity from above in (2) follows since $B_{n+1} \subset B_n$, $\lim_n B_n = B$ and $P(B_1) < \infty$. The continuity from below in (3) follows by defining $C_m \triangleq \cup_{k=n}^m A_k$, $C = \cup_{k=n}^\infty A_k$. Then $C_{m+1} \supset C_m$, $\lim_m C_m = C$, so $P(C) = \lim_{m\to\infty} P(C_m)$.)

Now we work with complements to apply independence

$$P\left(\bigcup_{k=n}^m A_k\right)^c \overset{\text{De Morgan}}{=} P\left(\bigcap_{k=n}^m A_k^c\right) \overset{\text{indep}}{=} \prod_{k=n}^m P(A_k^c) = \prod_{k=n}^m 1 - P(A_k)$$
$$\overset{1}{\leq} \prod_{k=n}^m \exp(-P(A_k)) \tag{18.1.3}$$

The inequality in (1) holds since $1-x \leq \exp(-x)$ on $[0,1]$. This can easily be shown by derivatives. Set $\ell(x) = 1-x$, $f(x) = \exp(-x)$. We have $\ell(x=0) = f(x=0) = 1$. But $\ell'(x) = 1 \leq \exp(-x) = f'(x)$ on $[0,1]$. That is, the graph of $\ell(x)$ lies below the graph of $f(x)$ on [0,1]. We construct this equality to apply the below about infinite products.

Now we use a fact about infinite products:

$$\sum_{k=1}^\infty \log b_k = \infty \implies \prod_{k=1}^\infty b_k = 0. \tag{18.1.4}$$

Setting $\{b_k\} = \{\exp(-P(A_r))\}_{r=n}^\infty$, the LHS of Eq. (18.1.4) becomes $\sum_{k=1}^\infty -P(A_k) = -\infty$ which holds by hypothesis, so Eq. (18.1.4) gives that $\prod_{k=n}^\infty \exp(-P(A_k)) = 0$. Applying this to Eq. (18.1.3), we find

$$\lim_{m\to\infty} P\left(\bigcup_{k=n}^m A_k\right)^c = 0 \implies \lim_{m\to\infty} P\left(\bigcup_{k=n}^m A_k\right) = 1$$

(The implication holds since $P(\bigcup_{k=n}^m A_k)^c = 1 - P(\bigcup_{k=n}^m A_k)$ and by additivity of limits.)

We substitute this result into Eq. (18.1.2) to prove the claim. □

18.2 § 6.2 Convergence Theorems

Theorem 18.2.1. *Let $X_1, X_2, \ldots$ be independent random variables with finite expectation. If $\sum_{n=1}^\infty \mathrm{Var}[X_n] < \infty$, then $\sum_{n=1}^\infty (X_n - \mathbb{E}[X_n])$ converges a.e.*

Proof. See [Ash et al., 2000, Thm. 6.2.1]. □

Theorem 18.2.2. ***(Kolmogorov Strong Law of Large Numbers.)*** *Let $X_1, X_2, \ldots$ be independent random variables with finite mean and variance. Let $\{b_n\}$ be an increasing sequence of positive real numbers such that $b_n \to \infty$. If*

$$\sum_{n=1}^\infty \frac{\mathrm{Var}[X_n]}{b_n^2} < \infty$$

Then (with $S_n \triangleq X_1 + \ldots X_n$)

$$\frac{S_n - \mathbb{E}[S_n]}{b_n} \to 0 \quad \textit{a.e.}$$

Proof. See [Ash et al., 2000, Thm. 6.2.2]. The proof uses Theorem 18.2.1. □

Example 18.2.1. (*Sample means converge to the population mean for independent random variables with finite means and variance.*) Let $X_1, X_2, \ldots$ be independent random variables with finite mean μ and variance σ^2. Then

$$\frac{S_n}{n} \to \mu \quad \text{a.e.}$$

△

Proof. Take $b_n = n$ in Theorem 18.2.2. The hypothesis holds since

$$\sum_{n=1}^{\infty} \frac{\text{Var}[X_n]}{b_n^2} = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Applying the conclusion of the theorem, we obtain

$$\frac{S_n - \mathbb{E}[S_n]}{b_n} = \frac{S_n - n\mu}{n} = \frac{S_n}{n} - \mu \to 0 \quad \text{a.e.}$$

□

Task 18.2.1. Compare the weak LLN to the strong LLN. △

In Lemma 18.2.1, we show that the expectation of a non-negative random variable can be bounded by Riemann sum approximations to the survival function.[105]. This lemma is used to prove the Strong Law of Large Numbers for the i.i.d case (Theorem 18.2.3), which does not require assumptions about variance.

Lemma 18.2.1. *If X is a non-negative random variable, then*

$$\sum_{n=1}^{\infty} P(X \geq n) \leq \mathbb{E}[X] \leq 1 + \sum_{n=1}^{\infty} P(X \geq n).$$

Proof. In Remark 13.5.1, we showed that

$$\mathbb{E}[X] = \int_0^{\infty} P(X > x)\, dx\ .$$

Since the survival function $P(X > x)$ is non-increasing in x, its integral $\int_0^{\infty} P(X > x)\, dx$ can be no greater than the left Riemann sum $\sum_{n=0}^{\infty} P(X \geq n) = 1 + \sum_{n=1}^{\infty} P(X \geq n)$ and no less than the right Riemann sum $\sum_{n=1}^{\infty} P(X \geq n)$. See Fig. 36. □

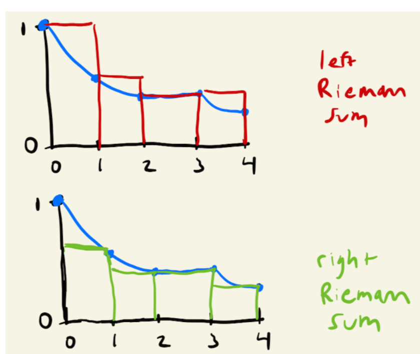


Figure 36: Approximating the area under a survival function with left and right Riemann sums.

Theorem 18.2.3. ***(Strong Law of Large Numbers, iid case.).*** *If $X_1, X_2, \ldots$ are iid random variables with finite expectation μ, and $S_n \triangleq X_1 + \ldots + X_n$, then $S_n/n \to \mu$ a.e.*

Remark 18.2.1. Theorem 18.2.3 shows that in the i.i.d case, we don't need to make an assumption about variance (as we did in Example 18.2.1). △

[105] We give a proof that is both more compact and more conceptual than the proof given by [Ash et al., 2000, Lemma 6.2.4].

Proof. See [Ash et al., 2000, Thm. 6.2.5]. The proof uses Lemma 18.2.1. $\square$

Remark 18.2.2. Theorem 18.2.3 still holds if the random variables have infinite expectation; see [Durrett, 2010, Theorem 2.4.5]. $\triangle$

Example 18.2.2. **Renewal theory.** [Durrett, 2010, pp.75] Let $X_1, X_2, \ldots$ be i.i.d with $0 < X_i < \infty$. In the context of a janitor replacing lightbulbs once they burn out, we have

$$\underbrace{T_n}_{\text{time of } n\text{-th lightbulb burning out}} \triangleq X_1 + \ldots + \underbrace{X_i}_{\text{lifetime of } i\text{-th lightbulb}} + \ldots + X_n$$

$$\underbrace{N_t}_{\text{\# of lightbulbs burned out at time } t} \triangleq \sup\{n : T_n \leq t\}$$

If $\mathbb{E}[X_i] = \mu < \infty$, then

$$\underbrace{\frac{N_t}{t}}_{\text{burnout rate}} \to \frac{1}{\mu} \quad \text{a.e.} \quad \text{as} \quad t \to \infty.$$

$\triangle$

Proof. By definition of N_t, we have

$$T_{N_t} \leq t < T_{N_t+1}.$$

Dividing by N_t, we obtain

$$\underbrace{\frac{T_{N_t}}{N_t}}_{\to \mu} \leq \frac{t}{N_t} < \underbrace{\frac{T_{N_t+1}}{N_t+1}}_{\to \mu} \cdot \underbrace{\frac{N_t+1}{N_t}}_{\to 1} \qquad (1)$$

where the convergences hold because $X_i < \infty \implies T_n < \infty \implies N_t \to \infty$ as $t \to \infty$. Convergence follows immediately for the third term, and for the first two we apply the SLLN, $\frac{T_n}{n} \to \mu$ a.e.

Thus, (1) gives us that

$$\frac{t}{N_t} \to \mu \quad \text{a.e.},$$

so by limit properties,

$$\frac{N_t}{t} \to \frac{1}{\mu} \quad \text{a.e.}$$

$\square$

18.3 § 6.3 Martingales

18.3.1 Filtrations

Definitions 18.3.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then

- A **filtration** $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ is an increasing sequence of sub σ-fields of $\mathcal{F}$.
- A sequence of random variables $\{X_n\}$ on $(\Omega, \mathcal{F}, P)$ is said to be **adapted to the filtration** if each X_n is $\mathcal{F}_n$-measurable.

$\triangle$

Example 18.3.1. (*A filtration can be constructed from the σ-fields generated by projecting random sequences of binary values onto their first n coordinates.*) Let $(\Omega = \{0,1\}^{\mathbb{R}}, \mathcal{F}, P)$ be a probability space of binary sequences.[106] Define the random variable Y_n by $\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) \mapsto \omega_n$, that is, Y_n gives the n-th coordinate of any $\omega \in \Omega$. Then by Example 17.4.2, since any random vector $(Y_1, \ldots, Y_n)$ takes on only finitely many values, we have that

- The σ-field $\mathcal{F}_1 \triangleq \sigma(Y_1)$ consists of any union of the empty set and elements in the partition of Ω according to the first entry of each outcome; i.e. $\Omega = \Omega_0 \uplus \Omega_1$, where

$$\begin{aligned} \Omega_0 &\triangleq (0, \cdot, \cdot, \cdot, \ldots) \\ \Omega_1 &\triangleq (1, \cdot, \cdot, \cdot, \ldots) \end{aligned}$$

  In other words, $\sigma(Y_1) = \big\{\emptyset, \Omega_0, \Omega_1, \Omega\big\}$.

- The σ-field $\mathcal{F}_2 \triangleq \sigma(Y_1, Y_2)$ consists of any union of the empty set and elements in the partition of Ω according to the first two entries of each outcome; i.e. $\Omega = \Omega_{00} \uplus \Omega_{01} \uplus \Omega_{10} \uplus \Omega_{11}$, where each cell in the partition is defined similarly as above; namely.

$$\begin{aligned} \Omega_{00} &\triangleq (0, 0, \cdot, \cdot, \ldots) \\ \Omega_{01} &\triangleq (0, 1, \cdot, \cdot, \ldots) \\ \Omega_{10} &\triangleq (1, 0, \cdot, \cdot, \ldots) \\ \Omega_{11} &\triangleq (1, 1, \cdot, \cdot, \ldots) \end{aligned}$$

  In other words, $\sigma(Y_1, Y_2) = \sigma(\Omega_{00}, \Omega_{01}, \Omega_{10}, \Omega_{11}) = \big\{\emptyset, \Omega_{00}, \Omega_{01}, \Omega_{10}, \Omega_{11}, \Omega_{:,0}, \Omega_{:,1}, \Omega_{0,:}, \Omega_{1,:}, \Omega\big\}$, where the colon : designates that any value is allowed at that index.

So the sequence $\mathcal{F}_1, \mathcal{F}_2, \ldots$ is a filtration.

Moreover, by defining $X_1 = Y_1, X_2 = (Y_1, Y_2), \ldots$ (or simply using the Y_n's themselves!), we also obtain an example of a sequence of RV's adapted to the filtration. $\triangle$

18.3.2 Martingales

Definition 18.3.1. Let $\{X_n\}$ be a sequence of random variables and $\{\mathcal{F}_n\}$ be a filtration. If

a) The sequence $\{X_n\}$ is adapted to the filtration $\{\mathcal{F}_n\}$.

b) Each X_n is integrable.[107]

c) $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \quad$ for all n.

Then we say that $\{X_n\}$ is a **martingale** relative to $\{\mathcal{F}_n\}$.

If, in the last definition, $=$ is replaced by $\leq$ or $\geq$, then $\{X_n\}$ is said to be a **supermartingale** or **submartingale**, respectively. $\triangle$

Proposition 18.3.1. (Properties of martingales)

a) (Multi-step expectations.) $\mathbb{E}[X_{n+k} \mid \mathcal{F}_n] = X_n \quad$ *for* $\quad n, k = 1, 2, \ldots$
(with corresponding statements for super- and sub-martingales).

[106] In the case where the bits are independent, we can find more detail on $\mathcal{F}$ and P, including a justification of the existence of P, by examining [Ash et al., 2000, Thm. 4.11.1]. TODO: But do we really need to assume independence here? The reason I did was because the options in Sec 4.11 were independence or Markov chains. It seems like this restriction should be unnecessary; consider e.g. Gaussian processes. The construction of infinite product measures in Sec 2.7 seems to be more general. So why does Sec 4.11 exist at all?

[107] Recall that we can characterize the integrability of X_n by $\mathbb{E}|X_n| < \infty$; see Cor. 7.2.1.

Proof. a) (Multi-step expectations.) We prove the statement for martingales. We want to show that $\mathbb{E}[X_{n+k} \mid \mathcal{F}_n] = X_n$ for $n, k = 1, 2, \ldots$.

To see this, note that

$$\begin{aligned}
\mathbb{E}[X_{n+2} \mid \mathcal{F}_n] &= \mathbb{E}\big[\mathbb{E}[X_{n+1} \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n\big] && \text{“Smaller } \sigma\text{-field always wins”, a.k.a. the tower property; see Theorem 17.5.1 (l).} \\
&= \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] && \text{Def martingale (at term } n+2) \\
&= X_n && \text{Def martingale (at term } n+1)
\end{aligned}$$

The general statement follows by induction.

□

Example 18.3.2. (*Simple random walk.*) Let $Y_1, Y_2, \ldots$ be independent random variables.[108] Define

$$X_n = \sum_{k=1}^{n} Y_k$$
$$\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$$

Then $\{X_n, \mathcal{F}_n\}$ is a

$$\begin{aligned}
&\text{martingale,} && \text{if } \mathbb{E}[Y_k] = 0 \\
&\text{supermartingale,} && \text{if } \mathbb{E}[Y_k] < 0 \\
&\text{submartingale,} && \text{if } \mathbb{E}[Y_k] > 0
\end{aligned}$$

for all $k = 1, 2, \ldots$. △

Proof. The first two conditions of Def. 18.3.1 follow immediately. For Def. 18.3.1 (c), we have

$$\begin{aligned}
\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[X_n + Y_{n+1} \mid \mathcal{F}_n] && \text{hypothesis} \\
&= \mathbb{E}[X_n \mid \mathcal{F}_n] + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] && \text{linearity} \\
&= \quad X_n \quad + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] && \text{“taking out what is known” [Thm. 17.5.1 (m)], since } X_n \text{ is } \mathcal{F}_n\text{-measurable} \\
& && \text{note that this is the “perfect information” case of conditional expectation (see Example 17.4.1)} \\
&= X_n + \mathbb{E}[Y_{n+1} \mid Y_1, \ldots, Y_n] && \text{hypothesis/notation} \\
&= X_n + \mathbb{E}[Y_{n+1}] && \text{independence}
\end{aligned}$$

and the RHS is equal to, less than, or greater than X_n depending on the signs of $\mathbb{E}[Y_k]$. □

Remark 18.3.1. (*When betting, there is nothing “super" about a supermartingale.*) If the increments Y_k in Example 18.3.2 are taken to be returns in a betting game, then a supermartingale corresponds to an unfavorable game. In this context, as Durrett [2010, pp.232] says, *there is nothing “super" about a supermartingale.*

△

Example 18.3.3. (*An “increasing information" process.*) Let Y be an integrable random variable and $\mathcal{F}_n$ be a filtration on $(\Omega, \mathcal{F}, P)$. Define

$$X_n \triangleq \mathbb{E}[Y \mid \mathcal{F}_n]$$

Then X_n is a martingale with respect to $\mathcal{F}_n$. △

[108] Note that random walks are sometimes defined as sums of independent *and identically distributed* random increments (e.g. see [Durrett, 2010, pp.179]), whereas here we merely refer to sums of independent random increments. Since the identically distributed assumption is not needed here to construct a martingale, we discard it here.

Proof. The first two conditions of Def. 18.3.1 follow immediately. For Def. 18.3.1 (c), note that

$$\begin{aligned}
\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\big[\mathbb{E}[Y \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n\big] && \text{def. } \{X_n\} \\
&= \mathbb{E}[Y \mid \mathcal{F}_n] && \text{“Smaller } \sigma\text{-field always wins", a.k.a. the tower property; see Theorem 17.5.1 (l).} \\
&= X_n && \text{def. } \{X_n\}
\end{aligned}$$

□

Remark 18.3.2. (*Concrete example of an increasing information process.*) In Example 18.3.3, we constructed an “increasing information process" and showed that it is a martingale. Let us now give a concrete example of an increasing information process.

We define a probability space as follows:

$$\begin{aligned}
\Omega &= [0,1]^2 && \text{(the unit square)} \\
\mathcal{F} &= \mathcal{B}([0,1]^2) \\
P &= \text{Uniform distribution}
\end{aligned}$$

To construct a filtration, we begin by constructing a sequence $\{\Omega_n\}$ of increasingly refined partitions of Ω. We set

$$\Omega_0 = \Big\{\emptyset, \Omega\Big\}$$

Ω_n as the partition of Ω formed by splitting cells of Ω_{n-1} in half, vertically if n odd and horizontally if n even

TODO: Refactor into the filtration section, so that we can more generally speak about filtrations generated by refinements of partitions.

We then define a filtration by setting $\mathcal{F}_n = \sigma(\Omega_n)$, i.e. each σ-field $\mathcal{F}_n$ consists of the sets that can be formed by taking unions of some subset of the cells in the partition Ω_n.

Now let Y be an integrable random variable on $(\Omega, \mathcal{F}, P)$, and define $X_n = \mathbb{E}[Y \mid \mathcal{F}_n]$. An example of such a sequence $\{X_n\}$ of random variables is given in Fig. 37.

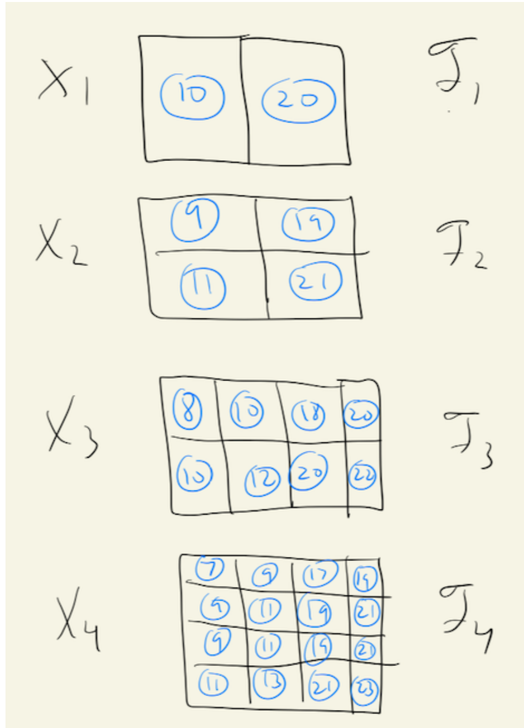


Figure 37: *An example of an increasing information process.* For the first few elements in the sequence ($n = 1, \ldots, 4$), we show the values of the random variable $X_n \triangleq \mathbb{E}[Y \mid \mathcal{F}_n]$ over each element of a partition Ω_n from which the σ-field $\mathcal{F}_n$ is generated.

Note that as n increases, X_n gives more information about the values of Y on the square Ω. (In particular, X_n gives us the average value of Y over a grid of sub-rectangles that is a refinement of the corresponding grid over which X_{n-1} gave averages.) Every time we split a cell in half, the original value a splits into two values b and c such that $a = \frac{b+c}{2}$. There are of course infinitely many possible choices for b and c given a, and we don't know what those values are until we observe the next random variable in the sequence.

Going the other way, as n decreases (so we move *up* along Fig. 37), the values of X_{n-1} are *determined* by the values of X_n. (In Fig. 37, the values in a square are determined by the values in the square immediately beneath it.) So there is no new information given. This must be true by the definition of conditional probability, which demands both (a) $\int_A \mathbb{E}[X_{n+1} \mid \mathcal{F}_n]\, dP = \int_A X_{n+1} dP$ for all $A \in \mathcal{F}_n$ and (b) $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]$ must be $\mathcal{F}_n$-measurable. In the Figure, the conditional expectation gives predictions about the cell values in the $(n+1)$-st subdivided square from the cell values in the n-th subdivided square. Criterion (b) requires that our predictions must be constant over the cell values in row n. Criterion (a) requires that our prediction for any cell must equal the average of the cell splits given in row $n+1$. For this reason, our sequence $\{X_n, \mathcal{F}_n\}$ is a martingale; that is, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$. TODO: Perhaps edit this paragraph. $\triangle$

References

Apostol, T. (1974). *Mathematical Analysis (2nd edition)*.

Ash, R. B., Robert, B., Doleans-Dade, C. A., and Catherine, A. (2000). *Probability and measure theory*. Academic Press.

Bartle, R. G. (2014). *The elements of integration and Lebesgue measure*. John Wiley & Sons.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Cheng, K. C., Miller, E. L., Hughes, M. C., and Aeron, S. (2020). On matched filtering for statistical change point detection. *IEEE Open Journal of Signal Processing*, 1:159–176.

Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.

Folland, G. B. (1999). *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons.

Makalic, E. and Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.

Matthews, A. G. d. G. (2017). *Scalable Gaussian process inference using variational methods*. PhD thesis, University of Cambridge.

Rosenthal, J. S. (2006). *First Look At Rigorous Probability Theory, A*. World Scientific Publishing Company.

Rudin, W. (1987). Real and complex analysis.

Stewart, I. and Tall, D. (2018). *Complex analysis*. Cambridge University Press.

Strichartz, R. S. (2000). *The way of analysis*. Jones & Bartlett Learning.

Zitkovic, G. (2013a). Lecture 10 in theory of probability i: Conditional expectation. `https://web.ma.utexas.edu/users/gordanz/notes/conditional_expectation.pdf`.

Zitkovic, G. (2013b). Lecture 6 in theory of probability i: Basic probability. `https://web.ma.utexas.edu/users/gordanz/notes/basic_probability.pdf`.

A Supremum and Infimum

Following are some definitions and propositions that we use in the notes.[109]

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 $a < m + \epsilon$.

△

[109]For a nice introductory overview, see `https://www.math.ucdavis.edu/~hunter/m125b/ch2.pdf`.

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. $\triangle$

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. $\triangle$

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*

$$\begin{aligned} \sup(A+B) &= \sup A + \sup B, \quad \inf(A+B) = \inf A + \inf B \\ \sup(A-B) &= \sup A - \inf B, \quad \inf(A-B) = \inf A - \sup B \end{aligned}$$

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. $\triangle$

A.3 Limits of monotone real-valued sequences

Proposition A.3.1. *Let $\{x_n\}$ be a sequence of real numbers.*

a) If $\{x_n\}$ is decreasing, then its infimum is the limit, i.e.

$$\text{If } x_n \downarrow x, \text{ then } x = \inf x_n$$

b) If $\{x_n\}$ is increasing, then its supremum is the limit, i.e.

$$\text{If } x_n \uparrow x, \text{ then } x = \sup x_n$$

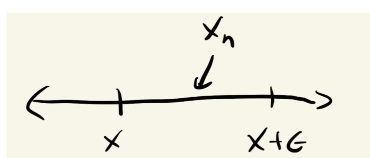
Proof. First note that if $\{x_n\}$ is a monotone sequence of real-numbers, then it always has a limit (possibly $\pm\infty$).

a) By the definition of the infimum, we need to show:

   i) x is a lower bound? ✓ .

   ii) x is the greatest lower bound (i.e. for all $\epsilon > 0$, $x + \epsilon$ is *not* a lower bound; see Remark A.1.2)?
   So we need to show

$$\forall \epsilon > 0, \exists x_n : \quad x_n < x + \epsilon \tag{A.3.1}$$



Well, note that $\lim_{n\to\infty} x_n = x$ means that

$$\forall \epsilon > 0, \exists N : \forall n \geq N, \quad |x_n - x| < \epsilon \qquad *$$

Since (x_n) is decreasing, the conclusion of $*$ becomes

$$x_n - x < \epsilon \qquad **$$

So given ϵ, choose x_N in $**$ to obtain a sequence element satisfying Eq. (A.3.1).

b) Can be argued similarly as part (a).

□

A.4 Limit inferior and limit superior of real valued sequences

Definition A.4.1. If a_n is a sequence of real numbers, then we define its limit inferior and limit superior as

$$\liminf_{n\to\infty} a_n := \lim_{n\to\infty} \inf_{m\geq n} a_m, \qquad \limsup_{n\to\infty} a_n := \lim_{n\to\infty} \sup_{m\geq n} a_m$$

△

Unlike the limit, the limit inferior and limit superior always exist, although they may be $\pm\infty$.

Remark A.4.1. (*Definitions of limit inferior and limit superior in quantified form.*) 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.
$$\begin{aligned} &\forall \epsilon > 0, \exists N \in \mathbb{N} : \forall n \geq N, \\ &x_n < S + \epsilon \end{aligned}$$

2.
$$\begin{aligned} &\forall \epsilon > 0, N \in \mathbb{N}, \exists n \geq N : \\ &x_n > S - \epsilon \end{aligned}$$

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). △

Now we give an alternate characterization of the limit superior and limit inferior

Proposition A.4.1. (Alternate characterization of the limit superior and limit inferior.) *If x_n is a sequence of real numbers, then we can express its limit inferior and limit superior as follows:*

a) $\liminf_{n\to\infty} x_n = \sup_n \inf_{k\geq n} x_k$

b) $\limsup_{n\to\infty} x_n = \inf_n \sup_{k\geq n} x_k$

Proof. a) Define $X_n := (x_k)_{k\geq n}$ as the tail of the sequence beginning at index n. Then clearly $X_{n+1} \subset X_n$. So by Proposition A.2.1,

$$\underbrace{\inf X_{n+1}}_{:= y_{n+1}} \geq \underbrace{\inf X_n}_{:= y_n}$$

Hence, (y_n) is an increasing sequence. So by Prop. A.3.1, $\lim y_n = \sup y_n$.

b) Can be argued similarly as in (a).

□

Proposition A.4.2. *a)* $\liminf_{n\to\infty} -x_n = -\limsup_{n\to\infty} x_n$

b) $\limsup_{n\to\infty} -x_n = -\liminf_{n\to\infty} x_n$

Proof. a)

$$\liminf_{n\to\infty} -x_n \overset{\textit{Prop. A.4.1}}{=} \lim_{n\to\infty} \inf_{k\geq n} -x_k \overset{\textit{Prop. A.2.2}}{=} \lim_{n\to\infty} -\sup_{k\geq n} x_k \overset{\textit{limit properties}}{=} -\lim_{n\to\infty} \sup_{k\geq n} x_k \overset{\textit{Prop. A.4.1}}{=} -\limsup_{n\to\infty} x_n.$$

b) Set $y_n = -x_n$ in (a) to obtain $\liminf_{n\to\infty} y_n = -\limsup_{n\to\infty} -y_n$. Then negate both sides of the equation.

□

Proposition A.4.3.

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

Proof.

$$\liminf_{n\to\infty} a_n \overset{\textit{def}}{=} \lim_{n\to\infty} \inf_{m\geq n} a_m \overset{*}{\leq} \lim_{n\to\infty} \sup_{m\geq n} a_m \overset{\textit{def}}{=} \limsup_{n\to\infty} a_n.$$

(Note: * holds because limits preserve non-strict inequalities.) □

A.5 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.5.1. If $f : A \to \mathbb{R}$ is a function, then

$$\sup_A f := \sup\{f(x) : x \in A\}, \qquad \inf_A f := \inf\{f(x) : x \in A\}$$

△

One useful result is that the limit of an increasing sequence of functions is its supremum.

Proposition A.5.1. *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.5.2. *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.5.1 Limit inferior and limit superior of functions

Definition A.5.2. If $\{f_n\}$ is a sequence of functions, then we define its limit inferior and limit superior as

$$\liminf_{n\to\infty} f := \lim_{n\to\infty} \inf_{m\geq n} f_m, \qquad \limsup_{n\to\infty} f := \lim_{n\to\infty} \sup_{m\geq n} f_m$$

△

Applying Proposition A.4.3 to a function pointwise, we have

$$\liminf_{n\to\infty} f \leq \limsup_{n\to\infty} f \tag{A.5.1}$$

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[110]

$$\begin{aligned} S^+(f,P) &:= \sum_{i=1}^n M_i(x_i - x_{i-1}) \\ &\quad \text{where } M_i \text{ is the supremum of } f \text{ on } (x_{i-1}, x_i]. \\ S^-(f,P) &:= \sum_{i=1}^n m_i(x_i - x_{i-1}) \\ &\quad \text{where } m_i \text{ is the infimum of } f \text{ on } (x_{i-1}, x_i]. \end{aligned}$$

- The *oscillation* Osc(f,P) of a function f over a partition P is given by the difference of the upper and lower sums; i.e. Osc$(f,P) := S^+(f,P) - S^-(f,P)$.
- The *maximum interval length* of P is defined by $\max_{i=1,\ldots,n} \big|[x_{i-1}, x_i]\big|$.

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 Complex Analysis

In the following, let $z \in \mathbb{C}$. So $z = x + iy$ for $x, y \in \mathbb{R}$.

C.1 The the complex modulus

The *modulus*, or absolute value, of $x \in \mathbb{R}$ is given by

$$|x| = \begin{cases} x, & \text{if} \quad x \geq 0 \\ -x, & \text{if} \quad x < 0 \end{cases}$$

We can define the modulus of $z \in \mathbb{C}$ by $|z| = +\sqrt{x^2 + y^2}$. We interpret $|z|$ as the Euclidean distance from the origin to z.

[110]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.

C.2 Bounding the complex modulus by real moduli

Let $z = x + iy, (x, y \in \mathbb{R})$. We can bound the complex modulus by moduli of real numbers as follows:

$$|x|, |y| \leq |z| = |x + iy| \leq |x| + |y| \tag{C.2.1}$$

where the latter inequality is the triangle inequality.

The bounds Eq. (**??**) are very useful. For instance, from them, we can quickly deduce that the problem of finding limits of sequences of complex numbers can be reduced directly to the real case.

Lemma C.2.1. *[Stewart and Tall, 2018, Lemma 3.2] Let (z_n) be a sequence of complex numbers, with $z_n = x_n + iy_n, (x_n, y_n \in \mathbb{R})$. Let $z = x + iy, (x, y \in \mathbb{R})$. Then*

$$\lim_{n\to\infty} z_n = z$$

if and only if

$$\lim_{n\to\infty} x_n = x \quad \textit{and} \quad \lim_{n\to\infty} y_n = y$$

D Miscellaneous

D.1 Right semi-closed intervals

Definition D.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 $-\infty \leq a < \infty$. △

D.2 DeMorgan's Law applies to relative complements

Remark D.2.1. DeMorgan's Law also holds for relative complements. That is, given a sequence of sets $A_1, A_2, ...$ that are subsets of another set X, we have:

$$X - \bigcap_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (X - A_n) \tag{D.2.1}$$

△

D.3 Vector space

Definition: A **vector space** consists of a set V (elements of V are called **vectors**), a field $\mathbb{F}$ (elements of $\mathbb{F}$ are called **scalars**), and two operations

- An operation called *vector addition* that takes two vectors $v, w \in V$, and produces a third vector, written $v + w \in V$.
- An operation called *scalar multiplication* that takes a scalar $c \in \mathbb{F}$ and a vector $v \in V$, and produces a new vector, written $cv \in V$.

which satisfy the following conditions (called *axioms*).

1. Associativity of vector addition: $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$.
2. Existence of a zero vector: There is a vector in V, written 0 and called the **zero vector**, which has the property that $u + 0 = u$ for all $u \in V$
3. Existence of negatives: For every $u \in V$, there is a vector in V, written $-u$ and called the **negative of** u, which has the property that $u + (-u) = 0$.
4. Associativity of multiplication: $(ab)u = a(bu)$ for any $a, b \in \mathbb{F}$ and $u \in V$.
5. Distributivity: $(a + b)u = au + bu$ and $a(u + v) = au + av$ for all $a, b \in \mathbb{F}$ and $u, v \in V$.
6. Unitarity: $1u = u$ for all $u \in V$.

Figure 38: Definition of a vector space. From `http://www.math.toronto.edu/gscott/WhatVS.pdf`.

In this document, the field $\mathbb{F}$ is always either the real numbers or the complex numbers.

D.4 Norm and semi-norm

Definition D.4.1. [Ash et al., 2000, pp.87]. A **norm** on a vector space V (over the real or complex field) is a real-valued function $||\cdot||$ on V such that for any $f, g \in V$,

1. $||f|| = 0 \implies f = 0$.
2. *(non-negativity)* $||f|| \geq 0$
3. *(scalar multiple)* $||af|| = |a|\ ||f||$for each scalar a.
4. *(triangle inequality)* $||f + g|| \leq ||f|| + ||g||$.

A **semi-norm** is a real-valued function that satisfies only conditions 2 through 4. △

D.5 Metric and psuedo-metric

A metric provides a notion of distance between elements of a vector space.

Definition D.5.1. [Ash et al., 2000, pp.87]. A **metric** on a vector space V (over the real or complex field) is a real-valued function d on pairs of elements $f, g \in V$ such that

1. $d(f, g) = 0 \iff f = g$.
2. *(non-negativity)* $d(f, g) \geq 0$.
3. *(symmetry)* $d(f, g) = d(g, f)$.

4. *(triangle inequality)* $d(f,h) \leq d(f,g) + d(g,h)$.

A **psuedometric** is a real-valued function that satisfies conditions 2 through 4 but only the $\Longleftarrow$ direction of condition 1. $\triangle$

A norm immediately induces a metric (and a semi-norm immediately induces a psuedo-metric) via the relation $d(f,g) := ||f-g||$.

D.6 Basic set theory definitions

Definition D.6.1. If $\{E_\alpha\}_{\alpha \in A}$ is an indexed collection of sets, the **union** and **intersection** are defined by

$$\bigcup_{\alpha \in A} E_\alpha := \{x : x \in E_\alpha \text{ for some } \alpha \in A\}$$
$$\bigcap_{\alpha \in A} E_\alpha := \{x : x \in E_\alpha \text{ for all } \alpha \in A\}$$

$\triangle$

Definition D.6.2. If X and Y are sets, their **Cartesian product** $X \times Y$ is the set of all ordered pairs (x,y) such that $x \in X$ and $y \in Y$. $\triangle$

For more basic set theory definitions, see Section 0.1 of [Folland, 1999].

D.7 Set theoretic definition of uniform convergence

Definition D.7.1. (Definition of uniform convergence.) Recall that $f_n \to f$ uniformly on set A means

$$\forall \epsilon < 0,\ \exists N_\epsilon :\ \forall n \geq N_\epsilon, x \in A$$
$$|f_n(x) - f(x)| < \epsilon$$

$\triangle$

Remark D.7.1. (*Set theoretic definition of uniform convergence*). By the density of the rationals, an equivalent condition to uniform convergence (Def. D.7.1) is

$$\forall k \in \mathbb{N},\ \exists N_k \in \mathbb{N} :\ \forall n \geq N_k, x \in A$$
$$|f_n(x) - f(x)| < \frac{1}{k}$$

Thus, set of points where $f_n \to f$ uniformly is given by

$$U := \underbrace{\bigcap_{k=1}^{\infty} \underbrace{\bigcap_{n=N_k}^{\infty} \left\{x : |f_n(x) - f(x)| < \frac{1}{k}\right\}}_{\text{points within a } \frac{1}{k} \text{ distance of target in tail starting at } N_k}}_{\text{...for all } k}$$

And if $f_n \to f$ uniformly on A then $A = U$. $\triangle$

D.8 Semicontinuous functions[111]

Semicontinuity is a weaker notion than continuity. Below we give the definition of [Ash et al., 2000, Appendix 2].

[111] This section follows [Ash et al., 2000, Appendix 2].

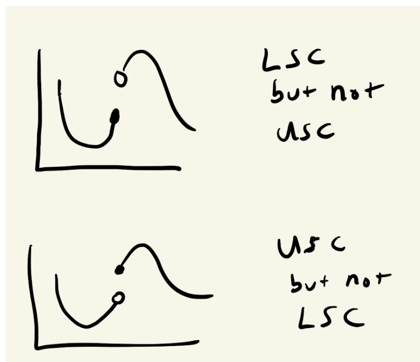
Definition D.8.1. Let Ω be a metric space. The function $\Omega \to \overline{\mathbb{R}}$ is said to be:

- **lower semi-continuous** (LSC) if $\{x \in \Omega : f(x) < a\}$ is open for all $a \in \overline{R}$.
- **upper semi-continuous** (USC) if $\{x \in \Omega : f(x) > a\}$ is open for all $a \in \overline{R}$.

$\triangle$

Remark D.8.1. An alternate characterization of semi-continuity is as follows:

- A function $f : \Omega \to \overline{\mathbb{R}}$ is LSC at $x_0 \in \Omega$ if for all $c < f(x_0)$, there is a neighborhood U of x_0 such that $f(x) > c$ for all $x \in U$.
- A function $f : \Omega \to \overline{\mathbb{R}}$ is USC at $x_0 \in \Omega$ if for all $c > f(x_0)$, there is a neighborhood U of x_0 such that $f(x) < c$ for all $x \in U$.



The alternate characterization clarifies the intuition behind the names:

- LSC functions have function values near x which are not much *lower* than $f(x)$ (so big jumps up are allowed, but not big jumps down).
- USC functions have function values near x which are not much *higher* than $f(x)$ (so big jumps down are allowed, but not big jumps up).

$\triangle$

Exercise D.8.1. Prove that the two characterizations of semicontinuity given in Def. D.8.1 and Remark D.8.1 are equivalent. $\triangle$

Now we give some useful properties of semicontinuous functions.

Proposition D.8.1. *If f is LSC and USC, then f is continuous.*

Proof. Continuous functions are characterized by the property that the inverse image of any open set A is open. Now since $A \subset \overline{\mathbb{R}}$, we can write it as a countable union of open intervals $A = \cup_{i=1}^{\infty}(a_i, b_i)$. So we have

$$f^{-1}\left(\bigcup_{i=1}^{\infty}(a_i, b_i)\right) = \bigcup_{i=1}^{\infty} f^{-1}(a_i, b_i) \qquad \text{Inverse images preserve set operations; see Sec. 6.2.4}$$
$$= \bigcup_{i=1}^{\infty} \underbrace{f^{-1}(y > a_i)}_{\text{open by LSC}} \bigcap \underbrace{f^{-1}(y < b_i)}_{\text{open by USC}}$$

which is open by the closure properties of open sets (finite intersections, countable unions) $\square$

Proposition D.8.2. *f is LSC iff $-f$ is USC.*

Proof.

$$\begin{aligned}
-f \text{ is USC} &\iff \{x : -f(x) < c\} \text{ is open } \forall c \in \overline{\mathbb{R}} && \text{def. USC}\\
&\iff \{x : f(x) > -c\} \text{ is open } \forall c \in \overline{\mathbb{R}} && \text{multiply by -1}\\
&\iff \{x : f(x) > c\} \text{ is open } \forall c \in \overline{\mathbb{R}} && \text{relabel}\\
&\iff f \text{ is LSC} && \text{def. LSC}
\end{aligned}$$

□

Now we give yet a third characterization of semicontinuous functions, which is useful when discussing weak convergence.

Theorem D.8.1. *The function $f : \Omega \to \overline{\mathbb{R}}$ is*

a) LSC iff $\liminf_{n\to\infty} f(x_n) \geq f(x)$.

b) USC iff $\limsup_{n\to\infty} f(x_n) \leq f(x)$.

for all sequences $\{x_n\} \subset \Omega$ converging to a point $x \in \Omega$.

Proof. a) • $\boxed{\implies}$. Let (1) f be LSC, (2) $x_n \to x$, (3) $c < f(x)$.

Since $x_n \to x$, eventually $x_n \in N_\epsilon(x)$ for any ϵ, where $N_\epsilon(x)$ denotes a neighborhood of radius ϵ centered on x.

Thus, eventually

$$\underbrace{f^{-1}(c,\infty]}_{\text{open}} \supset N_\epsilon(x) \ni x_n$$

where the openness holds since by (3), $x \in f^{-1}(c,\infty]$, which by (1) is an open subset of Ω.

So eventually,

$$f(x_n) \in (c,\infty] \quad \text{for any } c < f(x)$$

That is, eventually

$$f(x_n) > c \quad \text{for any } c < f(x)$$

So by characterization of the infimum (Remark A.1.2), we have

$$\liminf_{n\to\infty} f(x_n) \geq f(x)$$

• $\boxed{\impliedby}$. Suppose that (1) $x_n \to x$ implies $\liminf_{n\to\infty} f(x_n) \geq f(x)$. We need to show that $V := \{x : f(x) > c\}$ is open $\forall c \in \overline{\mathbb{R}}$. Let $x_n \to x$ for any x where (2) $f(x) > c$. Then

$$\liminf_{n\to\infty} f(x_n) \overset{(1)}{\geq} f(x) \overset{(2)}{>} c$$

So

$$f(x_n) > \underbrace{c}_{\text{eventual lower bound}} \quad \text{eventually}$$

So $x_n \in V$ eventually. So V is open.[112]

[112] This is a characterization of open sets: the tail of any sequence $\{x_n\}$ converging to $x \in V$ lies in V.

b) By Prop. D.8.2, we have that

$$f \text{ is USC} \iff -f \text{ is LSC.}$$

Therefore, by part (a), we have

$$\begin{aligned}
\liminf_{n\to\infty} -f(x_n) &\geq -f(x) & & \\
\lim_{n\to\infty} \inf_{k\geq n} -f(x_k) &\geq -f(x) & & \text{def. liminf} \\
\lim_{n\to\infty} -\sup_{k\geq n} f(x_k) &\geq -f(x) & & \text{Infimum of negated set; Prop. A.2.2} \\
-\lim_{n\to\infty} \sup_{k\geq n} f(x_k) &\geq -f(x) & & \text{Limit properties} \\
-\limsup_{n\to\infty} f(x_n) &\geq -f(x) & & \text{def. limsup} \\
\limsup_{n\to\infty} f(x_n) &\leq f(x) & & \text{multiply by -1}
\end{aligned}$$

□

D.9 Interior, boundary, and closure [113]

Definitions D.9.1. Let X be a metric space and $A \subseteq X$ a subset.

- We define the **interior** of A to be the set

$$A^\circ = \{a \in A | \text{ some } B_{r_a} \subseteq A, r_a > 0\}$$

consisting of points for which A is a "neighborhood".

- We define the **closure** of A to be the set

$$\overline{A} = \{x \in X | x = \lim_{n\to\infty} a_n, \text{ with } a_n \in A \text{ for all } n\}$$

consisting of limits of sequences in A.

△

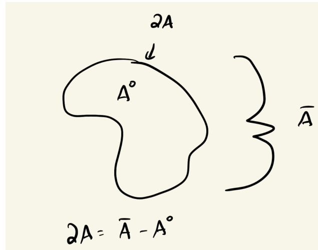
Remark D.9.1. (*Comments on interior and closure.*) In words, the interior consists of points in A for which all nearby points of X are also in A, whereas the closure allows for "points on the edge of A". Note that obviously

$$A^\circ \subseteq A \subseteq \overline{A}.$$

It can be shown that A° is the largest open set inside of A - that is, it is open and contains any open set lying inside of A (so in fact A is open if and only if $A = A^\circ$) - while $\overline{A}$ is the smallest closed set containing A, i.e. $\overline{(A)}$ is closed and lies inside of any closed set containing A (so in fact A is closed if and only if $A = \overline{A}$). △

Definition D.9.1. Let A be a subset of a metric space X. We define the **boundary** ∂A of A to be $\overline{A} - A^\circ$. △

[113] Large chunks of this section are lifted verbatim or almost verbatim from `http://math.stanford.edu/~conrad/diffgeomPage/handouts/closure.pdf`.



Example D.9.1. (*Boundary of a right semi-closed interval.*) The right semi-closed interval $I = (a, b]$ has boundary given by

$$\begin{aligned}\partial I &= \overline{I} - I^\circ \\ &= [a, b] - (a, b) \\ &= \{a\} \cup \{b\}\end{aligned}$$

$\triangle$

E Conditional Expectation: Supplemental

Below we provide another version of some properties of conditional expectation (Theorem 17.5.1), but where we use the “pairing" structure used by [Ash et al., 2000, Sec. 5.5]. I don’t think that it is necessary to use the pairing structure; other authors don’t, and there seems to be a generic way to obtain theorems about $E[Y|X = x]$ given a corresponding theorem about $E[Y \mid X]$; see [Ash et al., 2000, Sec. 5.5, Problem 3].

Theorem E.0.1. *If Y is a constant k a.e., then*

(a) (Constant property.) $\mathbb{E}[Y \mid \mathcal{G}] = k$ *a.e*

(a’) (Constant property.) $\mathbb{E}[Y \mid X = x] = k$ *a.e.* $[P_X]$

If $Y_1 \leq Y_2$ a.e., then

(b) (Monotonicity.) $\mathbb{E}[Y_1 \mid \mathcal{G}] \leq \mathbb{E}[Y_2 \mid \mathcal{G}]$ *a.e.*

(b’) (Monotonicity.) $\mathbb{E}[Y_1 \mid X = x] \leq \mathbb{E}[Y_2 \mid X = x]$ *a.e.* $[P_X]$

If $a, b \in \mathbb{R}$ and $a\mathbb{E}[Y_1] + b\mathbb{E}[Y_2]$ is well-defined (not of the form $\infty - \infty$), then

(c) (Linearity.) $\mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}] = a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}]$

(c’) (Linearity.) $\mathbb{E}[aY_1 + bY_2 \mid X = x] = a\mathbb{E}[Y_1 \mid X = x] + b\mathbb{E}[Y_2 \mid X = x]$

Finally, we have

(d) (Triangle inequality.) $|\mathbb{E}[Y \mid \mathcal{G}]| \leq \mathbb{E}[|Y| \mid \mathcal{G}]$ *a.e.*

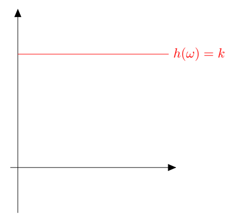
(d’) (Triangle inequality.) $|\mathbb{E}[Y \mid X = x]| \leq \mathbb{E}[|Y| \mid X = x]$ *a.e.* $[P_X]$

Proof. (a) Let $h = \mathbb{E}[Y \mid \mathcal{G}] : \Omega \to \mathbb{R}$ be defined a.e. by $h(\omega) = k$. We verify the two conditions for the existence (and essential uniqueness) of conditional expectation.

1. $h^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}(\overline{\mathbb{R}})$? ✓ . We have

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

and $\Omega, \emptyset \in \mathcal{G}$.



2. $\int_G Y \, dP = \int_G h \, dP$ for all $G \in \mathcal{G}$? ✓ . By hypothesis on Y and our guess on h, the condition becomes $\int_G k \, dP = \int_G k \, dP$, which obviously holds.

(a') Let $g = \mathbb{E}[Y \mid X = x] : \Omega' \to \mathbb{R}$ be defined a.e. by $g(x) = k$. We verify the two conditions for the existence (and essential uniqueness) of conditional expectation.

1. $g^{-1}(B) \in \mathcal{F}'$ for all $B \in \mathcal{B}(\overline{\mathbb{R}})$? ✓ . By a similar argument as in (a), $\{g^{-1}(B) : B \in \mathcal{B}(\overline{\mathbb{R}})\} = \{\Omega', \emptyset\}$.
2. $\int_{\{X \in A\}} Y \, dP = \int_A g \, dP_X$ for all $A \in \mathcal{F}'$? ✓ . By hypothesis on Y and our guess on g, the condition to be verified becomes $\int_{\{X \in A\}} k \, dP = \int_A k \, dP_X$. By integrals of indicators, LHS gives $kP\{X \in A\}$ and the RHS gives $kP_X(A)$. Recalling that $P_X(A) \triangleq \{X \in A\}$, the proof is complete.

(b) If $Y_1 \leq Y_2$ a.e., then

$$\begin{array}{rrlll} & Y_1 1_G & \leq Y_2 1_G & \text{a.e. } \forall G \in \mathcal{G} & \text{obvious} \\ \implies & \int_G Y_1 \, dP & \leq \int_G Y_2 \, dP & \forall G \in \mathcal{G} & \text{by monotonicity} \\ \implies & \int_G \mathbb{E}[Y_1 \mid \mathcal{G}] \, dP & \leq \int_G \mathbb{E}[Y_2 \mid \mathcal{G}] \, dP & \forall G \in \mathcal{G} & \text{by def. cond'l expect.} \\ \implies & \mathbb{E}[Y_1 \mid \mathcal{G}] & \leq \mathbb{E}[Y_2 \mid \mathcal{G}] & \text{a.e.} & \text{by monotonicity converse; Thm. 7.3.3} \end{array}$$

(b') The argument is almost identical to that given by (b). If $Y_1 \leq Y_2$ a.e., then

$$\begin{array}{rrlll} & Y_1 1_{\{X \in A\}} & \leq Y_2 1_{\{X \in A\}} & \text{a.e. } \forall A \in \mathcal{F}' & \text{obvious} \\ \implies & \int_{\{X \in A\}} Y_1 \, dP & \leq \int_{\{X \in A\}} Y_2 \, dP & \forall A \in \mathcal{F}' & \text{by monotonicity} \\ \implies & \int_A \mathbb{E}[Y_1 \mid X = x] \, dP_X & \leq \int_A \mathbb{E}[Y_2 \mid X = x] \, dP_X & \forall A \in \mathcal{F}' & \text{by def. cond'l expect.} \\ \implies & \mathbb{E}[Y_1 \mid X = x] & \leq \mathbb{E}[Y_2 \mid X = x] & \text{a.e.} - [P_X] & \text{by monotonicity converse; Thm. 7.3.3} \end{array}$$

(c) By the definition of conditional expectation,

$$\int_G (aY_1 + bY_2) \, dP = \int_G \mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}] \, dP \qquad \forall G \in \mathcal{G} \qquad (1)$$

On the other hand, $\forall \, G \in \mathcal{G}$, we have

$$\begin{aligned} \int_G (aY_1 + bY_2) \, dP &= a \int_G Y_1 \, dP + b \int_G Y_2 \, dP && \text{linearity} \\ &= a \int_G \mathbb{E}[Y_1 \mid \mathcal{G}] \, dP + b \int_G \mathbb{E}[Y_2 \mid \mathcal{G}] \, dP && \text{def. cond'l expect.} \\ &= \int_G \big(a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}]\big) \, dP && \text{linearity} \qquad (2) \end{aligned}$$

Combining ① and ②, we obtain

$$\int_G \mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}] \, dP \; = \int_G \left(a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}]\right) \, dP \qquad \forall G \in \mathcal{G}$$

Since the integrands are equal for all measurable sets, Cor. 7.3.1 yields the conclusion

$$\mathbb{E}[aY_1 + bY_2 \mid \mathcal{G}] \, dP \; = a\mathbb{E}[Y_1 \mid \mathcal{G}] + b\mathbb{E}[Y_2 \mid \mathcal{G}] \qquad a.e.$$

(c') Similar to (c).

(d) [114]

$$\begin{array}{rcccccl}
 & -|Y| & \leq & Y & \leq & Y & \text{obvious} \\
\Longrightarrow & \underbrace{\mathbb{E}[-|Y| \mid \mathcal{G}]}_{= -\mathbb{E}[|Y| \mid \mathcal{G}] \text{ by part (c)}} & \leq & \mathbb{E}[Y \mid \mathcal{G}] & \leq & \mathbb{E}[|Y| \mid \mathcal{G}] & \text{monotonicity (part b)} \\
\Longrightarrow & \left|\mathbb{E}[Y \mid \mathcal{G}]\right| & \leq & \mathbb{E}[|Y| \mid \mathcal{G}] & & & \text{def. absolute value}
\end{array}$$

(d') Same as (d), but replacing the conditioning set, and using (c') instead of (c).

□

[114] [Ash et al., 2000, pp.220-221] proves (d) before (c). I am not sure how this is possible.