Measure Theoretic Probability

Lecture notes

Instructor: Arijit Chakrabarty

Last updated: 6 January 2025

1 An impossibility

We start with showing (by giving an example) that a "measure" cannot be defined for all subsets, which is done in the following subsection.

1.1 The Vitali construction

For any set X, its power set is denoted throughout by 2^X.

Theorem 1.1.1. *There does not exist a function* $\lambda : 2^{[0,1)} \to [0,1]$ *such that* $\lambda([0,1)) = 1$ *and* λ *is*

1. *countably additive, that is, for disjoint* $A_1, A_2, A_3, \ldots \subset [0,1)$,
$$\lambda\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \lambda(A_n),$$
2. *and translation invariant, that is, if* $A \subset [0,1)$ *and* $x \in \mathbb{R}$ *are such that*
$$A + x := \{a + x : a \in A\} \subset [0,1),$$
then $\lambda(A + x) = \lambda(A)$.

The proof uses the **axiom of choice**, which says that for any collection $\mathcal{S}$ of non-empty sets, there exists a function
$$f : \mathcal{S} \to \bigcup_{A \in \mathcal{S}} A$$
such that $f(A) \in A$ for all $A \in \mathcal{S}$.

Proof of Theorem 1.1.1. Assume for the sake of contradiction such λ exists. For $x, y \in [0,1)$, say $x \sim y$ if $x - y \in \mathbb{Q}$. Then $\sim$ is an equivalence relation. Let $\mathcal{S}$ be the collection of equivalence classes under $\sim$, that is,

$$\mathcal{S} \subset 2^{[0,1)} \setminus \{\emptyset\},$$

$$\bigcup_{A \in \mathcal{S}} A = [0,1), \tag{1.1.1}$$

for $A, B \in \mathcal{S}$, either $A = B$ or $A \cap B = \emptyset$, and

$$\text{for all } x, y \in \mathcal{S},\ x \sim y \iff x, y \in A \text{ for some } A \in \mathcal{S}\,. \tag{1.1.2}$$

The axiom of choice implies there exists a function $f : \mathcal{S} \to [0,1)$ such that

$$f(A) \in A \text{ for all } A \in \mathcal{S}\,.$$

Define

$$V = f(\mathcal{S})\,. \tag{1.1.3}$$

Since this construction is due to Vitali, V will be later referred to as the "Vitali set". The definition of V ensures that

$$\#(V \cap A) = 1 \text{ for all } A \in \mathcal{S}\,. \tag{1.1.4}$$

For $x, y \in [0,1)$, define

$$x \oplus y = \begin{cases} x + y, & \text{if } x + y < 1, \\ x + y - 1, & \text{if } x + y \ge 1\,. \end{cases}$$

Trivially, $x \oplus y \in [0,1)$ for all $x, y \in [0,1)$. Hence for $x \in [0,1)$,

$$V \oplus x := \{v \oplus x : v \in V\} \subset [0,1)\,.$$

Our first claim is that

$$\lambda(V \oplus x) = \lambda(V),\ x \in [0,1)\,. \tag{1.1.5}$$

To see the above, write

$$V \oplus x = \big((V \cap [0, 1-x)) + x\big) \cup \big((V \cap [1-x, 1)) + x - 1\big),$$

and observe that

$$(V \cap [0, 1-x)) + x \subset [x, 1),$$
$$(V \cap [1-x, 1)) + x - 1 \subset [0, x)\,.$$

Countable additivity of λ implies

$$\begin{aligned}\lambda(V \oplus x) &= \lambda\big((V \cap [0, 1-x)) + x\big) + \lambda\big((V \cap [1-x, 1)) + x - 1\big) \\ &= \lambda\big((V \cap [0, 1-x)) \big) + \lambda\big((V \cap [1-x, 1)) \big) \\ &= \lambda(V),\end{aligned}$$

the second line being implied by translation invariance of λ and the last line again following from countable additivity; (1.1.5) thus follows.

We shall next show that

$$\bigcup_{r \in \mathbb{Q} \cap [0,1)} (V \oplus r) = [0, 1)\,. \tag{1.1.6}$$

Fix $x \in [0, 1)$. It follows from (1.1.1) that there exists $A \in \mathcal{S}$ such that $x \in A$. A consequence of (1.1.4) is that $A \cap V$ is a singleton set, say

$$A \cap V = \{v\}\,.$$

Thus x and v are in the same equivalence class which is A, and hence (1.1.2) implies $x \sim v$. In other words, $x - v \in \mathbb{Q}$. Define

$$r = \begin{cases} x - v, & \text{if } v \le x, \\ x - v + 1, & \text{if } v > x\,. \end{cases}$$

It is easy to check that $r \in \mathbb{Q} \cap [0, 1)$ and $x = v \oplus r$. Since $v \in V$, we get $x \in V \oplus r$. Thus

$$\bigcup_{r \in \mathbb{Q} \cap [0,1)} (V \oplus r) \supset [0, 1)\,.$$

The reverse inclusion being trivial, (1.1.6) follows.

We shall next prove that the left hand side of (1.1.6) is a disjoint union, that is,

$$(V \oplus r) \cap (V \oplus s) = \emptyset \text{ for distinct } r, s \in \mathbb{Q} \cap [0, 1)\,. \tag{1.1.7}$$

Fix distinct $r, s \in \mathbb{Q} \cap [0, 1)$ and if possible, let $x \in (V \oplus r) \cap (V \oplus s)$. Thus there exist $y, z \in V$ such that

$$y \oplus r = x = z \oplus s\,.$$

Since $y \oplus r$ equals either $y + r$ or $y + r - 1$ and likewise for $z \oplus s$, one of the following three must hold:

$$y + r = z + s, \tag{1.1.8}$$

$$y + r = z + s - 1, \tag{1.1.9}$$

$$\text{or, } y + r - 1 = z + s\,. \tag{1.1.10}$$

In all the above cases, $y - z \in \mathbb{Q}$, that is, $y \sim z$. By (1.1.2), there exists $A \in \mathcal{S}$ such that $y, z \in A$. Hence $y, z \in A \cap V$; (1.1.4) implies $y = z$. Since $r, s \in [0, 1)$, neither (1.1.9) nor (1.1.10) can hold because $|r - s| < 1$. Therefore, (1.1.8) holds, which shows $r = s$ and thus leads to a contradiction. This proves (1.1.7).

Countable additivity of λ in conjunction with (1.1.6) and (1.1.7) shows

$$\begin{aligned}\lambda\big([0,1)\big) &= \sum_{r \in \mathbb{Q} \cap [0,1)} \lambda(V \oplus r) \\ &= \sum_{r \in \mathbb{Q} \cap [0,1)} \lambda(V),\end{aligned}$$

(1.1.5) implying the second line. Obviously,

$$\sum_{r \in \mathbb{Q} \cap [0,1)} \lambda(V) = \begin{cases} 0, & \text{if } \lambda(V) = 0, \\ \infty, & \text{if } \lambda(V) > 0\,. \end{cases}$$

This contradicts $\lambda([0, 1)) = 1$ and thus completes the proof. $\square$

2 Measure

2.1 Definition of measure and its properties

Theorem 1.1.1 tells us that the domain of a "measure" on a set X has to be much smaller than 2^X. This calls for the following definition.

Definition. *Given a non-empty set Ω, $\mathcal{A} \subset 2^\Omega$ is a* σ-field on Ω *if $\Omega \in \mathcal{A}$, $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$, and $A_1, A_2, \ldots \in \mathcal{A}$ implies $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$. If $\mathcal{A}$ is a σ-field on Ω, then $(\Omega, \mathcal{A})$ is a* measurable space.

Now we are in a position to define measure. Before that let us fix the convention for adding and subtracting ∞. Define

$$x + \infty = \infty + x = \infty \text{ for all } x \in (-\infty, \infty]$$

and

$$x - \infty = -\infty + x = -\infty \text{ for all } x \in [-\infty, \infty)\,.$$

Neither $\infty - \infty$ nor $-\infty + \infty$ is defined. For $a_n \in \overline{\mathbb{R}} = [-\infty, \infty]$, say $a_n \to \infty$ if for all $M \in \mathbb{R}$, there exists N such that

$$a_n > M \text{ for all } n \geq N\,.$$

Similarly, $a_n \to -\infty$ if for all $M \in \mathbb{R}$, there exists N such that

$$a_n < M \text{ for all } n \geq N\,.$$

In view of the above definition, it is easy to check that for $a_1, a_2, \ldots \in [0, \infty]$, there exists $s \in [0, \infty]$ such that

$$\sum_{i=1}^{n} a_i \to s,\, n \to \infty\,.$$

For such a_n and s, we define

$$\sum_{n=1}^{\infty} a_n = s\,.$$

Definition. *Given a measurable space $(\Omega, \mathcal{A})$, a function $\mu : \mathcal{A} \to [0, \infty]$ is a* measure *if $\mu(\emptyset) = 0$ and μ is countably additive, that is,*

$$\mu(A_1 \cup A_2 \cup \ldots) = \sum_{n=1}^{\infty} \mu(A_n) \text{ for all disjoint } A_1, A_2, A_3, \ldots \in \mathcal{A}\,.$$

The tuple $(\Omega, \mathcal{A}, \mu)$ is a measure space. *We say μ is a*

- probability measure *if $\mu(\Omega) = 1$ and in this case $(\Omega, \mathcal{A}, \mu)$ is a probability space (a probability measure is usually denoted by P),*
- finite measure *if $\mu(\Omega) < \infty$ and in this case $(\Omega, \mathcal{A}, \mu)$ is a finite measure space,*
- σ-finite measure *if there exist $A_1, A_2, \ldots \in \mathcal{A}$ such that*

  $$\Omega = \bigcup_{n=1}^{\infty} A_n \text{ and } \mu(A_n) < \infty,\, n \geq 1,$$

  and in this case $(\Omega, \mathcal{A}, \mu)$ is a σ-finite measure space.

Exercise 2.1.1. *Suppose $(\Omega, \mathcal{A}, \mu)$ is a measure space. If $A \subset B$ and $A, B \in \mathcal{A}$, show that $\mu(A) \leq \mu(B)$.*

Exercise 2.1.2. *Suppose $(\Omega, \mathcal{A}, \mu)$ is a σ-finite measure space. Show that there exist disjoint $A_1, A_2, \ldots \in \mathcal{A}$ such that*

$$\Omega = \bigcup_{n=1}^{\infty} A_n \text{ and } \mu(A_n) < \infty,\, n \geq 1\,.$$

Let us give a few examples of measure spaces.

Example 2.1.1. *Let Ω be an uncountable set. Define*

$$\mathcal{A} = \{A \subset \mathcal{A} : \text{Either } A \text{ or } A^c \text{ is countable.}\}\,.$$

Clearly, $\mathcal{A}$ is a σ-field, usually called the countable-cocountable *σ-field. A set is called cocountable if its complement is countable. Fix $0 \le \alpha \le \infty$ and let*

$$\mu(A) = \begin{cases} \alpha, & \text{if } A \text{ is countable,} \\ 0, & \text{if } A \text{ is cocountable.} \end{cases}$$

Then μ is a measure on $(\Omega, \mathcal{A})$. Furthermore, μ is not σ-finite if $\alpha = \infty$.

Example 2.1.2. *Suppose $\Omega \supset \boldsymbol{C} \neq \emptyset$. Define*

$$\mu(A) = \#(A \cap \boldsymbol{C}),\ A \in 2^\Omega.$$

Then μ is a measure on $(\Omega, 2^\Omega)$, called the "counting measure" on $\boldsymbol{C}$. If $\boldsymbol{C}$ is countable, then μ is σ-finite.

The following property of a measure is known as "continuity".

Theorem 2.1.1. *Suppose $(\Omega, \mathcal{A}, \mu)$ is a measure space. For $A_1, A_2, \ldots \in \mathcal{A}$,*

- *(continuity from below) $A_n \uparrow A$ implies $\mu(A_n) \uparrow \mu(A)$,*
- *(continuity from above) $A_n \downarrow A$ and $\mu(A_1) < \infty$ imply $\mu(A_n) \downarrow \mu(A)$.*

Proof. Assume first that $A_n \uparrow A$. That $\mu(A_1) \le \mu(A_2 \le \ldots$ follows from monotonicity of μ. Denoting $A_0 = \emptyset$, it is immediate that

$$A = \bigcup_{n=1}^\infty (A_n \setminus A_{n-1}),$$

and that the sets on the right hand side are disjoint. Thus,

$$\begin{aligned} \mu(A) &= \sum_{n=1}^\infty \mu(A_n \setminus A_{n-1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i \setminus A_{i-1}) \\ &= \lim_{n\to\infty} \mu(A_n), \end{aligned}$$

the last line following from the observation that

$$A_n = \bigcup_{i=1}^n (A_i \setminus A_{i-1}).$$

Thus $\mu(A_n) \uparrow \mu(A)$, that is, continuity from below follows.

Now suppose that $A_n \downarrow A$ and $\mu(A_1) < \infty$. Once again, $\mu(A_1) \ge \mu(A_2) \ge \ldots$ follows from monotonicity. Clearly, $A_n^c \uparrow A^c$ and hence $(A_1 \setminus A_n) \uparrow (A_1 \setminus A)$. Continuity of μ from below implies that

$$\mu(A_1 \setminus A_n) \uparrow \mu(A_1 \setminus A).$$

Since $\mu(A_1) < \infty$, the above is the same as

$$\mu(A_1) - \mu(A_n) \uparrow \mu(A_1) - \mu(A)\,.$$

Thus $\mu(A_n) \downarrow \mu(A)$. This completes the proof. $\square$

Exercise 2.1.3. *If $\Omega = \mathbb{N}$ and μ is the counting measure on $\mathbb{N}$, show that*

$$\{n, n+1, \ldots\} \downarrow \emptyset$$

and

$$\mu\left(\{n, n+1, \ldots\}\right) \not\to 0,\ n \to \infty\,.$$

Definition. *For $\Omega \neq \emptyset$ and $\mathcal{G} \subset 2^\Omega$, the* σ-field generated by $\mathcal{G}$, *denoted by $\sigma(\mathcal{G})$, is defined as*

$$\sigma(\mathcal{G}) = \bigcap_{\mathcal{A}:\, \mathcal{G} \subset \mathcal{A} \subset 2^\Omega,\, \mathcal{A} \text{ is a } \sigma\text{-field}} \mathcal{A}\,.$$

In other words, $\sigma(\mathcal{G})$ is the intersection of all σ-fields on Ω containing $\mathcal{G}$.

Exercise 2.1.4. *If Ω and $\mathcal{G}$ are as above, show that $\sigma(\mathcal{G})$ is the smallest σ-field containing $\mathcal{G}$, that is,*

1. *$\mathcal{G} \subset \sigma(\mathcal{G})$,*
2. *$\sigma(\mathcal{G})$ is a σ-field,*
3. *and if $\mathcal{A}$ is any σ-field on Ω with $\mathcal{A} \supset \mathcal{G}$, then $\sigma(\mathcal{G}) \subset \mathcal{A}$.*

Definition. *The σ-field on $\mathbb{R}$ generated by the collection of all open subsets of $\mathbb{R}$ is called the* Borel σ-field on $\mathbb{R}$ *and is denoted by $\mathcal{B}(\mathbb{R})$. Elements of $\mathcal{B}(\mathbb{R})$ are called* Borel sets.

Exercise 2.1.5. *Show that $A \subset \mathbb{R}$ is a Borel set if A is*

1. *an open set*
2. *a closed set*
3. *a countable set*
4. *or an interval, that is, A is one of (a, b), $(a, b]$, $[a, b)$ or $[a, b]$ for some $-\infty < a \leq b < \infty$.*

2.2 The Carathéodory extension theorem

For constructing useful measures, the usual method is to first define a "countably additive set function" on a "field", and then use an extension theorem to extend it to the generated σ-field. This is done with the help of the Carathéodory extension theorem, stated and proved below. This result is therefore of fundamental importance in measure theory.

Definition. *For $\Omega \neq \emptyset$, $\mathcal{F} \subset 2^\Omega$ is a* field *if $\emptyset \in \mathcal{F}$, $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$ and $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$.*

Exercise 2.2.1. *Show that $\mathcal{F}$ is a field on Ω if and only if $\Omega \in \mathcal{F}$, $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements.*

Exercise 2.2.2. *If $\mathcal{F}$ is a field, show that for $A_1, \ldots, A_n \in \mathcal{F}$,*

$$A_1 \cup \ldots \cup A_n \in \mathcal{F} \text{ and } A_1 \cap \ldots \cap A_n \in \mathcal{F}\,.$$

Definition. *Given a field $\mathcal{F}$ on $\Omega \neq \emptyset$, a function $\mu : \mathcal{F} \to [0, \infty]$ is a* countably additive set function *if $\mu(\emptyset) = 0$ and*

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n)$$

whenever $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint and $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$.

Theorem 2.2.1 (Carathéodory extension theorem)**.** *If $\mathcal{F}$ is a field on $\Omega \neq \emptyset$ and μ is a countably additive set function on $\mathcal{F}$, then there exists a measure μ^* on $\big(\Omega, \sigma(\mathcal{F})\big)$ such that*

$$\mu^*(A) = \mu(A) \text{ for all } A \in \mathcal{F}\,.$$

Let μ be a countably additive set function on $\mathcal{F}$. We start with defining μ^* as the "outer measure" of μ as follows:

$$\mu^*(E) = \inf\left\{\sum_{n=1}^\infty \mu(A_n) : E \subset \bigcup_{n=1}^\infty A_n \text{ and } A_1, A_2, \ldots \in \mathcal{F}\right\},\ E \in 2^\Omega\,. \tag{2.2.1}$$

Lemma 2.2.1. *For all $A \in \mathcal{F}$, $\mu^*(A) = \mu(A)$.*

Proof. Fix $A \in \mathcal{F}$. Letting $A_1 = A$ and $A_2 = A_3 = \ldots = \emptyset$, it is immediate that

$$\mu^*(A) \leq \sum_{n=1}^\infty \mu(A_n) = \mu(A)\,.$$

For the reverse inequality, suppose that

$$A \subset \bigcup_{n=1}^\infty A_n \text{ for some } A_1, A_2, \ldots \in \mathcal{F}\,. \tag{2.2.2}$$

Define $B_1 = A \cap A_1$, $B_2 = A \cap A_2 \cap A_1^c$, $B_3 = A \cap A_3 \cap (A_1 \cup A_2)^c$ and in general

$$B_n = A \cap A_n \cap \left(\bigcup_{i=1}^{n-1} A_i \right)^c \text{ for all } n \geq 1\,.$$

Since $\mathcal{F}$ is a field, $B_1, B_2, \ldots \in \mathcal{F}$. Further, (2.2.2) shows

$$B_1 \cup B_2 \cup \ldots = A\,.$$

Obviously, $B_1, B_2, \ldots$ are disjoint. Countable additivity of μ implies that

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n),$$

the inequality following from the trivial fact that any countably additive set function is necessarily monotone. Since this holds for any $A_1, A_2, \ldots$ satisfying (2.2.2), we get

$$\mu(A) \leq \mu^*(A)\,.$$

This along with the already proven reverse inequality completes the proof. $\square$

Lemma 2.2.2. *For all $A_1, A_2, \ldots \in 2^{\Omega}$,*

$$\mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)\,.$$

In other words, μ^* is countably subadditive.

Proof. Let $A_1, A_2, \ldots \in 2^{\Omega}$. The claim would follow if it can be shown that for all $\varepsilon > 0$,

$$\mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \varepsilon + \sum_{n=1}^{\infty} \mu^*(A_n)\,. \tag{2.2.3}$$

Fix $\varepsilon > 0$. By definition of μ^*, there exists $A_{n1}, A_{n2}, \ldots \in \mathcal{F}$ such that

$$A_n \subset A_{n1} \cup A_{n2} \cup \ldots \text{ and } \sum_{i=1}^{\infty} \mu(A_{ni}) \leq \mu^*(A_n) + 2^{-n}\varepsilon\,.$$

Thus

$$\bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} \bigcup_{i=1}^{\infty} A_{ni}\,.$$

Hence

$$\mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \mu(A_{ni}) \leq \sum_{n=1}^{\infty} \left(\mu^*(A_n) + 2^{-n}\varepsilon \right) = \varepsilon + \sum_{n=1}^{\infty} \mu^*(A_n)\,.$$

This shows (2.2.3) from which the proof follows. $\square$

Lemma 2.2.3. *Define*

$$\mathcal{A} = \{A \in 2^\Omega : \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \in 2^\Omega\}. \quad (2.2.4)$$

Then $\mathcal{A} \supset \mathcal{F}$.

Proof. Lemma 2.2.2 shows that

$$\mu^*(E) \le \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E, A \in 2^\Omega.$$

Hence (2.2.4) becomes

$$\mathcal{A} = \{A \in 2^\Omega : \mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \in 2^\Omega\}. \quad (2.2.5)$$

Let $A \in \mathcal{F}$ and $E \in 2^\Omega$. Fix $\varepsilon > 0$ and let $A_1, A_2, \ldots \in \mathcal{F}$ be such that

$$E \subset \bigcup_{n=1}^\infty A_n \text{ and } \mu^*(E) + \varepsilon \ge \sum_{n=1}^\infty \mu(A_n).$$

The definition of μ^* implies

$$\mu^*(E \cap A) \le \sum_{n=1}^\infty \mu(A_n \cap A),$$

and

$$\mu^*(E \cap A^c) \le \sum_{n=1}^\infty \mu(A_n \cap A^c).$$

Combining the two yields

$$\begin{aligned}\mu^*(E \cap A) + \mu^*(E \cap A^c) &\le \sum_{n=1}^\infty \mu(A_n \cap A) + \sum_{n=1}^\infty \mu(A_n \cap A^c) \\ (\text{countable additivity of } \mu) &= \sum_{n=1}^\infty \mu(A_n) \\ &\le \varepsilon + \mu^*(E).\end{aligned}$$

Since ε is arbitrary, it follows that

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) \le \mu^*(E).$$

As this holds for all $E \in 2^\Omega$, (2.2.5) shows $A \in \mathcal{A}$ and hence completes the proof. □

Lemma 2.2.4. *The collection* $\mathcal{A}$*, as in* (2.2.4)*, is a field and*

$$\mu^*\left(E \cap \left(\bigcup_{n=1}^\infty A_n\right)\right) = \sum_{n=1}^\infty \mu^*(E \cap A_n) \text{ for disjoint } A_1, A_2, \ldots \in \mathcal{A},\ E \in 2^\Omega.$$

Proof. The definition (2.2.4) shows $\Omega \in \mathcal{A}$ and that $\mathcal{A}$ is closed under complements. Suppose $A, B \in \mathcal{A}$ and $E \in 2^{\Omega}$. Define

$$F = E \cap (A \cup B)\,.$$

Since $A \in \mathcal{A}$,

$$\mu^*(F) = \mu^*(F \cap A) + \mu^*(F \cap A^c) = \mu^*(F \cap A) + \mu^*(E \cap B \cap A^c)\,.$$

Since $B \in \mathcal{A}$ and $F \cap A \in 2^{\Omega}$,

$$\mu^*(F \cap A) = \mu^*((F \cap A) \cap B) + \mu^*((F \cap A) \cap B^c) = \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c),$$

which implies

$$\mu^*\left(E \cap (A \cup B)\right) = \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap B \cap A^c)\,. \tag{2.2.6}$$

Once again, $A \in \mathcal{A}$ implies

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\
(\text{as } B \in \mathcal{A}) &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) \\
&\quad + \mu^*(E \cap A^c \cap B^c) \\
(\text{by } (2.2.6)) &= \mu^*\left(E \cap (A \cup B)\right) + \mu^*(E \cap A^c \cap B^c) \\
&= \mu^*\left(E \cap (A \cup B)\right) + \mu^*\left(E \cap (A \cup B)^c\right)\,.
\end{aligned}$$

Since this holds for all E, $A \cup B \in \mathcal{A}$, showing that $\mathcal{A}$ is a field.

Fix $E \in 2^{\Omega}$ and disjoint $A_1, A_2, \ldots \in \mathcal{A}$. Let

$$F = E \cap (A_1 \cup A_2 \cup \ldots)\,.$$

Then

$$\begin{aligned}
\mu^*(F) &= \mu^*(F \cap A_1) + \mu^*(F \cap A_1^c) \\
&= \mu^*(E \cap A_1) + \mu^*\left(E \cap (A_2 \cup A_3 \cup \ldots)\right)\,.
\end{aligned}$$

Proceeding inductively, it can be shown that for all $n = 1, 2, \ldots$,

$$\mu^*(F) = \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*\left(E \cap (A_{n+1} \cup \ldots)\right) \geq \sum_{i=1}^{n} \mu^*(E \cap A_i)\,.$$

Since this is true for all n, we get

$$\mu^*\left(E \cap (A_1 \cup A_2 \cup \ldots)\right) \geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i)\,.$$

Lemma 2.2.2 yields the reverse inequality and thus completes the proof. $\square$

Lemma 2.2.5. *The collection $\mathcal{A}$ is a σ-field and μ^* is a measure on $(\Omega, \mathcal{A})$.*

Proof. Since $\mathcal{A}$ has already been shown to be a field in Lemma 2.2.4, to show $\mathcal{A}$ is a σ-field it suffices to prove it is closed under disjoint countable union. Let $A_1, A_2, \ldots \in \mathcal{A}$ be disjoint. Let

$$A = A_1 \cup A_2 \cup \ldots \text{ and } F_n = A_1 \cup \ldots \cup A_n,\ n \geq 1\,.$$

Since $F_n \in \mathcal{A}$ by Lemma 2.2.4, for any $E \in 2^\Omega$,

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap F_n) + \mu^*(E \cap F_n^c) \\ \text{(by Lemma 2.2.4)} &= \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \cap F_n^c) \\ &\geq \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \cap A^c), \end{aligned}$$

the last line following from the observation that $F_n^c \subset A^c$ and the fact that μ^* is monotone, that is,

$$\mu^*(A) \leq \mu^*(B) \text{ if } A \subset B \subset \Omega,$$

which follows from the very definition of μ^*. Let $n \to \infty$ to get

$$\mu^*(E) \geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap A^c) = \mu^*(E \cap A) + \mu^*(E \cap A^c),$$

the second equality following once again from Lemma 2.2.4. Thus $A \in \mathcal{A}$ by (2.2.5). Hence $\mathcal{A}$ is a σ-field. Taking $E = \Omega$ in Lemma 2.2.4 shows μ^* is a measure. This completes the proof. □

Proof of Theorem 2.2.1. Lemmas 2.2.1, 2.2.3 and 2.2.5 complete the proof. □

2.3 Uniqueness of the extension

After Theorem 2.2.1, the most pertinent question which arises subsequently is whether the extension is unique. The following theorem answers this.

Theorem 2.3.1 (Uniqueness)**.** *If μ and ν are finite measures on $(\Omega, \sigma(\mathcal{F}))$, where $\mathcal{F}$ is a field, and*

$$\mu(A) = \nu(A) \textit{ for all } A \in \mathcal{F},$$

then μ, ν agree on $\sigma(\mathcal{F})$.

The proof uses the following definition and the "monotone class theorem" which will be proved a moment later.

Definition. *For $\Omega \neq \emptyset$, $\mathcal{M} \subset 2^\Omega$ is a* monotone class *if*

$$A_n \in \mathcal{M} \text{ and } A_n \uparrow A \text{ imply } A \in \mathcal{M},$$

and

$$A_n \in \mathcal{M} \text{ and } A_n \downarrow A \text{ imply } A \in \mathcal{M}\,.$$

Theorem 2.3.2 (Monotone class theorem)**.** *If $\mathcal{F}$ is a field, $\mathcal{F} \subset \mathcal{M}$ and $\mathcal{M}$ is a monotone class, then*

$$\sigma(\mathcal{F}) \subset \mathcal{M}\,.$$

Although 'MCT' is an abbreviation for the above theorem, the same is being saved for the monotone convergence theorem to be stated and proved later.

Proof of Theorem 2.3.1. Define

$$\mathcal{G} = \{A \in \sigma(\mathcal{F}) : \mu(A) = \nu(A)\}\,.$$

The hypothesis implies $\mathcal{F} \subset \mathcal{G}$. Finiteness of $\mu(\Omega)$ and $\nu(\Omega)$ in conjunction with Theorem 2.1.1 implies that

$$\mu(A_n) \to \mu(A) \text{ and } \nu(A_n) \to \nu(A)$$

if either $A_n \uparrow A$ or $A_n \downarrow A$. Thus $A \in \mathcal{G}$ if $A_n \in \mathcal{G}$ and either $A_n \uparrow A$ or $A_n \downarrow A$. In other words, $\mathcal{G}$ is a monotone class. Theorem 2.3.2 shows $\mathcal{G} \supset \sigma(\mathcal{F})$, from which the proof follows. □

Remark 2.3.1. *The technique employed in the proof of Theorem 2.3.1 is a variant of the so-called "good set principle" and is ubiquitous in measure theory, which is the following. For showing that a property holds for every set in a σ-field generated by a collection $\mathcal{H}$, we first define a set to be "good" if it has the said property and then show the following.*

- *The collection of all good sets, say $\mathcal{G}$, is a σ-field*
- *and that $\mathcal{H} \subset \mathcal{G}$.*

The above would ensure from the definition of a σ-field generated by $\mathcal{H}$ that

$$\sigma(\mathcal{H}) \subset \mathcal{G}$$

which is tautologically the same as that the desired property holds for every set in $\sigma(\mathcal{H})$. The tools needed for showing $\mathcal{G}$ to be a σ-field vary with context. In the proof of Theorem 2.3.1, for example, the monotone class theorem was used to essentially show that $\mathcal{G}$ is a σ-field. In other situations, $\mathcal{G}$ is shown to be a σ-field from first principles.

The following stronger version of Theorem 2.3.1 can be proven along similar lines, whose proof is left as an exercise.

Theorem 2.3.3. *Suppose $(\Omega, \mathcal{A})$ is a measurable space on which, μ_1, μ_2 are measures. Assume $\mathcal{F} \subset \mathcal{A}$ is a field. If μ_1, μ_2 are σ-finite on $\mathcal{F}$, that is, there exist $A_1, A_2, \ldots \in F$ such that*

$$\bigcup_{n=1}^{\infty} A_n = \Omega \text{ and } \mu_i(A_n) < \infty \text{ for all } i = 1, 2,\ n = 1, 2, \ldots,$$

and

$$\mu_1(A) = \mu_2(A) \text{ for all } A \in \mathcal{F},$$

then

$$\mu_1(A) = \mu_2(A) \text{ for all } A \in \sigma(\mathcal{F})\,.$$

Proof of Theorem 2.3.2. Let $\mathcal{M}_0$ be the intersection of all monotone classes containing $\mathcal{F}$, that is, the smallest monotone class containing $\mathcal{F}$. Clearly, it suffices to show that

$$\sigma(\mathcal{F}) \subset \mathcal{M}_0\,.$$

To achieve that end, we shall show that $\mathcal{M}_0$ is a σ-field, which will be done via the following steps.
Step 1. If $A \in \mathcal{M}_0$, then $A^c \in \mathcal{M}_0$.

Proof of Step 1. Define

$$\mathcal{G} := \{A \in \mathcal{M}_0 : A^c \in \mathcal{M}_0\}\,.$$

Notice that $\mathcal{F} \subset \mathcal{G}$ because $\mathcal{F}$ is a field. Next observe that if $A_1, A_2, \ldots \in \mathcal{G}$ and $A_n \uparrow A$, then $A \in \mathcal{M}_0$ because $\mathcal{M}_0$ is closed under monotone unions. Further, $A_n^c \in \mathcal{M}_0$ and $A_n^c \downarrow A^c$. As $\mathcal{M}_0$ is closed under monotone intersection, $A^c \in \mathcal{M}_0$. Therefore, $A \in \mathcal{G}$, showing that $\mathcal{G}$ is closed under monotone union. Similarly, it can be shown that $\mathcal{G}$ is closed under monotone intersection. Thus, $\mathcal{G}$ is a monotone class containing $\mathcal{F}$. Hence, $\mathcal{G} \supset \mathcal{M}_0$, thereby proving Step 1. □

Step 2. If $A \in \mathcal{F}$ and $B \in \mathcal{M}_0$, then $A \cup B \in \mathcal{M}_0$.

Proof. Define

$$\mathcal{H} := \{B \in \mathcal{M}_0 : A \cup B \in \mathcal{M}_0 \text{ for all } A \in \mathcal{F}\}\,.$$

By virtue of being a field, $\mathcal{F} \subset \mathcal{H}$. Routine verification will ensure that $\mathcal{H}$ is a monotone class, and hence contains $\mathcal{M}_0$. This proves Step 2. □

Step 3. If $A, B \in \mathcal{M}_0$, then $A \cup B \in \mathcal{M}_0$.

Proof. Define

$$\mathcal{I} := \{B \in \mathcal{M}_0 : A \cup B \in \mathcal{M}_0 \text{ for all } A \in \mathcal{M}_0\}.$$

By Step 2, it follows that $\mathcal{F} \subset \mathcal{I}$. Once again, similar ideas will ensure that $\mathcal{I}$ is a monotone class, and thus prove Step 3. ☐

Steps 1 and 3, along with the fact that $\mathcal{M}_0$ is a monotone class establishes that it is a σ-field, and thus completes the proof of the monotone class theorem. ☐

The statement of Theorem 2.3.3 should not be misinterpreted as the following: if μ, ν are σ-finite measures on $(\Omega, \sigma(\mathcal{F}))$, where $\mathcal{F}$ is a field, which agree on $\mathcal{F}$, then they agree on $\sigma(\mathcal{F})$. As shown by the following example, the claim just made is false.

Example 2.3.1. *Define measures μ and ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by*

$$\begin{aligned}\mu(A) &= \#(A \cap \mathbb{Q}),\\ \nu(A) &= 2\#(A \cap \mathbb{Q}),\end{aligned}$$

for all $A \in \mathcal{B}(\mathbb{R})$. Let

$$\mathcal{F} = \Big\{\mathbb{R} \cap ((a_1, b_1] \cup \ldots \cup (a_n, b_n]) : -\infty \le a_1 < b_1 < \ldots < a_n < b_n \le \infty,$$
$$n = 0, 1, 2, \ldots\Big\}.$$

Then $\mathcal{F}$ is a field and $\sigma(\mathcal{F}) = \mathcal{B}(\mathbb{R})$. Further, μ and ν agree on $\mathcal{F}$, are σ-finite on $\mathcal{B}(\mathbb{R})$, and do not agree on $\mathcal{B}(\mathbb{R})$.

2.4 Extension from semi-field to field

In practice, it is usually easier to define a set function on a "semi-field" and then extend it to the generated field.

Definition. *For $\Omega \neq \emptyset$, $\mathcal{S} \subset 2^\Omega$ is a* semi-field *on Ω if $\Omega \in \mathcal{S}$,*

$$A, B \in \mathcal{S} \text{ implies } A \cap B \in \mathcal{S},$$

and for all $A \in \mathcal{S}$,

$$A^c = A_1 \cup \ldots \cup A_n,$$

for some disjoint $A_1, \ldots, A_n \in \mathcal{S}$.

Example 2.4.1. *Let*

$$\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}.$$

Then $\mathcal{S}$ is a semi-field on $\mathbb{R}$. It can be checked that $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$.

Theorem 2.4.1. *If $\mathcal{S}$ is a semi-field on $\Omega \neq \emptyset$, then*

$$\mathcal{F} = \{A_1 \cup \ldots \cup A_n : A_1, \ldots, A_n \in \mathcal{S} \text{ are disjoint, } n = 0, 1, 2, \ldots\}$$

is the smallest field containing $\mathcal{S}$.

Proof. Trivially, any field containing $\mathcal{S}$ contains $\mathcal{F}$. Thus, all that has to be shown is $\mathcal{F}$ is a field. Since $\mathcal{S} \subset \mathcal{F}$, $\Omega \in \mathcal{F}$ is automatic.

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

$$A = A_1 \cup \ldots \cup A_m \text{ and } B = B_1 \cup \ldots \cup B_n$$

where $A_1, \ldots, A_m \in \mathcal{S}$ are disjoint and so are $B_1, \ldots, B_n$. Thus

$$A \cap B = \bigcup_{i=1}^{m} \bigcup_{j=1}^{n} (A_i \cap B_j)$$

where $\{(A_i \cap B_j) : i = 1, \ldots, m,\ j = 1, \ldots, n\}$ is a collection of disjoint sets. Since $\mathcal{S}$ is a semi-field,

$$A_i \cap B_j \in \mathcal{S},\ i = 1, \ldots, m,\ j = 1, \ldots, n\,.$$

Hence $A \cap B \in \mathcal{F}$. That is, $\mathcal{F}$ is closed under finite intersections.

To show that $\mathcal{F}$ is closed under complements, let $A \in \mathcal{F}$. Then

$$A = A_1 \cup \ldots \cup A_n$$

for some disjoint $A_1, \ldots, A_n \in \mathcal{S}$. Since $A_i \in \mathcal{S}$ for $i = 1, \ldots, n$,

$$A_i^c = \bigcup_{j=1}^{k_i} B_{ij} \text{ for some disjoint } B_{i1}, \ldots, B_{ik_i} \in \mathcal{S}\,.$$

Thus

$$A^c = A_1^c \cap \ldots \cap A_n^c = \bigcup_{j_1=1}^{k_1} \ldots \bigcup_{j_n=1}^{k_n} \left(B_{1j_1} \cap \ldots \cap B_{nj_n}\right)\,. \tag{2.4.1}$$

Clearly,

$$B_{1j_1} \cap \ldots \cap B_{nj_n} \in \mathcal{S},\ (j_1, \ldots, j_n) \in \{1, \ldots, k_1\} \times \ldots \times \{1, \ldots, k_n\},$$

and

$$(B_{1i_1} \cap \ldots \cap B_{ni_n}) \cap (B_{1j_1} \cap \ldots \cap B_{nj_n}) = \emptyset \text{ if } (i_1, \ldots, i_n) \neq (j_1, \ldots, j_n)\,.$$

Hence (2.4.1) shows $A^c \in \mathcal{F}$. Exercise 2.2.1 shows that $\mathcal{F}$ is a field, and hence the proof follows. □

Definition. *Suppose $\mathcal{S}$ is a semi-field on $\Omega \neq \emptyset$. A function $\mu : \mathcal{S} \to [0,\infty]$ is a* finitely additive set function *if $\mu(\emptyset) = 0$ and*

$$\mu(A_1 \cup \ldots \cup A_n) = \mu(A_1) + \ldots + \mu(A_n)$$

whenever $A_1, \ldots, A_n \in \mathcal{S}$ are disjoint such that $A_1 \cup \ldots \cup A_n \in \mathcal{S}$. If in addition,

$$\mu(A_1 \cup A_2 \cup \ldots) = \sum_{n=1}^{\infty} \mu(A_n)$$

whenever $A_1, A_2, \ldots \in \mathcal{S}$ are disjoint such that $A_1 \cup A_2 \cup \ldots \in \mathcal{S}$, then μ is a countably additive set function.

Theorem 2.4.2. *Suppose $\Omega \neq \emptyset$, $\mathcal{S}$ is a semi-field on Ω and μ is a finitely additive set function on $\mathcal{S}$. Let $\mathcal{F}$ be the field generated by $\mathcal{S}$. Define $\mu : \mathcal{F} \to [0,\infty]$ by*

$$\mu(A) = \sum_{i=1}^{k} \mu(A_i), \text{ if } A = A_1 \cup \ldots \cup A_k \text{ for disjoint } A_1, \ldots, A_k \in \mathcal{S}\,. \tag{2.4.2}$$

Then μ is well defined on $\mathcal{F}$, that is, different representations of A lead to the same definition of $\mu(A)$, definition of μ remains unchanged on $\mathcal{S}$ and μ is finitely additive on $\mathcal{F}$. If in addition, μ is countably additive on $\mathcal{S}$, then so it is on $\mathcal{F}$.

Proof. Assume μ is finitely additive on $\mathcal{S}$. The first step is to show that the right hand side of (2.4.2) remains invariant under the choice of $A_1, \ldots, A_k$. In other words, if for some $A \in \mathcal{F}$,

$$A_1 \cup \ldots \cup A_m = A = B_1 \cup \ldots \cup B_n,$$

where $A_1, \ldots, A_m$ are disjoint sets in $\mathcal{S}$ and so are $B_1, \ldots, B_n$, then

$$\sum_{i=1}^{m} \mu(A_i) = \sum_{j=1}^{n} \mu(B_j)\,. \tag{2.4.3}$$

For a fixed $i = 1, \ldots, m$,

$$A_i = A_i \cap A = A_i \cap (B_1 \cup \ldots \cup B_n) = \bigcup_{j=1}^{n} (A_i \cap B_j)\,.$$

Since $B_1, \ldots, B_n$ are disjoint, so are $A_i \cap B_1, \ldots, A_i \cap B_n$. Finite additivity of μ shows

$$\mu(A_i) = \sum_{j=1}^{n} \mu(A_i \cap B_j),$$

and hence

$$\sum_{i=1}^{m} \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mu(A_i \cap B_j) .$$

A similar argument shows that

$$\sum_{j=1}^{n} \mu(B_j) = \sum_{j=1}^{n} \sum_{i=1}^{m} \mu(A_i \cap B_j) ;$$

(2.4.3) follows by comparing the above two equalities. In other words, μ is well defined by (2.4.2). A trivial consequence is that (2.4.2) keeps μ unchanged on $\mathcal{S}$.

Finite additivity of μ on $\mathcal{F}$ would follow if it shown that

$$\mu(A \cup B) = \mu(A) + \mu(B) \text{ for disjoint } A, B \in \mathcal{F} .$$

Fix $A, B \in \mathcal{F}$ disjoint. Then

$$A = A_1 \cup \ldots \cup A_m \text{ and } B = B_1 \cup \ldots \cup B_n$$

for disjoint $A_1, \ldots, A_m, B_1, \ldots, B_n \in \mathcal{S}$. It follows from (2.4.2) that

$$\mu(A \cup B) = \sum_{i=1}^{m} \mu(A_i) + \sum_{j=1}^{n} \mu(B_j) = \mu(A) + \mu(B),$$

establishing μ is finitely additive on $\mathcal{F}$.

For the final claim, suppose μ is countably additive on $\mathcal{S}$. Suppose that $A_1, A_2, \ldots \in \mathcal{F}$ and

$$\bigcup_{n=1}^{\infty} A_n = A_\infty \in \mathcal{F} .$$

Then for $n = 1, 2, \ldots, \infty$, there exist disjoint $B_{n1}, \ldots, B_{nk_n} \in \mathcal{S}$ such that

$$A_n = B_{n1} \cup \ldots \cup B_{nk_n} ,$$

where $k_1, k_2, \ldots, k_\infty \in \mathbb{N}$. Thus

$$\mu(A_\infty) = \sum_{i=1}^{k_\infty} \mu\left(B_{\infty i}\right) . \tag{2.4.4}$$

For $i = 1, \ldots, k_\infty$,

$$\begin{aligned} B_{\infty i} &= B_{\infty i} \cap A_\infty \\ &= B_{\infty i} \cap \left(\bigcup_{n=1}^{\infty} \bigcup_{j=1}^{k_n} B_{nj} \right) \\ &= \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{k_n} \left(B_{\infty i} \cap B_{nj}\right) . \end{aligned}$$

Since $B_{\infty i} \cap B_{nj} \in S$ for all $n = 1, 2, \ldots$ and $j = 1, \ldots, k_n$, and the right hand side above is a disjoint union, countable additivity of μ on $\mathcal{S}$ shows that

$$\mu(B_{\infty i}) = \sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \mu(B_{\infty i} \cap B_{nj}) .$$

Invoking (2.4.4), it follows that

$$\begin{aligned}
\mu(A_\infty) &= \sum_{i=1}^{k_\infty} \sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \mu(B_{\infty i} \cap B_{nj}) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k_\infty} \sum_{j=1}^{k_n} \mu(B_{\infty i} \cap B_{nj}) \\
&= \sum_{n=1}^{\infty} \mu(A_n),
\end{aligned}$$

where the last line again uses (2.4.2) along with the observation that

$$A_n = A_\infty \cap A_n = \bigcup_{i=1}^{k_\infty} \bigcup_{j=1}^{k_n} (B_{\infty i} \cap B_{nj}),$$

and that the union on the extreme right hand side is disjoint. This shows μ is countably additive on $\mathcal{F}$ and hence completes the proof. $\square$

The following is a trivial consequence of Theorems 2.2.1 and 2.4.2.

Corollary 2.4.1. *If $\Omega \neq \emptyset$, $\mathcal{S}$ is a semi-field on Ω and $\mu : \mathcal{S} \to [0, \infty]$ is countably additive, then μ can be extended to a measure on $\big(\Omega, \sigma(\mathcal{S})\big)$.*

Theorem 2.4.3. *If $\Omega \neq \emptyset$, $\mathcal{F}$ is a field on Ω, and μ is a finitely additive set function on $\mathcal{F}$, then μ is*

(a) monotone, that is, $\mu(A) \leq \mu(B)$ for $A, B \in \mathcal{F}$ with $A \subset B$,

(b) finitely subadditive, that is,

$$\mu(A_1 \cup \ldots \cup A_n) \leq \mu(A_1) + \ldots + \mu(A_n),\ A_1, \ldots, A_n \in \mathcal{F},$$

(c) and countably superadditive, that is,

$$\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n) \text{ if } A_1, A_2, \ldots \in \mathcal{F} \text{ disjoint, } A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F} .$$

Proof. The first two claims are trivial. For (c), use (a) to argue for a finite N,

$$\mu(A) \geq \mu\left(A_1 \cup \ldots \cup A_N\right) = \sum_{n=1}^{N} \mu(A_n)\,.$$

Letting $N \to \infty$, (c) follows, which completes the proof. $\square$

Regarding the question of uniqueness, the following is a trivial consequence of Theorems 2.3.3 and 2.4.1.

Theorem 2.4.4. *Suppose $(\Omega, \mathcal{A})$ is a measurable space on which, μ_1, μ_2 are measures. Assume $\mathcal{S} \subset \mathcal{A}$ is a semi-field. If μ_1, μ_2 are σ-finite on $\mathcal{S}$, that is, there exist $A_1, A_2, \ldots \in S$ such that*

$$\bigcup_{n=1}^{\infty} A_n = \Omega \text{ and } \mu_i(A_n) < \infty \text{ for all } i = 1, 2,\ n = 1, 2, \ldots,$$

and

$$\mu_1(A) = \mu_2(A) \text{ for all } A \in \mathcal{S},$$

then

$$\mu_1(A) = \mu_2(A) \text{ for all } A \in \sigma(\mathcal{S})\,.$$

Proof. Exercise. $\square$

2.5 Riemann-Stieltjes measure on $\mathbb{R}$

Let us adopt the following convention. For any function $F : \mathbb{R} \to \mathbb{R}$,

$$F(\infty) = \lim_{x \to \infty} F(x) \text{ if it exists,}$$

and

$$F(-\infty) = \lim_{x \to -\infty} F(x) \text{ if it exists.}$$

Definition. *Given a function $F : \mathbb{R} \to \mathbb{R}$, a measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the* Riemann-Stieltjes measure corresponding to F *if*

$$\mu\big((a, b]\big) = F(b) - F(a) \text{ for all } -\infty < a \leq b < \infty\,. \tag{2.5.1}$$

Due to its importance, the following result is often referred to as the fundamental theorem in probability.

Theorem 2.5.1. *For any non-decreasing right continuous function $F : \mathbb{R} \to \mathbb{R}$, a unique Riemann-Stieltjes measure μ corresponding to F exists. If in addition, $F(\infty) = 1$ and $F(-\infty) = 0$, then μ is a probability measure.*

Proof. Let $\mathcal{S}$ be as in Example 2.4.1. Define $\mu : \mathcal{S} \to [0, \infty]$ by

$$\mu(A) = \begin{cases} F(b) - F(a), & A = (a, b] \cap \mathbb{R} \text{ for some } -\infty \le a < b \le \infty, \\ 0, & \text{otherwise, that is, if } A = \emptyset\,; \end{cases}$$

$F(b) - F(a)$ is defined whenever $a < b$ because then neither $F(b)$ nor $-F(a)$ equals $-\infty$ (the sum of two quantities in $[-\infty, \infty]$ is undefined if and only if one of them is ∞ and the other one $-\infty$).

The first step is to show that μ is finitely additive on $\mathcal{S}$. Let $A_1, \ldots, A_n \in \mathcal{S}$ be disjoint such that

$$A = A_1 \cup \ldots \cup A_n \in \mathcal{S}\,.$$

Without loss of generality, assume these are non-empty, that is,

$$A_i = (a_i, b_i] \cap \mathbb{R} \text{ for some } -\infty \le a_i < b_i \le \infty\,.$$

Since $A_i \cap A_j = \emptyset$ for all $1 \le i < j \le n$, either $b_i \le a_j$ or $b_j \le a_i$. By a relabelling, it can be assumed without loss of generality that

$$a_1 < b_1 \le a_2 < \ldots < b_{n-1} \le a_n < b_n\,.$$

Since $A \in \mathcal{S}$, it is necessary that $A = (a_1, b_n] \cap \mathbb{R}$ and hence $b_1 = a_2, \ldots, b_{n-1} = a_n$. Thus

$$\begin{aligned} \sum_{i=1}^{n} \mu(A_i) &= \sum_{i=1}^{n} \big(F(b_i) - F(a_i)\big) \\ \text{(because } b_1 = a_2, \ldots, b_{n-1} = a_n) &= F(b_n) - F(a_1) \\ &= \mu(A), \end{aligned}$$

showing μ is finitely additive on $\mathcal{S}$.

Let $\mathcal{F}$ be the field generated by $\mathcal{S}$. Extend μ to $\mathcal{F}$ by (2.4.2). Theorem 2.4.2 shows μ is finitely additive on $\mathcal{F}$. Theorem 2.4.3 tells us μ is monotone, finitely subadditive and countably superadditive on $\mathcal{F}$.

Our next task is to show μ is countably additive on $\mathcal{S}$. Let $A_1, A_2, \ldots \in \mathcal{S}$ be disjoint such that

$$\bigcup_{n=1}^{\infty} A_n = A \in \mathcal{S}\,.$$

Countable superadditivity of μ on $\mathcal{F}$ implies

$$\mu(A) \ge \sum_{n=1}^{\infty} \mu(A_n)\,.$$

Once it is shown that

$$\mu(A) \le \varepsilon + \sum_{n=1}^{\infty} \mu(A_n) \text{ for all } \varepsilon > 0, \tag{2.5.2}$$

countable additivity of μ on $\mathcal{S}$ would follow.

Fix $\varepsilon > 0$. Suppose $A_n = (a_n, b_n] \cap \mathbb{R}$ for some $a_n < b_n$. Let $\delta_n > 0$ be such that

$$F(b_n + \delta_n) \le \varepsilon 2^{-n} + F(b_n),\ n \ge 1,$$

which exists if $b_n < \infty$ by right continuity of F, and trivially holds for any $\delta_n > 0$ if $b_n = \infty$.

Suppose $A = (a, b] \subset \mathbb{R}$ for some $a < b$. The proof of (2.5.2) will be given separately for the cases $b < \infty$ and $b = \infty$. First assume $b < \infty$. Let $a' \in (a, b]$. In this case, $-\infty \le a < a' \le b < \infty$. Since

$$[a', b] \subset (a, b] = \bigcup_{n=1}^{\infty} \big((a_n, b_n] \cap \mathbb{R}\big) \subset \bigcup_{n=1}^{\infty} (a_n, b_n + \delta_n), \tag{2.5.3}$$

the Heine-Borel theorem implies

$$[a', b] \subset \bigcup_{n=1}^{N} (a_n, b_n + \delta_n)$$

for some finite N. Thus

$$(a', b] \subset [a', b] \subset \bigcup_{n=1}^{N} (a_n, b_n + \delta_n) \subset \bigcup_{n=1}^{N} \big((a_n, b_n + \delta_n] \cap \mathbb{R}\big).$$

Monotonicity of μ on $\mathcal{F}$ implies

$$\begin{aligned}
\mu\big((a', b]\big) &\le \mu\left(\bigcup_{n=1}^{N} \big((a_n, b_n + \delta_n] \cap \mathbb{R}\big)\right) \\
\text{(finite subadditiviety on } \mathcal{F}) &\le \sum_{n=1}^{N} \mu\big((a_n, b_n + \delta_n] \cap \mathbb{R}\big) \\
&= \sum_{n=1}^{N} \big(F(b_n + \delta_n) - F(a_n)\big) \\
&\le \sum_{n=1}^{\infty} \big(\varepsilon 2^{-n} + F(b_n) - F(a_n)\big) \\
&= \varepsilon + \sum_{n=1}^{\infty} \mu(A_n).
\end{aligned}$$

That is,

$$F(b) - F(a') \le \varepsilon + \sum_{n=1}^{\infty} \mu(A_n).$$

Since

$$\lim_{a' \downarrow a} F(a') = F(a), \tag{2.5.4}$$

which follows from right continuity of F is $a > -\infty$ and the definition if $a = -\infty$, (2.5.2) follows for the case $b < \infty$.

To prove (2.5.2) in the case $b = \infty$, fix $a < a' \le b' < b = \infty$. The arguments from (2.5.3) to (2.5.4) with b replaced by b', using the Heine-Borel theorem, imply

$$F(b') - F(a) \le \varepsilon + \sum_{n=1}^{\infty} \mu(A_n)\,.$$

Letting $b' \uparrow \infty$, (2.5.2) follows for the case $b = \infty$.

As argued before, (2.5.2) shows μ is countably additive on $\mathcal{S}$. Invoking Corollary 2.4.1, μ can be extended to a measure on $(\mathbb{R}, \sigma(\mathcal{S}))$. Clearly, $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$, and hence μ is a Riemann-Stieltjes measure corresponding to F. It is trivial that $\mu(\mathbb{R}) = 1$ if $F(\infty) = 1$ and $F(-\infty) = 0$.

For uniqueness, suppose μ_1 and μ_2 are Riemann-Stieltjes measure corresponding to F. Then

$$\mu_1\big((a,b]\big) = \mu_2\big((a,b]\big) \text{ for all } -\infty < a \le b < \infty\,.$$

Continuity from below implies

$$\mu_1(A) = \mu_2(A) \text{ for all } A \in \mathcal{S}\,.$$

Further,

$$\mathbb{R} = \bigcup_{n\in\mathbb{Z}} (n, n+1] \text{ and } \mu_i\big((n, n+1]\big) < \infty \text{ for all } n \in \mathbb{Z},\ i = 1, 2\,.$$

Theorem 2.4.4 implies μ_1 and μ_2 agree on $\sigma(\mathcal{S})$ which is $\mathcal{B}(\mathbb{R})$. Thus uniqueness follows, which completes the proof. □

Theorem 2.5.2. *Given a function $F : \mathbb{R} \to \mathbb{R}$, a Riemann-Stieltjes measure corresponding to F exists if and only if F is non-decreasing and right continuous.*

Proof. The "if" part is precisely the claim of Theorem 2.5.1. For the converse part, assume a Riemann-Stieltjes measure μ corresponding to F exists. Then for $-\infty < a < b < \infty$,

$$F(b) - F(a) = \mu\big((a,b]\big) \ge 0,$$

showing that F is non-decreasing. For right continuity, assume that $x_n \downarrow x_\infty$ for real numbers $x_1, \ldots, x_\infty$. Then

$$(x_\infty, x_n] \downarrow \emptyset\,.$$

Since $\mu\big((x_\infty, x_1]\big) = F(x_1) - F(x_\infty) < \infty$, continuity from above implies

$$\mu\big((x_\infty, x_n]\big) \downarrow 0\,.$$

In other words,

$$\lim_{n\to\infty} F(x_n) = F(x_\infty)\,.$$

As this holds for any $x_n \downarrow x_\infty$, F is right continuous at x_∞. Hence the "only if" part follows, which completes the proof. $\square$

Definition. *The Riemann-Stieltjes measure corresponding to the identity function F, that is,*

$$F(x) = x \text{ for all } x \in \mathbb{R},$$

is the Lebesgue measure *on $\mathbb{R}$.*

Exercise 2.5.1. 1. *Show that the Lebesgue measure λ on $\mathbb{R}$ is translation invariant, that is,*

$$\lambda(A+x) = \lambda(A) \text{ for all } x \in \mathbb{R},\ A \in \mathcal{B}(\mathbb{R}),$$

where $A + x = \{a + x : a \in A\}$.

2. *Hence or otherwise, prove that the Vitali set V, as in* (1.1.3)*, is not a Borel set.*

Definition. *A measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a* Radon measure *if $\mu(K) < \infty$ for every compact set $K \subset \mathbb{R}$.*

Exercise 2.5.2. *Suppose μ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Show that there exists a non-decreasing right continuous $F : \mathbb{R} \to \mathbb{R}$ such that μ is the Riemann-Stieltjes measure corresponding to F if and only if μ is a Radon measure.*

2.6 Measurable functions

Definition. *Suppose $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces. A function $f : \Omega_1 \to \Omega_2$ is* measurable *with respect to $\mathcal{A}_1/\mathcal{A}_2$ if*

$$f^{-1}(A) \in \mathcal{A}_1 \text{ for all } A \in \mathcal{A}_2,$$

where $f^{-1}(A) = \{\omega \in \Omega_1 : f(\omega) \in A\}$. Whenever either $\mathcal{A}_1$ or $\mathcal{A}_2$ is obvious from the context, it will be suppressed in the mention of measurability.

Theorem 2.6.1. *Suppose $\Omega_1 \neq \emptyset$ and $(\Omega_2, \mathcal{A}_2)$ is a measurable space. Then*

$$\left\{ f^{-1}A : A \in \mathcal{A}_2 \right\}$$

is a σ-field.

Proof. Let

$$\mathcal{A}_1 = \left\{ f^{-1}A : A \in \mathcal{A}_2 \right\} .$$

Clearly, $\Omega_1 = f^{-1}\Omega_2 \in \mathcal{A}_1$. It is easy to check that for any collection $(A_\alpha : \alpha \in I)$ of subsets of Ω_2,

$$f^{-1}\left(\bigcup_{\alpha \in I} A_\alpha \right) = \bigcup_{\alpha \in I} f^{-1}A_\alpha \,. \tag{2.6.1}$$

Hence if $B_1, B_2, \ldots \in \mathcal{A}_1$, that is, $B_n = f^{-1}A_n$ for some $A_n \in \mathcal{A}_2$, then

$$B_1 \cup B_2 \cup \ldots = f^{-1}(A_1 \cup A_2 \cup \ldots) \in \mathcal{A}_1 \text{ because } A_1 \cup A_2 \cup \ldots \in \mathcal{A}_2 \,.$$

Finally,

$$\left(f^{-1}A \right)^c = f^{-1}(A^c) \text{ for all } A \subset \Omega_2, \tag{2.6.2}$$

which shows $\mathcal{A}_1$ is closed under complements. Thus $\mathcal{A}_1$ is a σ-field, which completes the proof. □

Theorem 2.6.2. *Suppose* $(\Omega_1, \mathcal{A}_1)$ *and* $(\Omega_2, \mathcal{A}_2)$ *are measurable spaces and* $f : \Omega_1 \to \Omega_2$ *is a function. If* $\mathcal{G} \subset \mathcal{A}_2$ *is such that* $\sigma(\mathcal{G}) = \mathcal{A}_2$ *and*

$$f^{-1}A \in \mathcal{A}_1 \textit{ for all } A \in \mathcal{G},$$

then f *is measurable.*

Proof. Let

$$\mathcal{F} = \{ A \subset \Omega_2 : f^{-1}A \in \mathcal{A}_1 \} .$$

The hypothesis implies $\mathcal{G} \subset \mathcal{F}$. Clearly $\Omega_2 \in \mathcal{F}$ because $\mathcal{A}_1$ is a σ-field and hence $f^{-1}\Omega_2 = \Omega_1 \in \mathcal{A}_1$. Next (2.6.1) and (2.6.2) show $\mathcal{F}$ is a σ-field. Thus $\mathcal{F} \supset \sigma(\mathcal{G}) = \mathcal{A}_2$. Hence the proof follows. □

Theorem 2.6.3. *Suppose* $(\Omega_1, \mathcal{A}_1, \mu)$ *is a measure space,* $(\Omega_2, \mathcal{A}_2)$ *is a measurable space and* $T : \Omega_1 \to \Omega_2$ *is a measurable function. Define*

$$\nu(A) = \mu\left(T^{-1}A \right), \, A \in \mathcal{A}_2 \,. \tag{2.6.3}$$

Then ν *is a measure on* $(\Omega_2, \mathcal{A}_2)$.

Proof. Follows from (2.6.1). □

The following definition is motivated by the above theorem.

Definition. *Under the hypotheses of Theorem 2.6.3,* ν *defined by* (2.6.3) *is the* push-forward measure *of* μ *under* T*, and is denoted by*

$$\nu = \mu \circ T^{-1} \,.$$

Definition. *If* $(\Omega, \mathcal{A}, P)$ *is a probability space, a function* $X : \Omega \to \mathbb{R}$ *is a* random variable *if* X *is* $\mathcal{A}/\mathcal{B}(\mathbb{R})$*-measurable. For a random variable* X *defined on the probability space* $(\Omega, \mathcal{A}, P)$*, its* cumulative distribution function *(CDF)* F *is a function from* $\mathbb{R}$ *to* $[0, 1]$ *defined by*

$$F(x) = P\left(X^{-1}(-\infty, x]\right), \, x \in \mathbb{R} .$$

The following is an immediate consequence of Theorem 2.5.2.

Theorem 2.6.4. *Given a function* $F : \mathbb{R} \to \mathbb{R}$*, there exists a random variable* X *defined on some probability space* $(\Omega, \mathcal{A}, P)$ *such that* F *is its CDF if and only if* F *is right continuous, non-decreasing and satisfies* $F(\infty) = 1$ *and* $F(-\infty) = 0$*.*

Proof. For the "if" part, assume F has the given properties. Let P be the Riemann-Stieltjes measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ corresponding to F, that is,

$$P\big((a, b]\big) = F(b) - F(a), \; -\infty < a < b < \infty \, ;$$

P is a probability measure because $F(\infty) = 1$ and $F(-\infty) = 0$. Keeping b fixed and letting $a \to -\infty$ shows

$$P\big((-\infty, b]\big) = F(b), \; b \in \mathbb{R} .$$

Letting X be the identity function from $\mathbb{R}$ to $\mathbb{R}$, the above immediately shows that F is the CDF of X which is a random variable defined on the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$.

Conversely, suppose that F is the CDF of a random variable X defined on some probability space $(\Omega, \mathcal{A}, P)$. Let μ be the push-forward of P under X, that is, μ is the measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by $\mu = P \circ X^{-1}$. It is easy to check that μ is the Riemann-Stieltjes measure corresponding to F. Hence F is non-decreasing and right continuous by Theorem 2.5.2. Since

$$F(n) = P\left(X^{-1}(-\infty, n]\right), \, n \in \mathbb{Z},$$

$X^{-1}(-\infty, n] \uparrow \Omega$ and $X^{-1}(-\infty, -n] \downarrow \emptyset$, it follows that $F(\infty) = 1$ and $F(-\infty) = 0$. This proves the "only if" part follows, which completes the proof. $\square$

Exercise 2.6.1. *Show that*

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\{(-\infty, x) : x \in \mathbb{R}\}\right) = \sigma\left(\{(-\infty, x] : x \in \mathbb{R}\}\right) .$$

Theorem 2.6.5. *Suppose* $(\Omega, \mathcal{A})$ *is a measurable space and* f, g *are* $\mathcal{A}/\mathcal{B}(\mathbb{R})$*-measurable functions from* Ω *to* $\mathbb{R}$*. Then* $f + g$ *is* $\mathcal{A}/\mathcal{B}(\mathbb{R})$*-measurable.*

Proof. In view of Exercise 2.6.1, it suffices to show that

$$(f+g)^{-1}(-\infty, x) \in \mathcal{A} \text{ for all } x \in \mathbb{R}\,.$$

Fix $x \in \mathbb{R}$. We claim that

$$(f+g)^{-1}(-\infty, x) = \bigcup_{r \in \mathbb{Q}} \left(f^{-1}(-\infty, r)\right) \cap \left(g^{-1}(-\infty, x-r)\right)\,.$$

That the right hand side is a subset of the left hand side is obvious. For the reverse inclusion, fix ω in the left hand side, that is,

$$f(\omega) + g(\omega) < x\,.$$

Then $f(\omega) < x - g(\omega)$. Thus there exists $r \in \mathbb{Q}$ such that

$$f(\omega) < r < x - g(\omega)\,.$$

Hence $\omega \in \left(f^{-1}(-\infty, r)\right) \cap \left(g^{-1}(-\infty, x-r)\right)$. This shows the claimed set theoretic equality and hence the proof follows. $\square$

Exercise 2.6.2. *If $(\Omega, \mathcal{A})$ is a measurable space and $f : \Omega \to \mathbb{R}$ is measurable, show that for any $\alpha \in \mathbb{R}$, αf is measurable.*

Definition. *A function $f : \mathbb{R} \to \mathbb{R}$ is a* Borel function *if it is $\mathcal{B}(\mathbb{R})/\mathcal{B}(\mathbb{R})$-measurable.*

Theorem 2.6.6. *1. A continuous function is a Borel function.*

2. A monotone function is a Borel function.

Proof. 1. If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then $f^{-1}U$ is an open set for every open set $U \subset \mathbb{R}$. Since

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\{U \subset \mathbb{R} : U \text{ is open}\}\right),$$

f is Borel.

2. Suppose $f : \mathbb{R} \to \mathbb{R}$ is non-decreasing. Fix $x \in \mathbb{R}$ and let

$$\alpha = \sup\left(f^{-1}(-\infty, x]\right)\,.$$

It is immediate that $f^{-1}(-\infty, x]$ is either $(-\infty . \alpha)$ or $(-\infty, \alpha]$. In both cases,

$$f^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})\,.$$

Since this holds for all x, f is Borel. A similar argument holds for a non-increasing function.

$\square$

Theorem 2.6.7. *Suppose $(\Omega_i, \mathcal{A}_i)$ is a measurable space for $i = 1, 2, 3$. If $f : \Omega_1 \to \Omega_2$ and $g : \Omega_2 \to \Omega_3$ are measurable, then $g \circ f : \Omega_1 \to \Omega_3$ is measurable.*

Proof. Follows from the definition. $\square$

Theorem 2.6.8. *Suppose $(\Omega, \mathcal{A})$ is a measurable space and f, g are measurable functions from Ω to $\mathbb{R}$. Then fg is measurable.*

Proof. Write

$$fg = \left(\frac{f+g}{2}\right)^2 - \left(\frac{f-g}{2}\right)^2 .$$

Theorem 2.6.5 shows $(f+g)/2$ and $(f-g)/2$ are measurable. The function $x \mapsto x^2$ is Borel by Theorem 2.6.6. Theorem 2.6.7 shows $\big((f \pm g)/2\big)^2$ is measurable. The proof follows by applying Theorem 2.6.5 once more. $\square$

Theorem 2.6.9. *If f, g are measurable functions from some measurable space $(\Omega, \mathcal{A})$ to $\mathbb{R}$, then $f \vee g$ and $f \wedge g$ are measurable.*

Proof. Exercise. $\square$

3 Integration

3.1 Integration of non-negative functions

Henceforth, $(\Omega, \mathcal{A}, \mu)$, where $\mu(\Omega) > 0$, will be the underlying measure space. Unless mentioned otherwise, functions talked about are $\mathcal{A}$-measurable. The theory of integration is first developed for all non-negative measurable functions on Ω which possibly take the value infinity. To that end, define

$$\overline{\mathbb{R}} = [-\infty, \infty]$$

and

$$\mathcal{B}(\overline{\mathbb{R}}) = \sigma\left(\mathcal{B}(\mathbb{R}) \cup \big\{\{-\infty\}, \{\infty\}\big\}\right) .$$

That is, $\mathcal{B}(\overline{\mathbb{R}})$ is the smallest σ-field on $\overline{\mathbb{R}}$ which contains the singleton sets $\{-\infty\}$ and $\{\infty\}$, in addition to every Borel subset of $\mathbb{R}$.

Exercise 3.1.1. *Show that*

$$\mathcal{B}(\overline{\mathbb{R}}) = \sigma\left(\{[-\infty, x) : x \in \mathbb{R}\}\right) .$$

Definition. *A measurable function $s : \Omega \to \overline{\mathbb{R}}$ is a* simple function *if range of s is a finite set, that is,*

$$\# s(\Omega) < \infty .$$

First we shall define integral of simple functions, for which, the conventions of multiplication of two numbers which are possibly infinite have to be set. Define

$$x \cdot \infty = \infty \cdot x = \begin{cases} -\infty, & \text{if } -\infty \le x < 0, \\ 0, & \text{if } x = 0, \\ \infty, & \text{if } 0 < x \le \infty, \end{cases}$$

and

$$x \cdot (-\infty) = (-\infty) \cdot x = \begin{cases} \infty, & \text{if } -\infty \le x < 0, \\ 0, & \text{if } x = 0, \\ -\infty, & \text{if } 0 < x \le \infty\,. \end{cases}$$

It is worth emphasizing that $\pm\infty \cdot 0$ has been defined to be 0.

Definition. *If $s : \Omega \to [0, \infty]$ is a simple function whose range is $\{\alpha_1, \ldots \ldots, \alpha_n\}$, where $\alpha_1, \ldots, \alpha_n$ are distinct by convention, the integral of s with respect to μ, denoted by $\int s\, d\mu$, is defined by*

$$\int s\, d\mu = \sum_{i=1}^{n} \alpha_i \mu \left(s^{-1}\{\alpha_i\} \right) \,. \tag{3.1.1}$$

Theorem 3.1.1. *If $s = \sum_{i=1}^{n} \beta_i \mathbf{1}_{A_i}$, where $A_1, \ldots, A_n \in \mathcal{A}$ are disjoint and $\beta_1, \ldots, \beta_n \in [0, \infty]$, then*

$$\int s\, d\mu = \sum_{i=1}^{n} \beta_i \mu(A_i)\,.$$

Proof. Zero times anything, including ∞, is zero. Hence it can be assumed without loss of generality that $A_i \ne \emptyset$ for $i = 1, \ldots, n$. For $\omega \in A_i$, which exists as $A_i \ne \emptyset$,

$$s(\omega) = \beta_i,$$

because $A_1, \ldots, A_n$ are disjoint. Thus $\beta_1, \ldots, \beta_n \in s(\Omega)$. Further, for all $\alpha \in s(\Omega)$, either $\alpha = 0$ or $\alpha = \beta_i$ for some i. In the latter case,

$$s^{-1}\{\alpha\} = \bigcup_{i:\beta_i = \alpha} A_i,$$

and hence for $\alpha \in s(\Omega) \setminus \{0\}$,

$$\mu \left(s^{-1}\{\alpha\} \right) = \sum_{i:\beta_i = \alpha} \mu(A_i)\,. \tag{3.1.2}$$

The definition implies

$$\begin{aligned}
\int s\,d\mu &= \sum_{\alpha\in s(\Omega)} \alpha\mu\left(s^{-1}\{\alpha\}\right)\\
&= \sum_{\alpha\in s(\Omega)\setminus\{0\}} \alpha\mu\left(s^{-1}\{\alpha\}\right)\\
\big(\text{by } (3.1.2)\big) &= \sum_{\alpha\in s(\Omega)\setminus\{0\}} \alpha \sum_{i:\beta_i=\alpha} \mu(A_i)\\
&= \sum_{\alpha\in s(\Omega)\setminus\{0\}} \sum_{i:\beta_i=\alpha} \beta_i\mu(A_i)\\
&= \sum_{i:\beta_i\neq 0} \beta_i\mu(A_i)\\
&= \sum_{i=1}^{n} \beta_i\mu(A_i)\,.
\end{aligned}$$

This completes the proof. $\square$

Theorem 3.1.2. *If s, t are simple functions with $0 \le s \le t$, then*

$$\int s\,d\mu \le \int t\,d\mu\,.$$

Proof. Let

$$\left\{s^{-1}\{\alpha\}\cap t^{-1}\{\beta\} : \alpha\in s(\Omega),\ \beta\in t(\Omega)\right\} = \{A_1,\ldots,A_n\}\,.$$

For $i = 1,\ldots,n$, if $A_i \neq \emptyset$, fix $\omega \in A_i$ and denote

$$\alpha_i = s(\omega),\ \beta_i = t(\omega),$$

and let $\alpha_i = \beta_i = 0$ if $A_i = \emptyset$. Thus $\alpha_i \le \beta_i$ for all i, $A_1,\ldots,A_n$ are disjoint sets in $\mathcal{A}$ and

$$s = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i},\ t = \sum_{i=1}^{n} \beta_i \mathbf{1}_{A_i}\,.$$

Theorem 3.1.1 implies

$$\begin{aligned}
\int s\,d\mu &= \sum_{i=1}^{n} \alpha_i\mu(A_i)\\
&\le \sum_{i=1}^{n} \beta_i\mu(A_i)\\
&= \int t\,d\mu,
\end{aligned}$$

which completes the proof. $\square$

Definition. *For a measurable function $f : \Omega \to [0, \infty]$, define*

$$\int f\, d\mu = \sup\left\{\int s\, d\mu : 0 \leq s \leq f,\ s \textit{ is a simple function}\right\}.$$

Theorem 3.1.2 ensures that for a simple function $t \geq 0$, the above definition is consistent with (3.1.1), that is,

$$\sum_{\alpha \in t(\Omega)} \alpha\mu\left(t^{-1}\{\alpha\}\right) = \sup\left\{\int s\, d\mu : 0 \leq s \leq t,\ s \text{ is a simple function}\right\}.$$

The following theorem is a trivial consequence of the definition given above.

Theorem 3.1.3. *For measurable f, g with $0 \leq f \leq g$,*

$$\int f\, d\mu \leq \int g\, d\mu\,.$$

Proof. Exercise. □

Theorem 3.1.4. *Suppose $f_1, f_2, \ldots$ are measurable functions from $\Omega \to \overline{\mathbb{R}}$. Then*

$$g = \inf_n f_n \textit{ and } h = \sup_n f_n$$

are measurable.

Proof. Follows trivially from the observation that

$$[g < \alpha] = \bigcup_{n=1}^{\infty} [f_n < \alpha]$$

and Exc 3.1.1 and likewise for h. □

Exercise 3.1.2. *For measurable functions $f_1, f_2, \ldots$ from Ω to $\overline{\mathbb{R}}$, show the following are measurable:*

1. $f_1 + f_2$, *if it is defined,*
2. $f_1 f_2$,
3. $\liminf_{n\to\infty} f_n$,
4. $\limsup_{n\to\infty} f_n$,
5. $\lim_{n\to\infty} f_n$, *if it exists.*

The next theorem is of utmost importance in measure theory. As mentioned earlier, 'MCT' will refer to the following result and not the monotone class theorem (Theorem 2.3.2).

Theorem 3.1.5 (Monotone convergence theorem (MCT))**.** *If $f_n \geq 0$ and $f_n \uparrow f$, then*

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

f being measurable by Theorem 3.1.4.

The following exercise will be used for proving the above theorem.

Exercise 3.1.3. *If $\alpha_1, \ldots, \alpha_\infty, \beta_1, \ldots, \beta_\infty$ are such that $0 \leq \alpha_n \uparrow \alpha_\infty$ and $0 \leq \beta_n \uparrow \beta_\infty$, then*

$$\alpha_n \beta_n \uparrow \alpha_\infty \beta_\infty \,.$$

Proof of Theorem 3.1.5 (MCT). Theorem 3.1.3 implies that

$$\int f_1 \, d\mu \leq \int f_2 \, d\mu \leq \ldots \leq \int f \, d\mu \,.$$

Hence

$$\int f \, d\mu \geq \lim_{n \to \infty} \int f_n \, d\mu, \text{ which exists.}$$

To complete the proof, it suffices to show that for all

$$\alpha < \int f \, d\mu,$$

there exists n for which

$$\int f_n \, d\mu \geq \alpha \,. \tag{3.1.3}$$

Fix α as above. The definition of integral implies there exists a simple function s such that $0 \leq s \leq f$ and $\int s \, d\mu > \alpha$. Write

$$s = \sum_{i=1}^{k} \beta_i \mathbf{1}_{A_i}$$

where $A_1, \ldots, A_k \in \mathcal{A}$ are disjoint and $\beta_1, \ldots, \beta_k > 0$. Since

$$\begin{aligned} \alpha &< \int s \, d\mu \\ \text{(Theorem 3.1.1)} &= \sum_{i=1}^{k} \beta_i \mu(A_i) \\ &= \lim_{\gamma_1 \uparrow \beta_1, \ldots, \gamma_k \uparrow \beta_k} \sum_{i=1}^{k} \gamma_i \mu(A_i), \end{aligned}$$

Exc 3.1.3 implying the last line, there exist $0 \leq \gamma_1 < \beta_1, \ldots, 0 \leq \gamma_k < \beta_k$ such that

$$\sum_{i=1}^{k} \gamma_i \mu(A_i) > \alpha \,. \tag{3.1.4}$$

Let

$$A_{ni} = \{\omega \in A_i : f_n(\omega) > \gamma_i\},\ n = 1, 2, \ldots,\ \text{ and } i = 1, \ldots, k\,.$$

If $\omega \in A_i$, then

$$\lim_{n\to\infty} f_n(\omega) = f(\omega) \geq s(\omega) = \beta_i > \gamma_i\,.$$

Hence

$$A_{ni} \uparrow A_i,\ i = 1. \ldots, k\,.$$

Continuity from below implies

$$\mu\,(A_{ni}) \uparrow \mu(A_i),\ n \to \infty\,.$$

Exc 3.1.3 shows

$$\lim_{n\to\infty} \sum_{i=1}^{k} \gamma_i \mu\,(A_{ni}) = \sum_{i=1}^{k} \gamma_i \mu(A_i) > \alpha,$$

(3.1.4) implying the inequality.

Thus there exists n for which

$$\begin{aligned} \alpha &< \sum_{i=1}^{k} \gamma_i \mu(A_{ni}) \\ \text{(Theorem 3.1.1)} &= \int \left(\sum_{i=1}^{k} \gamma_i \mathbf{1}_{A_{ni}} \right) d\mu\,. \end{aligned}$$

The definition of A_{ni} implies

$$\sum_{i=1}^{k} \gamma_i \mathbf{1}_{A_{ni}} \leq f_n,$$

which in conjunction with the above shows

$$\alpha \leq \int f_n\, d\mu\,.$$

This show (3.1.3) from which the proof follows. □

The following is an immediate consequence of the MCT.

Theorem 3.1.6. *For $f, g \geq 0$ measurable and $\alpha \in [0, \infty]$,*

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

and

$$\int \alpha f\, d\mu = \alpha \int f\, d\mu\,.$$

Exercise 3.1.4. *For $f : \Omega \to [0, \infty]$ measurable, define*

$$s_n = n\mathbf{1}_{[f \geq n]} + \sum_{i=0}^{n2^n - 1} 2^{-n} i \mathbf{1}_{[2^{-n}i \leq f < 2^{-n}(i+1)]},\ n \geq 1\,.$$

Show that $0 \leq s_n \uparrow f$.

Proof of Theorem 3.1.6. For the first claim, let us first prove it for the case when f, g are simple. If

$$f = \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i} \text{ and } g = \sum_{j=1}^n \beta_j \mathbf{1}_{B_j},$$

where $A_1, \ldots, A_m \in \mathcal{A}$ are disjoint, $A_1 \cup \ldots \cup A_m = \Omega$ and likewise for $B_1, \ldots, B_n$, then

$$f + g = \sum_{i=1}^m \sum_{j=1}^n (\alpha_i + \beta_j) \mathbf{1}_{A_i \cap B_j}\,.$$

Theorem 3.1.1 shows

$$\int (f+g)\, d\mu = \sum_{i=1}^m \sum_{j=1}^n (\alpha_i + \beta_j) \mu(A_i \cap B_j) = \int f\, d\mu + \int g\, d\mu\,.$$

For general measurable f, g, Exc 3.1.4 and MCT prove the claim. The second claim follows in a similar way. □

Corollary 3.1.1. *If*

$$f = \sum_{i=1}^\infty \alpha_i \mathbf{1}_{A_i}$$

for $A_1, A_2, \ldots \in \mathcal{A}$ which are not necessarily disjoint and $\alpha_1, \alpha_2, \ldots \in [0, \infty]$, then

$$\int f\, d\mu = \sum_{i=1}^\infty \alpha_i \mu(A_i)\,.$$

Exercise 3.1.5. *Suppose $f : \Omega \to [0, \infty]$ is measurable with respect to a σ-field $\mathcal{A}_0 \subset \mathcal{A}$. Show that*

$$\int_{(\Omega, \mathcal{A}_0)} f\, d\mu = \int_{(\Omega, \mathcal{A})} f\, d\mu\,.$$

Exercise 3.1.6. *Suppose $f \geq 0$ is measurable.*

1. *If $\mu([f > 0]) > 0$, show that*

$$\int f\, d\mu > 0\,.$$

2. *If $\mu([f = \infty]) > 0$, show that*

$$\int f\, d\mu = \infty\,.$$

3.2 Integration of measurable functions

For all $x \in \mathbb{R}$,

$$x^+ = x \vee 0 \text{ and } x^- = (-x) \vee 0\,.$$

Definition. *For a measurable $f : \Omega \to \overline{\mathbb{R}}$, the integral of f with respect to μ, denoted by $\int f\,d\mu$, is defined by*

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

whenever either $\int f^+\,d\mu$ or $\int f^-\,d\mu$ is finite. If both $\int f^+\,d\mu$ and $\int f^-\,d\mu$ are finite, f is integrable.

The following notations will also be used to denote the integral of f with respect to μ:

$$\int_\Omega f\,d\mu,\ \int f(x)\,d\mu(x),\ \int f(x)\,\mu(dx) \text{ etc.}$$

Exercise 3.2.1. *1. Show that the following are equivalent for a measurable f:*

(a) f is integrable,

(b) $\int |f|\,d\mu < \infty$,

(c) the integral of f is defined and finite.

2. Show that 'f is integrable' and 'integral of f is defined' are two different claims, the former implies the latter, and the converse is false.

Theorem 3.2.1. *Suppose f, g are measurable whose integrals are defined and*

$$\int f\,d\mu + \int g\,d\mu \text{ is defined}. \tag{3.2.1}$$

Assume $f+g$ is defined, that is, there does not exist ω for which one of $f(\omega)$ and $g(\omega)$ is ∞ and the other one is $-\infty$. Then integral of $f+g$ is defined and

$$\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu\,.$$

Exercise 3.2.2. *For all $x, y \in \overline{\mathbb{R}}$ such that $x+y$ is defined, show that*

$$x^+ + y^+ + (x+y)^- = (x+y)^+ + x^- + y^-, \tag{3.2.2}$$

$$(x+y)^+ \le x^+ + y^+, \tag{3.2.3}$$

$$\text{and } (x+y)^- \le x^- + y^-\,. \tag{3.2.4}$$

Proof of Theorem 3.2.1. It follows from (3.2.2) that

$$f^+ + g^+ + (f+g)^- = (f+g)^+ + f^- + g^- .$$

Theorem 3.1.6 implies

$$\int f^+ \, d\mu + \int g^+ \, d\mu + \int (f+g)^- \, d\mu = \int (f+g)^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu . \tag{3.2.5}$$

If both sides of the above are finite, that is, f, g and $f+g$ are integrable, then $\int f^- \, d\mu < \infty$ and $\int g^- \, d\mu < \infty$ can be brought to the left hand side with a negative sign, and $\int (f+g)^- \, d\mu$ can be brought to the right hand side, again with a negative sign, from which, the claim would follow.

In case $\int f \, d\mu = \infty$, then (3.2.1) shows $\int g \, d\mu > -\infty$. Hence

$$\int f^+ \, d\mu = \infty, \int f^- \, d\mu < \infty \text{ and } \int g^- \, d\mu < \infty ;$$

(3.2.4) shows that

$$\int (f+g)^- \, d\mu < \infty .$$

Hence $f+g$ has a defined integral. In this case, (3.2.5) shows

$$\int (f+g)^+ \, d\mu = \infty,$$

from which it follows that

$$\int (f+g) \, d\mu = \infty = \int f \, d\mu + \int g \, d\mu .$$

The case $\int g \, d\mu = \infty$ can be proved similarly. The cases when either $\int f \, d\mu$ or $\int g \, d\mu$ is $-\infty$ can be proved by replacing f and g by $-f$ and $-g$, respectively. $\square$

Definition. *If $A \in \mathcal{A}$ is such that $\mu(A^c) = 0$, then anything which holds on A is said to hold* almost everywhere *or a.e.*

Theorem 3.2.2. *The following hold for f, g whose integrals are defined.*

1. *If $f \le g$ a.e., then*
$$\int f \, d\mu \le \int g \, d\mu .$$
2. *If $f = g$ a.e., then*
$$\int f \, d\mu = \int g \, d\mu .$$
3. *For $\alpha \in \mathbb{R}$,*
$$\int \alpha f \, d\mu = \alpha \int f \, d\mu .$$

4. It holds that

$$\left|\int f\,d\mu\right| \le \int |f|\,d\mu\,.$$

5. If f, g are integrable, then $f+g$ is defined a.e. and is integrable, and

$$\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu\,.$$

Proof. 1. If $f \le g$ a.e., then $f^+ \le g^+$ a.e. Thus

$$\begin{aligned}\int f^+\,dP &= \int_{[f^+\le g^+]} f^+\,dP \\ &\le \int g^+\,dP\,.\end{aligned}$$

A similar argument shows

$$\int f^-\,d\mu \ge \int g^-\,d\mu,$$

from which the claim follows.

2. Follows from 1.

3. Follows from the observation

$$(\alpha f)^+ = \alpha f^+ \text{ and } (\alpha f)^- = \alpha f^- \text{ if } \alpha \ge 0,$$

and

$$(\alpha f)^+ = -\alpha f^- \text{ and } (\alpha f)^- = -\alpha f^+ \text{ if } \alpha < 0\,.$$

4. The definition of integral implies

$$\begin{aligned}\left|\int f\,d\mu\right| &= \left|\int f^+\,d\mu - \int f^-\,d\mu\right| \\ &\le \int f^+\,d\mu + \int f^-\,d\mu \\ &= \int |f|\,d\mu,\end{aligned}$$

the last line following from Theorem 3.1.6.

5. Exc 3.2.1 implies that

$$\int |f|\,d\mu < \infty \text{ and } \int |g|\,d\mu < \infty\,.$$

Exc 3.1.6 shows f and g are finite a.e. Hence $f+g$ is defined a.e. The above along with the inequality $|f+g| \le |f| + |g|$ shows $f+g$ is integrable by Exc 3.2.1. The proof follows from Theorem 3.2.1.

□

Remark 3.2.1. *As is the case for $f+g$ in 5. of the above theorem, it suffices for a function to be defined a.e. On the set where it is not defined, the function may be redefined as zero, for example, which doesn't really matter as long as that set has zero measure.*

Theorem 3.2.3 (Fatou's lemma)**.** *For measurable $f_1, f_2, \ldots \geq 0$,*

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

Proof. Let

$$g_n = \inf_{k\geq n} f_k, \; n = 1, 2, \ldots,$$

and

$$g_\infty = \liminf_{n\to\infty} f_n \, .$$

Then $0 \leq g_n \uparrow g_\infty$. MCT implies

$$\begin{aligned} \int g_\infty \, d\mu &= \lim_{n\to\infty} \int g_n \, d\mu \\ &= \liminf_{n\to\infty} \int g_n \, d\mu \\ &\leq \liminf_{n\to\infty} \int f_n \, d\mu, \end{aligned}$$

the last line following from the obvious observation that $g_n \leq f_n$. This completes the proof. $\square$

Theorem 3.2.4 (Dominated convergence theorem)**.** *If $f_n \to f$ and $|f_n| \leq g$, where f_n, f, g are measurable and g is integrable, then f_n and f are integrable, and*

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

Proof. The assumptions imply $f_n + g \geq 0$. Further

$$\begin{aligned} \int (f+g) \, d\mu &= \int \liminf_{n\to\infty} (f_n + g) \, d\mu \\ \text{(Fatou's lemma)} &\leq \liminf_{n\to\infty} \int (f_n + g) \, d\mu \\ &= \int g \, d\mu + \liminf_{n\to\infty} \int f_n \, d\mu \, . \end{aligned}$$

Subtracting $\int g \, d\mu$ from both sides, which is finite, we get

$$\int f \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu \, . \tag{3.2.6}$$

A similar argument with Fatou's lemma shows

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

which implies

$$\int f\, d\mu \ge \limsup_{n\to\infty} \int f_n\, d\mu\,.$$

The proof follows by combining the above with (3.2.6). $\square$

3.3 Inequalities

As before, $(\Omega, \mathcal{A}, \mu)$ is a measure space, and all functions talked about are measurable functions from Ω to $\overline{\mathbb{R}}$.

Theorem 3.3.1 (Hölder)**.** *For $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\int |fg|\, d\mu \le \left(\int |f|^p\, d\mu\right)^{1/p} \left(\int |g|^q\, d\mu\right)^{1/q}.$$

The proof uses the following lemma.

Lemma 3.3.1. *For $0 < \lambda < 1$, $a, b \ge 0$ and $t > 0$,*

$$a^\lambda b^{1-\lambda} \le \lambda t^{\lambda-1} a + (1-\lambda) t^\lambda b\,.$$

Proof. Assume without loss of generality that $a, b > 0$ for else the left hand side vanishes. Define

$$f(x) = \lambda x^{\lambda-1} a + (1-\lambda) x^\lambda b\,, x > 0\,.$$

Differentiating with respect to x, we get

$$\begin{aligned} f'(x) &= \lambda(\lambda-1)x^{\lambda-2}a + (1-\lambda)\lambda x^{\lambda-1}b \\ &= \lambda(1-\lambda)x^{\lambda-2}(bx-a), \end{aligned}$$

where f' is the derivative of f. Thus $f' \ge 0$ on $[a/b, \infty)$ and $f' \le 0$ on $(0, a/b]$. Hence for any $t > 0$,

$$f(t) \ge f(a/b) = \lambda(a/b)^{\lambda-1}a + (1-\lambda)(a/b)^\lambda b = a^\lambda b^{1-\lambda},$$

which completes the proof. $\square$

Proof of Theorem 3.3.1. Fix measurable functions f, g and $p, q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} = 1\,. \tag{3.3.1}$$

Assume without loss of generality that

$$\int |f|^p \, d\mu > 0 \text{ and } \int |g|^q \, d\mu > 0,$$

because otherwise either $f = 0$ a.e. or $g = 0$ a.e. Next assume without loss of generality that the above quantities are finite as well, because otherwise the right hand side of the claimed inequality is ∞.

Use Lemma 3.3.1 with $\lambda = \frac{1}{p}$, $a = |f|^p$ and $b = |g|^q$ to get for a fixed $t > 0$

$$|f|\,|g| \le \frac{1}{p} t^{-1/q} |f|^p + \frac{1}{q} t^{1/p} |g|^q \, .$$

Thus,

$$\int |fg| \, d\mu \le \frac{1}{p} t^{-1/q} \int |f|^p \, d\mu + \frac{1}{q} t^{1/p} \int |g|^q \, d\mu \, .$$

The above holds for all $t > 0$. Putting

$$t = \left(\int |g|^q \, d\mu \right)^{-1} \int |f|^p \, d\mu,$$

the right hand side becomes

$$\frac{1}{p} \left(\int |f|^p \, d\mu \right)^{1-1/q} \left(\int |g|^q \, d\mu \right)^{1/q} + \frac{1}{q} \left(\int |f|^p \right)^{1/p} \left(\int |g|^q \, d\mu \right)^{1-1/p} .$$

Since (3.3.1) implies the above is the same as the right hand side of the claimed inequality, the proof follows. ☐

Theorem 3.3.2 (Cauchy-Schwarz)**.** *For f, g measurable,*

$$\int |fg| \, d\mu \le \left(\int f^2 \, d\mu \int g^2 \, d\mu \right)^{1/2} .$$

Proof. Follows from Theorem 3.3.1 by putting $p = q = 2$. ☐

Theorem 3.3.3 (Minkowski)**.** *For $1 \le p < \infty$,*

$$\left(\int |f + g|^p \, d\mu \right)^{1/p} \le \left(\int |f|^p \, d\mu \right)^{1/p} + \left(\int |g|^p \, d\mu \right)^{1/p} .$$

Proof. Without loss of generality, assume that the right hand side of the claimed inequality is finite. Then

$$\int |f + g|^p \, d\mu \le \int (|f| + |g|)^p \, d\mu \le 2^p \int (|f|^p + |g|^p) \, d\mu < \infty \, .$$

Denote

$$c = \left(\int |f + g|^p \, d\mu \right)^{1/p} < \infty \, .$$

Assume without loss of generality that $c > 0$ for else the claimed inequality is trivial. In addition, assume without loss of generality that $p > 1$ because for $p = 1$ the claim follows trivially from

$$|f + g| \le |f| + |g| \,.$$

Write

$$\begin{aligned}
c^p &= \int |f + g|^p \, d\mu \\
&= \int |f + g||f + g|^{p-1} \, d\mu \\
&\le \int |f||f + g|^{p-1} \, d\mu + \int |g||f + g|^{p-1} \, d\mu \\
\text{(Hölder inequality)} \quad &\le \left(\int |f|^p \, d\mu\right)^{1/p} \left(\int |f + g|^{(p-1)q} \, d\mu\right)^{1/q} \\
&\quad + \left(\int |g|^p \, d\mu\right)^{1/p} \left(\int |f + g|^{(p-1)q} \, d\mu\right)^{1/q} \\
&= c^{p/q} \left[\left(\int |f|^p \, d\mu\right)^{1/p} + \left(\int |g|^p \, d\mu\right)^{1/p}\right],
\end{aligned}$$

where $q = (1 - 1/p)^{-1}$. Therefore

$$\left(\int |f|^p \, d\mu\right)^{1/p} + \left(\int |g|^p \, d\mu\right)^{1/p} \ge c^{p-p/q} = c,$$

which completes the proof. □

Definition. *For $A \in \mathcal{A}$, define*

$$\int_A f \, d\mu = \int f \mathbf{1}_A \, d\mu$$

whenever the right hand side makes sense.

Theorem 3.3.4 (Markov)**.** *For $f \ge 0$ and $a \in (0, \infty)$,*

$$\mu\left([f \ge a]\right) \le \frac{1}{a} \int f \, d\mu \,.$$

Proof. By writing

$$\begin{aligned}
\int f \, d\mu &\ge \int_{[f \ge a]} f \, d\mu \\
&\ge \int_{[f \ge a]} a \, d\mu \\
&= a \, \mu\left([f \ge a]\right),
\end{aligned}$$

the proof follows. □

Definition. *For* $-\infty \le a < b \le \infty$, *a function* $\varphi : (a,b) \to \mathbb{R}$ *is* convex *if*

$$\varphi\left(\lambda x + (1-\lambda)y\right) \le \lambda\varphi(x) + (1-\lambda)\varphi(y) \text{ for all } x, y \in (a,b)\,.$$

Exercise 3.3.1. *If* $(a,b) \subset \mathbb{R}$ *and* $\varphi : (a,b) \to \mathbb{R}$ *is convex, show that*

1. φ *is continuous and hence Borel,*
2. *for all* $x \in (a,b)$,
$$\varphi'_+(x) = \lim_{h \downarrow 0} \frac{\varphi(x+h) - \varphi(x)}{h}$$
exists,
3. *and for all* $x, x_0 \in (a,b)$,
$$\varphi(x) \ge \varphi(x_0) + (x - x_0)\varphi'_+(x_0)\,. \tag{3.3.2}$$

Theorem 3.3.5 (Jensen)**.** *Suppose* $(\Omega, \mathcal{A}, P)$ *is a probability space,* $f : \Omega \to (a,b)$ *is measurable for some* $-\infty \le a < b \le \infty$ *and* $\varphi : (a,b) \to \mathbb{R}$ *is convex. If* f *and* $\varphi \circ f$ *are integrable functions, then*

$$a < \int f\,dP < b, \tag{3.3.3}$$

and

$$\int \varphi(f)\,dP \ge \varphi\left(\int f\,dP\right)\,.$$

Proof. Since $b > f$, Exc 3.1.6 shows that

$$0 < \int (b - f)\,dP = b - \int f\,dP,$$

the equality following from $P(\Omega) = 1$. As $\int f\,dP$ is a finite quantity, it can be taken to the left hand side, which shows

$$\int f\,dP < b\,.$$

A similar argument proves

$$\int f\,dP > a,$$

(3.3.3) follows from which.

Letting $x_0 = \int f\,dP$, (3.3.2) shows

$$\varphi(f) \ge \varphi(x_0) + (f - x_0)\varphi'_+(x_0)\,.$$

Thus

$$\begin{aligned}\int \varphi(f)\,dP &\geq \int \left(\varphi(x_0) + (f - x_0)\varphi'_+(x_0)\right)\,dP \\ \text{(since } P(\Omega) = 1) &= \varphi(x_0) + \varphi'_+(x_0)\left(\int f\,dP - x_0\right) \\ &= \varphi(x_0),\end{aligned}$$

and hence the proof. $\square$

3.4 The L^p space

As usual, $(\Omega, \mathcal{A}, \mu)$ is a measure space. All functions talked about are measurable functions from Ω to $\overline{\mathbb{R}}$, unless mentioned otherwise.

Definition. *For $1 \leq p < \infty$, define*

$$L^p(\Omega, \mathcal{A}, \mu) = \left\{ f : \Omega \to \overline{\mathbb{R}} \text{ measurable } : \int |f|^p\,d\mu < \infty \right\},$$

and for all $f \in L^p(\Omega, \mathcal{A}, \mu)$, define

$$\|f\|_p = \left(\int |f|^p\,d\mu\right)^{1/p}.$$

For $f, f_1, f_2, \ldots \in L^p(\Omega, \mathcal{A}, \mu)$, we say $f_n \to f$ in L^p if

$$\|f_n - f\|_p \to 0 \text{ as } n \to \infty.$$

The mention of $\mathcal{A}$ or μ in $L^p(\Omega, \mathcal{A}, \mu)$ will be suppressed whenever the same is obvious from the context. For $f, g \in L^p(\Omega)$, f and g will be identified with each other if $f = g$ a.e. In other words, $L^p(\Omega)$ will refer to the set of partitions induced by the equivalence relation $\sim$ which is defined as follows: for $f, g \in L^p(\Omega)$,

$$f \sim g \iff f = g \text{ a.e.}$$

Let us fix $1 \leq p < \infty$ for this subsection.

Theorem 3.4.1. *If*

$$d(f, g) = \|f - g\|_p,\ f, g \in L^p(\Omega),$$

then d is a metric on $L^p(\Omega)$, in which, two functions are identified if they are equal a.e.

Proof. The only non-trivial property of d to be checked is the triangle inequality, which follows from Theorem 3.3.3. $\square$

Definition. *For measurable functions $f_1, f_2, \ldots, f_\infty$ from Ω to $\overline{\mathbb{R}}$, say $f_n \to f_\infty$ a.e. if*

$$\mu\left(\left\{\omega \in \Omega : \lim_{n\to\infty} f_n(\omega) = f_\infty(\omega)\right\}^c\right) = 0\,.$$

Exercise 3.4.1. *If $f_n \to f$ a.e. and $|f_n| \le g$, where f_n, f, g are measurable and g is integrable, then show that*

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

Theorem 3.4.2 (Scheffé's lemma)**.** *Suppose $f, f_1, f_2, \ldots$ are measurable functions such that $0 \le f_n \to f$ a.e. Then $f_n \to f$ in L^1 if and only if*

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu < \infty\,.$$

Proof. For the 'if' part, write

$$\|f_n - f\|_1 = \int f_n\, d\mu + \int f\, d\mu - 2\int (f_n \wedge f)\, d\mu\,. \tag{3.4.1}$$

Exc 3.4.1 along with the fact $0 \le f_n \wedge f \le f$ implies

$$\lim_{n\to\infty} \int (f_n \wedge f)\, d\mu = \int f\, d\mu\,.$$

This along with the assumed hypothesis implies that the right hand side of (3.4.1) goes to zero as $n \to \infty$. Hence the 'if' part follows.

The 'only if' part follows trivially from

$$\left|\int f_n\, d\mu - \int f\, d\mu\right| \le \|f_n - f\|_1\,.$$

This completes the proof. $\square$

Definition. *For sets $A_1, A_2, A_3, \ldots$, define*

$$\limsup_n A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k,$$

and

$$\liminf_n A_n = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k\,.$$

Theorem 3.4.3 (Borel-Cantelli lemma)**.** *Suppose that $\sum_{n=1}^\infty \mu(A_n) < \infty$ for $A_1, A_2, \ldots \in \mathcal{A}$. Then*

$$\mu\left(\limsup_n A_n\right) = 0\,.$$

Proof. Let $B = \limsup A_n$ and

$$B_n = \bigcup_{k=n}^{\infty} A_k,\ n \geq 1\,.$$

Then $B_n \downarrow B$. Since

$$\mu(B_1) = \mu\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mu(A_k) < \infty,$$

it follows that $\mu(B_n) \downarrow \mu(B)$. Hence it suffices to show $\mu(B_n) \downarrow 0$.

Observing that for all $n = 1, 2, \ldots,$

$$\mu(B_n) \leq \sum_{k=n}^{\infty} \mu(A_k),$$

and that the right hand side goes to zero as $n \to \infty$, the proof follows. $\square$

Theorem 3.4.4. *For measurable $f, f_1, f_2, \ldots$ from Ω to $\mathbb{R}$, $f_n \to f$ a.e. if and only if*

$$\mu\left(\limsup_n \left[|f_n - f| > \varepsilon\right]\right) = 0 \text{ for all } \varepsilon > 0\,. \tag{3.4.2}$$

Proof. Define

$$g = \limsup_{n\to\infty} |f_n - f|\,.$$

Then $[f_n \to f] = [g = 0]$. For all $\varepsilon > 0$,

$$[g > \varepsilon] \subset [|f_n - f| > \varepsilon \text{ for infinitely many } n] = \limsup_n \left[|f_n - f| > \varepsilon\right] \tag{3.4.3}$$

$$\subset [g \geq \varepsilon]\,. \tag{3.4.4}$$

Thus if $g = 0$ a.e., then (3.4.4) implies

$$\mu\left(\limsup_n \left[|f_n - f| > \varepsilon\right]\right) \leq \mu([g \geq \varepsilon]) = 0,$$

which proves the 'only if' part.

For the 'if' part, assume (3.4.2). Let

$$\Omega_0 = \bigcap_{k=1}^{\infty} \left(\limsup_n \left[|f_n - f| > \frac{1}{k}\right]\right)^c\,.$$

Then

$$
\begin{aligned}
\mu\left(\Omega_0^c\right) &= \mu\left(\bigcup_{k=1}^{\infty}\left(\limsup_n \left[|f_n - f| > 1/k\right]\right)\right) \\
&\le \sum_{k=1}^{\infty} \mu\left[\left(\limsup_n \left[|f_n - f| > 1/k\right]\right)\right] \\
&= 0,
\end{aligned}
$$

(3.4.2) implying the last line. Finally if $\omega \in \Omega_0$, then for $k \ge 1$,

$$
\omega \in \left(\limsup_n \left[|f_n - f| > 1/k\right]\right)^c \subset [g \le 1/k],
$$

(3.4.3) implying the set inclusion. Thus $g(\omega) = 0$, which implies $f_n \to f$ on Ω_0. This proves the 'if' part. ☐

Theorem 3.4.5. *If $f_n \to f$ in L^p, then $\{f_n\}$ has a subsequence which converges to f a.e.*

Proof. The hypothesis implies the existence of $1 \le n_1 < n_2 < \ldots$ such that

$$
\|f_{n_k} - f\|_p \le 4^{-k},\ k \ge 1\,.
$$

Then

$$
\begin{aligned}
\mu\left(\left[|f_{n_k} - f| > 2^{-k}\right]\right) &= \mu\left(\left[|f_{n_k} - f|^p > 2^{-kp}\right]\right) && (3.4.5) \\
\text{(Markov inequality)} &\le 2^{kp}\int |f_{n_k} - f|^p \, d\mu && (3.4.6) \\
&= 2^{kp}\|f_{n_k} - f\|_p^p && (3.4.7) \\
&\le 2^{-kp}\,. && (3.4.8)
\end{aligned}
$$

Theorem 3.4.3 implies that

$$
|f_{n_k} - f| \le 2^{-k} \text{ for large enough } k,
$$

except on a set of measure zero. Hence $f_{n_k} \to f$ a.e. This completes the proof. ☐

Theorem 3.4.6. *The metric space $L^p(\Omega)$ is complete.*

Proof. Let $\{f_n\}$ be a Cauchy sequence in $L^p(\Omega)$, that is, for all $\varepsilon > 0$, there exists N such that

$$
\|f_m - f_n\|_p \le \varepsilon,\ m, n \ge N\,.
$$

This allows us to choose $1 \le n_1 < n_2 < \ldots$ such that

$$
\|f_i - f_j\|_p \le 4^{-k} \text{ for all } i, j \ge n_k\,. \tag{3.4.9}
$$

This by an appeal to Minkowski implies

$$\sup_{n\ge 1} \|f_n\|_p \le \frac{1}{4} + \max_{1\le i\le n_1} \|f_i\|_p < \infty\,. \tag{3.4.10}$$

Another immediate consequence of (3.4.9) is that

$$\left\|f_{n_{k+1}} - f_{n_k}\right\|_p \le 4^{-k},\ k \ge 1\,.$$

Arguments similar to (3.4.5)–(3.4.8) imply that

$$\mu\left(\left[|f_{n_{k+1}} - f_{n_k}| > 2^{-k}\right]\right) \le 2^{-kp},\ k \ge 1\,.$$

Theorem 3.4.3 implies

$$\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty \text{ a.e.}$$

Letting

$$\Omega_0 = \left[\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty\right],$$

it follows that for all $\omega \in \Omega_0$, $\{f_{n_k}(\omega) : k \ge 1\}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $\lim_{k\to\infty} f_{n_k}(\omega)$ exists for all $\omega \in \Omega_0$.

Let

$$f = \limsup_{k\to\infty} f_{n_k}\,.$$

Since $f_{n_k} \to f$ a.e.,

$$\begin{aligned}
\int |f|^p\, d\mu &= \int \liminf_{k\to\infty} |f_{n_k}|^p\, d\mu \\
\text{(Fatou's lemma)} &\le \liminf_{k\to\infty} \int |f_{n_k}|^p\, d\mu \\
&< \infty,
\end{aligned}$$

(3.4.10) implying the last line. Thus $f \in L^p(\Omega)$.

For a fixed $k = 1, 2, 3, \ldots$, (3.4.9) implies

$$\|f_{n_k} - f_{n_l}\|_p \le 4^{-k},\ l \ge k\,.$$

Thus

$$\begin{aligned}
\|f_{n_k} - f\|_p^p &= \int |f_{n_k} - f|^p\, d\mu \\
&= \int \liminf_{l\to\infty} |f_{n_k} - f_{n_l}|^p\, d\mu \\
\text{(Fatou's lemma)} &\le \liminf_{l\to\infty} \int |f_{n_k} - f_{n_l}|^p\, d\mu \\
&\le 4^{-kp}\,.
\end{aligned}$$

For $n \geq n_k$, $\|f_n - f_{n_k}\|_p \leq 4^{-k}$ by (3.4.9); Minkowski implies for such n,

$$\|f_n - f\|_p \leq \|f_n - f_{n_k}\|_p + \|f_{n_k} - f\|_p \leq 2^{1-2k}\,.$$

This shows $f_n \to f$ in L^p. In other words, every Cauchy sequence in $L^p(\Omega)$ is convergent. This completes the proof. $\square$

Definition. *For measurable $f : \Omega \to \overline{\mathbb{R}}$, define*

$$\|f\|_\infty = \inf\{0 \leq \alpha \leq \infty : |f| \leq \alpha \text{ a.e.}\}\,.$$

As before, $L^\infty(\Omega) = \{f : \|f\|_\infty < \infty\}$ with the understanding that two functions are considered the same if they are equal a.e. For $f, f_1, f_2, \ldots \in L^\infty(\Omega)$, $f_n \to f$ in L^∞ if $\|f_n - f\|_\infty \to 0$.

Exercise 3.4.2. *Show that $|f| \leq \|f\|_\infty$ a.e.*

Exercise 3.4.3. *Show that $L^\infty(\Omega)$ is a complete metric space.*

Exercise 3.4.4. *Show that Hölder's inequality holds with $p = 1$ and $q = \infty$, that is,*

$$\|fg\|_1 \leq \|f\|_1 \|g\|_\infty\,.$$

Exercise 3.4.5. *If $\mu(\Omega) < \infty$, show that*

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

Exercise 3.4.6. *Show the following for $f_1, \ldots, f_\infty$.*

1. *If $f_n \to f_\infty$ in L^∞, then show that there exist measurable functions $g_1, \ldots, g_\infty$ such that*

   $$\lim_{n \to \infty} \sup_{\omega \in \Omega} |g_n(\omega) - g_\infty(\omega)| = 0,$$

   and

   $$g_n = f_n \text{ a.e. for } n = 1, 2, \ldots, \infty\,.$$

   Since f_n and g_n are considered identical elements of $L^\infty(\Omega)$, convergence in L^∞ essentially means uniform convergence.

2. *If $f_n \to f_\infty$ a.e., then show that there exist measurable functions $g_1, \ldots, g_\infty$ such that*

   $$\lim_{n \to \infty} g_n(\omega) = g_\infty(\omega) \text{ for all } \omega \in \Omega,$$

   and

   $$g_n = f_n \text{ a.e. for } n = 1, 2, \ldots, \infty\,.$$

   In other words, a.e. convergence essentially means pointwise convergence.

3. *If $f_n \to f_\infty$ in L^∞, show that $f_n \to f_\infty$ a.e.*

Exercise 3.4.7. 1. *For $1 \le p < q \le \infty$, show that neither of $L^p(\mathbb{R}, \lambda)$ and $L^q(\mathbb{R}, \lambda)$ is a subset of the other, where λ is the Lebesgue measure.*

2. *If $\mu(\Omega) < \infty$, show that*

$$L^p(\Omega) \supset L^q(\Omega) \text{ if } 1 \le p \le q \le \infty\,.$$

3. *If $\mu(\Omega) = 1$, then show that*

$$\|f\|_p \le \|f\|_q \text{ if } 1 \le p \le q \le \infty\,.$$

This is known as Lyapunov's inequality.

4. *Let $\ell^p = L^p\big(\mathbb{N}, 2^{\mathbb{N}}, \mu\big)$ where μ is the counting measure, for $1 \le p \le \infty$. In other words,*

$$\ell^p = \left\{(x_1, x_2, x_3, \ldots) \in \mathbb{R}^{\mathbb{N}} : \sum_{n=1}^{\infty} |x_n|^p < \infty\right\} \text{ for } 1 \le p < \infty$$

and

$$\ell^\infty = \left\{(x_1, x_2, x_3, \ldots) \in \mathbb{R}^{\mathbb{N}} : \sup_{n\ge 1} |x_n| < \infty\right\}.$$

Show that

$$\ell^p \subsetneq \ell^q \text{ if } 1 \le p < q \le \infty\,.$$

Exercise 3.4.8. *Suppose $1 \le p \le q \le \infty$ and $f_1, f_2, \ldots \in L^p(\Omega) \cap L^q(\Omega)$. If*

$$f_n \to g_1 \text{ in } L^p,$$

$$f_n \to g_2 \text{ in } L^q$$

and

$$f_n \to g_3 \text{ a.e.},$$

show that $g_1 = g_2 = g_3$ a.e.

3.5 Lebesgue-Stieltjes integration

Let μ be the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. As in (2.2.1), denote by μ^* the outer measure of μ where $\mathcal{F}$ therein is replaced by $\mathcal{B}(\mathbb{R})$, that is,

$$\mu^*(E) = \inf\{\mu(A) : E \subset A,\, A \in \mathcal{B}(\mathbb{R})\},\, E \in 2^{\mathbb{R}}\,.$$

As in (2.2.4), let

$$\mathcal{L}(\mathbb{R}) = \left\{A \in 2^{\mathbb{R}} : \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \in 2^{\mathbb{R}}\right\}.$$

Lemma 2.2.5 shows that $\mathcal{L}(\mathbb{R})$ is a σ-field an μ^* is a measure on $(\mathbb{R}, \mathcal{L}(\mathbb{R}))$, which agrees with μ on $\mathcal{B}(\mathbb{R})$ by Lemma 2.2.1.

The following is a minor observation.

Theorem 3.5.1. *If $A \in \mathcal{B}(\mathbb{R})$ is such that $\mu(A) = 0$, then $B \in \mathcal{L}(\mathbb{R})$ for all $B \subset A$.*

Proof. Suppose $\mu(A) = 0$ for some $A \in \mathcal{B}(\mathbb{R})$ and $B \subset A$. Since μ^* is monotone by definition,

$$\mu^*(B) \le \mu^*(A) = 0\,.$$

Thus for any $E \in 2^{\mathbb{R}}$,

$$\mu^*(E \cap B) + \mu^*(E \cap B^c) \le \mu^*(B) + \mu^*(E \cap B^c) = \mu^*(E \cap B^c) \le \mu^*(E);$$

(2.2.5) shows that $B \in \mathcal{L}(\mathbb{R})$. This completes the proof. $\square$

Denote

$$\lambda(A) = \mu^*(A),\ A \in \mathcal{L}(\mathbb{R})\,.$$

Thus $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ is a measure space.

Exercise 3.5.1. *If $A \in \mathcal{B}(\mathbb{R})$ is such that $\lambda(A) = 0$ and $B \subset A$, show that for any function $g : \mathbb{R} \to \mathbb{R}$, $g\mathbf{1}_B$ is $\mathcal{L}(\mathbb{R})$-measurable.*

Soln.: If $h = g\mathbf{1}_B$, then for any $E \subset \mathbb{R} \setminus \{0\}$, $h^{-1}E \subset B \subset A$ and hence by Theorem 3.5.1, $h^{-1}E \in \mathcal{L}(\mathbb{R})$. For $E \subset \mathbb{R}$ such that $0 \in E$,

$$h^{-1}E = \left(h^{-1}(E^c)\right)^c \in \mathcal{L}(\mathbb{R})$$

because $h^{-1}(E^c) \in \mathcal{L}(\mathbb{R})$. This completes the solution.

Definition. *For $[a, b] \subset \mathbb{R}$, $f : [a, b] \to \overline{\mathbb{R}}$ is* Lebesgue measurable *if it is measurable with respect to the σ-field $\mathcal{L}([a, b])$, defined by*

$$\mathcal{L}([a, b]) = \{A \in \mathcal{L}(\mathbb{R}) : A \subset [a, b]\},$$

that is,

$$f^{-1}B \in \mathcal{L}([a, b]) \text{ for all } B \in \mathcal{B}(\overline{\mathbb{R}})\,.$$

If in addition, $\int_{[a,b]} |f|\, d\lambda < \infty$, then f is Lebesgue integrable *and $\int_{[a,b]} f\, d\lambda$ is the* Lebesgue integral *of f on $[a, b]$. The notations*

$$\int_a^b f(x)\, dx,\ \int_a^b f\, d\lambda \text{ etc.}$$

are also used for denoting the Lebesgue integral. As usual, define

$$\int_A f(x)\, dx = \int f\mathbf{1}_A\, d\lambda,\ A \in \mathcal{B}([a, b])\,.$$

In particular, if $A = [\alpha, \beta] \subset [a, b]$, then there is no conflict of notation in defining

$$\int_\alpha^\beta f\, d\lambda = \int_A f\, d\lambda\,.$$

Let us fix $[a,b] \subset \mathbb{R}$ for this subsection. Henceforth, "integrable on $[a,b]$" will mean Lebesgue integrable.

Exercise 3.5.2. *Suppose $f : [a,b] \to \mathbb{R}$ is Lebesgue integrable. Define*

$$g(x) = \int_a^x f(t)\,dt,\; a \le x \le b\,.$$

1. *Show that g is a continuous function.*
2. *If f is continuous at x, show that g is differentiable at x and its derivative at x equals $f(x)$.*

Theorem 3.5.2. *If $f : [a,b] \to \mathbb{R}$ is bounded and Riemann integrable, then f is Lebesgue integrable on $[a,b]$ and its Lebesgue integral equals its Riemann integral.*

Proof. Let $[a,b] \subset \mathbb{R}$ and $f : [a,b] \to \mathbb{R}$ be bounded and Riemann integrable. Let R denote its Riemann integral. Then

$$\lim_{n\to\infty} (b-a)2^{-n} \sum_{i=1}^{2^n} \inf\{f(x) : a+2^{-n}(i-1)(b-a) < x \le a+2^{-n}i(b-a)\} = R \tag{3.5.1}$$

and

$$\lim_{n\to\infty} (b-a)2^{-n} \sum_{i=1}^{2^n} \sup\{f(x) : a+2^{-n}(i-1)(b-a) < x \le a+2^{-n}i(b-a)\} = R\,. \tag{3.5.2}$$

Defining

$$P_n = \big\{\{a\}, (a, a+2^{-n}(b-a)], \ldots, (b-2^{-n}(b-a), b]\big\}\,,$$

$$\begin{aligned} f_n^L &= \sum_{A\in P_n} \mathbf{1}_A \inf_{x\in A} f(x), \\ \text{and } f_n^U &= \sum_{A\in P_n} \mathbf{1}_A \sup_{x\in A} f(x), \end{aligned}$$

(3.5.1) and (3.5.2) become

$$\lim_{n\to\infty} \int_{[a,b]} f_n^L\,d\lambda = R = \lim_{n\to\infty} \int_{[a,b]} f_n^U\,d\lambda\,.$$

This in view of the observation

$$f_1^L \le f_2^L \le \ldots \le f \le \ldots \le f_2^U \le f_1^U, \tag{3.5.3}$$

show that $\{g_n\}$, defined by

$$g_1 = f_1^L,\ g_2 = f_1^U,\ g_3 = f_2^L,\ g_4 = f_2^U, \ldots,$$

is a Cauchy sequence in $L^1\left([a,b], \mathcal{B}([a,b]), \lambda\right)$.

Theorem 3.4.6 shows there exists $g \in L^1([a,b], \mathcal{B}([a,b]), \lambda)$ such that $g_n \to g$ in L^1. Theorem 3.4.5 shows that each of $\{g_{2n-1}\}_{n\geq 1}$ and $\{g_{2n}\}_{n\geq 1}$ has a subsequence that converges to g in L^1. In other words, there exist $1 \leq m_1 < m_2 < \ldots$ such that

$$f_{m_k}^L \to g \text{ a.e.}$$

and $1 \leq n_1 < n_2 < \ldots$ such that

$$f_{n_k}^U \to g \text{ a.e.}$$

Denoting $A = [f_{m_k}^L \to g] \cap [f_{n_k}^U \to g]$, (3.5.3) shows that $f = g$ on A. Further, $A \in \mathcal{B}([a,b])$ and $\lambda([a,b] \setminus A) = 0$. Thus

$$f = g\mathbf{1}_A + f\mathbf{1}_{[a,b]\setminus A};$$

Exc 3.5.1 shows $f\mathbf{1}_{[a,b]\setminus A}$ is $\mathcal{L}([a,b])$-measurable. Thus so is f. Further, $f = g$ a.e. Hence $g_n \to f$ in L^1. This shows

$$\int_{[a,b]} f\, d\lambda = \lim_{n\to\infty} \int_{[a,b]} g_n\, d\lambda = R,$$

which completes the proof. $\square$

Definition. *A function $f : [a,b] \to \overline{\mathbb{R}}$ is* upper semi-continuous *at $x \in [a,b]$ if for all $\alpha > f(x)$, there exists $\varepsilon > 0$ such that*

$$f(y) \leq \alpha \text{ for all } y \in [x-\varepsilon, x+\varepsilon] \cap [a,b]\,.$$

On the other hand, f is lower semi-continuous *at x if for all $\beta < f(x)$, there exists $\varepsilon > 0$ such that*

$$f(y) \geq \beta \text{ for all } y \in [x-\varepsilon, x+\varepsilon] \cap [a,b]\,.$$

Exercise 3.5.3. *1. For $f : [a,b] \to \overline{\mathbb{R}}$ and $x \in [a,b]$, show that*

(a) f is upper semi-continuous at x if and only if for all $x_n \in [a,b]$ with $x_n \to x$,

$$\limsup_{n\to\infty} f(x_n) \leq f(x),$$

(b) and f is lower semi-continuous at x if and only if for all $x_n \in [a,b]$ with $x_n \to x$,

$$\liminf_{n\to\infty} f(x_n) \geq f(x)\,.$$

2. *For $f : [a, b] \to \overline{\mathbb{R}}$, show that f is upper semi-continuous if and only if $-f$ is lower semi-continuous.*

3. *If f, g are upper semi-continuous functions on $[a, b]$, show that so if $f + g$.*

4. *If $K_1, \ldots, K_n \subset \mathbb{R}$ are compact sets and $\alpha_1, \ldots, \alpha_n \geq 0$, show that*
$$\sum_{i=1}^{n} \alpha_i \mathbf{1}_{K_i \cap [a,b]}$$
is an upper semi-continuous function on $[a, b]$.

5. *If $V_1, V_2, \ldots \subset \mathbb{R}$ are open sets and $\alpha_1, \alpha_2, \ldots \geq 0$, show that*
$$\sum_{i=1}^{\infty} \alpha_i \mathbf{1}_{V_i \cap [a,b]}$$
is a lower semi-continuous function on $[a, b]$.

Theorem 3.5.3 (Vitali-Carathéodory)**.** *If $f \in L^1([a, b], \mathcal{B}([a, b]), \lambda)$, then for all $\varepsilon > 0$ there exist an upper semi-continuous u and a lower semi-continuous v on $[a, b]$ such that u, v are integrable, $u \leq f \leq v$ and*
$$\int_a^b (v - u)\, d\lambda < \varepsilon\,.$$

The proof uses the following exercise.

Exercise 3.5.4. *For all $A \in \mathcal{B}(\mathbb{R})$ with $\lambda(A) < \infty$ and $\varepsilon > 0$, there exist a compact K and an open V with $K \subset A \subset V$ and $\lambda(V \setminus K) < \varepsilon$.*

Proof of Theorem 3.5.3. Let us first prove this for the case $f \geq 0$. Since f is a Borel function, there exist simple Borel functions s_n such that $0 \leq s_n < \infty$ and $s_n \uparrow f$. Denote $s_0 = 0$ and let
$$t_n = s_n - s_{n-1}, n = 1, 2, \ldots,$$
so that we get
$$f = \sum_{n=1}^{\infty} t_n\,.$$
Since each t_n is a non-negative simple Borel function, there exist $c_1, c_2, \ldots > 0$ and $A_1, A_2, \ldots \in \mathcal{B}([a, b])$ such that
$$f = \sum_{i=1}^{\infty} c_i \mathbf{1}_{A_i}\,. \tag{3.5.4}$$

MCT implies

$$\int_a^b f\,d\lambda = \sum_{i=1}^\infty c_i\lambda(A_i)\,.$$

Since f is integrable, the series on the right hand side is finite. Fix $\varepsilon > 0$ and let n be such that

$$\sum_{i=n+1}^\infty c_i\lambda(A_i) < \frac{\varepsilon}{2}\,.$$

Exc 3.5.4 shows there exist for $i = 1, 2, \ldots$, a compact K_i and open V_i with $K_i \subset A_i \subset V_i$ and

$$\lambda(V_i \setminus K_i) < \frac{1}{c_i} 2^{-i-1}\varepsilon\,.$$

Letting

$$u = \sum_{i=1}^n c_i \mathbf{1}_{K_i} \text{ and } v = \sum_{i=1}^\infty c_i \mathbf{1}_{V_i\cap[a,b]},$$

(3.5.4) shows $u \le f \le v$. Exc 3.5.3 shows u and v are respectively upper and lower semi-continuous. Finally,

$$\begin{aligned}
\int_a^b (v-u)\,d\lambda &= \int_a^b \left(\sum_{i=1}^n c_i \mathbf{1}_{V_i\cap[a,b]\cap K_i^c} + \sum_{i=n+1}^\infty c_i \mathbf{1}_{V_i\cap[a,b]}\right) d\lambda \\
&\le \sum_{i=1}^n c_i\lambda(V_i \setminus K_i) + \sum_{i=n+1}^\infty c_i\lambda(V_i) \\
&< \sum_{i=1}^n 2^{-i-1}\varepsilon + \sum_{i=n+1}^\infty c_i\lambda(V_i) \\
&= \sum_{i=1}^n 2^{-i-1}\varepsilon + \sum_{i=n+1}^\infty c_i\lambda(A_i) + \sum_{i=n+1}^\infty c_i\lambda(V_i \setminus A_i) \\
&< \sum_{i=1}^n 2^{-i-1}\varepsilon + \frac{\varepsilon}{2} + \sum_{i=n+1}^\infty 2^{-i-1}\varepsilon \\
&= \varepsilon\,.
\end{aligned}$$

This completes the proof for the case $f \ge 0$.

For an integrable f, not necessarily non-negative, use the above to get $0 \le u_1 \le f^+ \le v_1$ and $0 \le u_2 \le f^- \le v_2$, where u_1, u_2 are upper semi-continuous, v_1, v_2 are lower semi-continuous and

$$\int_a^b (v_i - u_i)\,d\lambda < \frac{\varepsilon}{2},\ i = 1, 2\,.$$

Letting $u = u_1 - v_2$ and $v = v_1 - u_2$, the proof follows. □

Exercise 3.5.5. *If $f : [a,b] \to [0,\infty]$ is a Borel function such that*

$$\int_a^b f \, d\lambda = \infty,$$

show that

$$\sup\left\{\int_a^b u \, d\lambda : 0 \le u \le f, \ u \text{ upper semi-continuous integrable on } [a,b]\right\} = \infty \,. \tag{3.5.5}$$

Soln.: If $f : [a,b] \to [0,\infty]$ is a Borel non-integrable function, then (3.5.4) still holds for some $c_1, c_2, \ldots \in [0,\infty)$. MCT implies

$$\sum_{i=1}^{\infty} c_i \lambda(A_i) = \infty \,.$$

Thus for any $\alpha < \infty$, there exists n_0 such that

$$\sum_{i=1}^{n_0} c_i \lambda(A_i) > \alpha \,.$$

As before, compact sets $K_1, \ldots, K_{n_0}$ can be chosen such that $K_i \subset A_i$ for $i = 1, \ldots, n_0$, which is possible because $A_1, \ldots, A_{n_0}$ are bounded sets and hence has finite Lebesgue measure, and

$$\sum_{i=1}^{n_0} c_i \lambda(K_i) \ge \alpha \,.$$

Setting

$$u = \sum_{i=1}^{n_0} c_i \mathbf{1}_{K_i},$$

it follows that $u \ge 0$ is upper semi-continuous and

$$\alpha \le \int_a^b u(x) \, dx = \sum_{i=1}^{n_0} c_i \lambda(K_i) < \infty \,.$$

Since $\alpha < \infty$ is arbitrary, (3.5.5) follows, which completes the solution.

Exercise 3.5.6. *If $F : [a,b] \to \mathbb{R}$ is differentiable, that is, its right derivative exists on $[a,b)$, the left derivative exists on $(a,b]$ and the two agree on (a,b), then show that the derivative of F is a Borel function.*

Unless mentioned otherwise, $f'(x)$ denotes the derivative of f at x if it exists.

Theorem 3.5.4 (Fundamental theorem of calculus)**.** *Suppose $f : [a,b] \to \mathbb{R}$ is differentiable and f' is integrable on $[a,b]$, that is,*

$$\int_a^b |f'(x)|\,dx < \infty\,.$$

Then

$$\int_a^b f'(x)\,dx = f(b) - f(a)\,.$$

The following lemma is the main content of the proof.

Lemma 3.5.1. *If $f : [a,b] \to \mathbb{R}$ is differentiable and $u : [a,b] \to \mathbb{R}$ is integrable and upper semi-continuous on $[a,b]$ with $u < f'$, then*

$$\int_a^b u(x)\,dx \le f(b) - f(a)\,.$$

Proof. Let u, f be as given in the hypothesis. For any $x \in [a,b)$, fixed for a moment, $u(x) < f'(x)$. Upper semi-continuity of u at x implies there exists $0 < \delta_x^{(1)} < b - x$ such that

$$u(t) \le f'(x) \text{ for all } x \le t \le x + \delta_x^{(1)}\,.$$

Fix $\varepsilon > 0$. The definition of derivative implies there exists $0 < \delta_x^{(2)} < b-x$ such that

$$\frac{f(t) - f(x)}{t - x} \ge f'(x) - \varepsilon \text{ for all } x < t \le x + \delta_x^{(2)}\,.$$

Define $\delta_x = \delta_x^{(1)} \wedge \delta_x^{(2)}$.

For all $x \in [a,b]$, define

$$F(x) = \int_a^x u(t)\,dt - f(x) + f(a) - \varepsilon(x - a)\,.$$

For fixed $x \in [a,b)$ and $x \le y \le x + \delta_x$,

$$\begin{aligned}
F(y) - F(x) &= \int_x^y u(t)\,dt - f(y) + f(x) - \varepsilon(y - x) \\
(\text{choice of } \delta_x^{(1)}) &\le \int_x^y f'(x)\,dt - f(y) + f(x) - \varepsilon(y - x) \\
&= (y - x)f'(x) - f(y) + f(x) - \varepsilon(y - x) \\
(\text{choice of } \delta_x^{(2)}) &\le (y - x)\left(\frac{f(y) - f(x)}{y - x} + \varepsilon\right) - f(y) + f(x) - \varepsilon(y - x) \\
&= 0\,.
\end{aligned}$$

That is,

$$F(x) \geq F(y) \text{ whenever } x \leq y \leq x + \delta_x \,. \tag{3.5.6}$$

Let $\alpha = a + \delta_a$ and fix $\alpha \leq \beta < b$. Since

$$[\alpha, \beta] \subset \bigcup_{x \in [a,b)} (x, x + \delta_x),$$

the Heine-Borel theorem implies that there exist $a \leq x_1 < x_2 < \ldots < x_n < b$ such that

$$[\alpha, \beta] \subset \bigcup_{i=1}^{n} (x_i, x_i + \delta_{x_i}) \;.$$

Thus $\beta \in (x_{i_1}, x_{i_1} + \delta_{x_{i_1}})$ for some i_1, which along with (3.5.6) shows that

$$F(\beta) \leq F(x_{i_1}) \,.$$

Either $i_1 = 1$ or there exists $i_2 < i_1$ such that

$$F(x_{i_1}) \leq F(x_{i_2}) \,.$$

Proceeding inductively, it can be shown that

$$F(\beta) \leq F(x_1) \leq F(a)$$

by choice of δ_a because $x_1 < \alpha = x + \delta_a$.

Letting $\beta \uparrow b$ and using continuity of F, which follows from Exc 3.5.2, shows $F(b) \leq F(a) = 0$. In other words,

$$\int_a^b u(t)\, dt \leq f(b) - f(a) + \varepsilon(b - a) \,.$$

Since ε is arbitrary, the proof follows. □

Proof of Theorem 3.5.4. Fix $\varepsilon > 0$. Since f' is an integrable Borel function on $[a, b]$, Theorem 3.5.3 shows there exists an upper semi-continuous integrable u on $[a, b]$ such that $u \leq f'$ and

$$\int_a^b (f' - u)\, d\lambda < \varepsilon \,.$$

Since $u - \varepsilon < f'$, Lemma 3.5.1 shows that

$$\begin{aligned} f(b) - f(a) &\geq \int_a^b (u - \varepsilon)\, d\lambda \\ &= \int_a^b u\, d\lambda - \varepsilon(b - a) \\ &> \int_a^b f'\, d\lambda - \varepsilon(1 + b - a) \,. \end{aligned}$$

Since ε is arbitrary, it follows that

$$f(b) - f(a) \geq \int_a^b f' \, d\lambda \, .$$

Replacing f by $-f$, the reverse inequality follows, which completes the proof. $\square$

The following is another version of the fundamental theorem of calculus, and is important for probability theory.

Theorem 3.5.5. *If $F : [a, b] \to \mathbb{R}$ is non-decreasing and differentiable, then*

$$\int_a^b F'(x) \, dx = F(b) - F(a) \, .$$

Proof. Let F be as given. In view of Theorem 3.5.4, it suffices to show that

$$\int_a^b F'(x) \, dx < \infty \, .$$

Exc 3.5.5 would imply the above once it is shown that

$$\sup \left\{ \int_a^b u \, d\lambda : \, 0 \leq u \leq F', \, u \text{ upper semi-continuous integrable on } [a, b] \right\} \tag{3.5.7}$$
$$< \infty \, .$$

For any upper semi-continuous integrable u with $0 \leq u \leq F'$, Lemma 3.5.1 applied to $u - 1$ shows that

$$\int_a^b (u - 1) \, d\lambda \leq F(b) - F(a) \, .$$

Hence the quantity in (3.5.7) is at most $F(b) - F(a) + b - a$ and hence finite, from which the proof follows. $\square$

Definition. *Let $F : [a, b] \to \mathbb{R}$ be a non-decreasing right continuous function. Extend F to $\mathbb{R}$ by*

$$F(x) = \begin{cases} F(a), & \text{if } x < a, \\ F(b), & \text{if } x > b, \end{cases} \tag{3.5.8}$$

and denote by μ_F the Riemann-Stieltjes measure of F as in (2.5.1). For any Borel function $f : [a, b] \to \overline{\mathbb{R}}$, the Stieltjes integral of f with respect to F is defined by

$$\int_{[a,b]} f(x) F(dx) = \int_{[a,b]} f \, d\mu_F,$$

whenever the right hand side is defined. The Stieltjes integral on the left hand side is also denoted by $\int_{[a,b]} f \, dF$, $\int_{[a,b]} f(x) \, dF(x)$ etc.

It should be noted that when F is the identity function, that is, $F(z) = z$ for all $z \in \mathbb{R}$, μ_F is the Lebesgue measure. Hence the Lebesgue integral is a special case of the Stieltjes integral (for Borel functions). Furthermore, the use of the notation $\int_{[a,b]} f(x)\,dx$ for the Lebesgue integral is also justified by taking F to be the identity function (and hence interpreting $F(dx)$ as dx) in the Stieltjes integral.

Recall that $P = (x_0, x_1, \dots, x_n)$ is a partition of $[a, b]$ if $a = x_0 < x_1 < \dots < x_n = b$ whose "mesh" is

$$\min_{1 \le i \le n} (x_i - x_{i-1}) \,.$$

For such P and a function $f : [a, b] \to \mathbb{R}$, the upper and lower Stieltjes sum, denoted respectively by $U(P, F, f)$ and $L(P, F, f)$, is defined by

$$U(P, F, f) = \sum_{i=1}^{n} (F(x_i) - F(x_{i-1})) \sup_{x_{i-1} \le x \le x_i} f(x),$$

and

$$L(P, F, f) = \sum_{i=1}^{n} (F(x_i) - F(x_{i-1})) \inf_{x_{i-1} \le x \le x_i} f(x) \,.$$

When F is the identity function, the upper and lower Stieltjes sums become the respective Riemann sums.

Theorem 3.5.6. *Suppose $F : [a, b] \to \mathbb{R}$ is a non-decreasing right continuous function, and P_n is a sequence of partitions of $[a, b]$ whose mesh goes to zero as $n \to \infty$. Then for any continuous $f : [a, b] \to \mathbb{R}$,*

$$\lim_{n\to\infty} U(P_n, F, f) = \int_{[a,b]} f(x)F(dx) = \lim_{n\to\infty} L(P_n, F, f) \,.$$

Proof. Fix $n = 1, 2, \dots$ and denote $P_n = (x_0, \dots, x_k)$. Define

$$f_n(x) = f(a)\mathbf{1}_{\{a\}}(x) + \sum_{i=1}^{k} \mathbf{1}_{(x_{i-1},x_i]}(x) \sup_{x_{i-1} \le z \le x_i} f(z),\ a \le x \le b \,.$$

Denote by μ_F the Riemann-Stieltjes measure of F which is extended to the whole of $\mathbb{R}$ by (3.5.8). It is easy to see that

$$\int_{[a,b]} f_n \, d\mu_F = U(P_n, F, f) \,.$$

Fix $\varepsilon > 0$. Since f is continuous on $[a, b]$, it is uniformly continuous. Hence there exists $\delta > 0$ such that

$$|f(x) - f(y)| \le \varepsilon \text{ if } |x - y| \le \delta,\ x, y \in [a, b] \,.$$

For n large enough such that mesh of P_n smaller than δ, it is immediate that

$$|f_n(x) - f(x)| \le \varepsilon,\ a \le x \le b\,.$$

Hence for such n,

$$\left| U(P_n, F, f) - \int_{[a,b]} f(x)\, dF(x) \right| = \left| \int_{[a,b]} (f_n - f)\, d\mu_F \right| \le \varepsilon\,(F(b) - F(a))\ .$$

Since ε is arbitrary, it follows that

$$\lim_{n\to\infty} U(P_n, F, f) = \int_{[a,b]} f(x) F(dx)\,.$$

A similar argument works for $L(P_n.F, f)$ and completes the proof. $\square$

Exercise 3.5.7. *Suppose* $(\Omega, \mathcal{A}, \mu)$ *is a measure space and* $f : \Omega \to [0, \infty]$ *is measurable. Define*

$$\nu(A) = \int_A f\, d\mu,\ A \in \mathcal{A}\,. \tag{3.5.9}$$

1. *Show that* ν *is a measure on* $(\Omega, \mathcal{A})$*.*
2. *Show that for any measurable function* $g : \Omega \to \overline{\mathbb{R}}$,

$$\int g\, d\nu = \int gf\, d\mu, \tag{3.5.10}$$

whenever either side is defined. This is known as the "change of measure" formula.

Definition. *Suppose* $(\Omega, \mathcal{A})$ *is a measurable space on which* μ *and* ν *are measures. If there exists* $f : \Omega \to [0, \infty]$ *such that* (3.5.9) *holds, then* f *is a* density *of* ν *with respect to* μ*, and we usually denote*

$$f = \frac{d\nu}{d\mu}\,.$$

In view of the above definition, (3.5.10) becomes

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

Exercise 3.5.8. *Suppose* $(\Omega, \mathcal{A})$ *is a measurable space on which* μ *and* ν *are measures. Assume* ν *is a* σ*-finite measure. Show that if* f *and* g *are densities of* ν *with respect to* μ*, then* $f = g$ *a.e.*

Exercise 3.5.9. *If μ and ν are finite measures on $\big([a,b], \mathcal{B}([a,b])\big)$ satisfying*

$$\mu\big((\alpha,\beta]\big) = \nu\big((\alpha,\beta]\big) \text{ for all } a \le \alpha < \beta \le b,$$

then show that μ and ν are identical.

The following theorem justifies the substitution

$$dF(x) = F'(x)\,dx$$

whenever F is differentiable.

Theorem 3.5.7. *If $F : [a,b] \to \mathbb{R}$ is differentiable and non-decreasing, then for a Borel function $f : [a,b] \to \overline{\mathbb{R}}$,*

$$\int_{[a,b]} f(x)\,dF(x) = \int_a^b f(x)F'(x)\,dx,$$

whenever either side is defined.

Proof. Let F be extended to the whole of $\mathbb{R}$ by (3.5.8) and μ_F be the Riemann-Stieltjes measure of F. Let ν be the measure on $\big([a,b], \mathcal{B}([a,b])\big)$ defined by

$$\nu(A) = \int_A F'(x)\,dx,\ A \in \mathcal{B}([a,b])\,.$$

Theorem 3.5.5 shows that for $a \le \alpha < \beta \le b$,

$$\nu\left((\alpha,\beta]\right) = \int_\alpha^\beta F'(x)\,dx = F(\beta) - F(\alpha) = \mu_F\left((\alpha,\beta]\right)\,.$$

Exc 3.5.9 shows $\mu_F = \nu$ on $[a,b]$.

For $f : [a,b] \to \overline{\mathbb{R}}$ Borel, (3.5.10) shows

$$\int_{[a,b]} f\,d\nu = \int_a^b f(x)F'(x)\,dx,$$

whenever either side makes sense. Since $\nu = \mu_F$, the left hand side is the same as $\int_{[a,b]} f\,dF$, from which the proof follows. $\square$

Exercise 3.5.10. *Suppose $(\Omega_1, \mathcal{A}_1, \mu)$ is a measure space, $(\Omega_2, \mathcal{A}_2)$ is a measurable space and $T : \Omega_1 \to \Omega_2$ is a measurable map. Then for a measurable $f : \Omega_2 \to \overline{\mathbb{R}}$, show that*

$$\int_{\Omega_1} f\left(T(\omega)\right)\,\mu(d\omega) = \int_{\Omega_2} f(x)\,\mu\circ T^{-1}(dx), \tag{3.5.11}$$

whenever either side makes sense. This is the so-called change of variables formula for push-forward measures.

Theorem 3.5.8. *Suppose $F : [a,b] \to [c,d]$ is strictly increasing and continuous. Then for a Borel $f : [c,d] \to \overline{\mathbb{R}}$,*

$$\int_{[c,d]} f(y)\, dy = \int_{[a,b]} f \circ F(x)\, dF(x)\,.$$

Proof. If μ_F is the Riemann-Stieltjes measure of F, (3.5.11) shows

$$\int_{[a,b]} f \circ F(x)\, dF(x) = \int_{[c,d]} f\, d\mu_F \circ F^{-1}, \tag{3.5.12}$$

whenever either side is defined.

For $c \le \alpha < \beta \le d$,

$$\begin{aligned}
\mu_F \circ F^{-1}\left((\alpha,\beta]\right) &= \mu_F\left(F^{-1}(\alpha,\beta]\right) \\
(F \text{ is a bijection}) &= \mu_F\left(\left(F^{-1}(\alpha), F^{-1}(\beta)\right]\right) \\
&= F\left(F^{-1}(\beta)\right) - F\left(F^{-1}(\alpha)\right) \\
&= \beta - \alpha \\
&= \lambda\left((\alpha,\beta]\right),
\end{aligned}$$

where λ is the Lebesgue measure. Exc 3.5.9 shows $\mu_F \circ F^{-1}$ and λ agree on $[c,d]$. Hence the right hand side of (3.5.12) equals $\int_c^d f(y)\, dy$ and hence the proof follows. $\square$

The next theorem, which follows from Theorems 3.5.7 and 3.5.8, justifies the following substitution:

$$y = F(x), \quad dy = |F'(x)|\, dx\,.$$

Theorem 3.5.9 (Change of variables formula for Lebesgue integration)**.** *Suppose $U \subset \mathbb{R}$ and $V \subset \mathbb{R}$ are open sets and $F : U \to V$ is a differentiable bijection. Then for $f : V \to \overline{\mathbb{R}}$ Borel,*

$$\int_V f(y)\, dy = \int_U f \circ F(x) |F'(x)|\, dx \tag{3.5.13}$$

whenever either side is defined.

Proof. Let us first consider the case where $f \ge 0$, $U = (a,b)$ and $V = (c,d)$. For any $a < a' < b' < b$ and $[c',d'] = F([a,b])$, Theorems 3.5.7 and 3.5.8 show that

$$\int_{[c',d']} f(y)\, dy = \begin{cases} \int_{[a',b']} f \circ F(x) F'(x)\, dx, & F \text{ increasing}, \\ \int_{[a',b']} f \circ F(x) (-F'(x))\, dx, & F \text{ decreasing}\,. \end{cases}$$

In both the cases, the right hand side equals

$$\int_{[a',b']} f \circ F(x) |F'(x)|\, dx\,.$$

Letting $a' \downarrow a$ and $b' \uparrow b$ with the help of MCT implies

$$\int_a^b f \circ F(x)|F'(x)|\,dx = \int_c^d f(y)\,dy\,.$$

This proves (3.5.13) when $U = (a, b)$ and $V = (c, d)$. Since any open set in $\mathbb{R}$ is the union of countably many disjoint open intervals, (3.5.13) follows for open sets U and V when $f \ge 0$. Finally for a general f measurable, the standard argument of $f = f^+ - f^-$ proves it whenever either side of (3.5.13) is defined, which completes the proof. $\square$

Remark 3.5.1. *The formula* (3.5.13) *becomes wrong if the modulus sign on* $F'(x)$ *on the right hand side is removed.*

3.6 Expectation

In this subsection, $(\Omega, \mathcal{A}, P)$ is the probability space on which the random variables talked about are defined, unless specified otherwise.

Definition. *For a random variable* X*, its* expectation $\mathrm{E}(X)$ *is defined by*

$$\mathrm{E}(X) = \int_\Omega X\,dP,$$

whenever the right hand side is defined. As usual, X *is an* integrable random variable *if*

$$\int_\Omega |X|\,dP < \infty\,.$$

*The word "*mean*" is an often used synonym of "expectation".*

Theorem 3.6.1. *For a random variable* $X \ge 0$*,*

$$\mathrm{E}(X) = \int_0^\infty P(X > x)\,dx\,.$$

Proof. We first show this when X is finite and simple, that is, it takes values

$0 \le s_1 < \ldots < s_n < \infty$. In this case,

$$
\begin{aligned}
\int_0^\infty P(X > x)\, dx &= \int_0^{s_n} P(X > x)\, dx \\
(\text{put } s_0 = 0) &= \sum_{i=1}^n \int_{s_{i-1}}^{s_i} P(X > x)\, dx \\
&= \sum_{i=1}^n (s_i - s_{i-1}) P(X > s_{i-1}) \\
&= \sum_{i=1}^n (s_i - s_{i-1}) \sum_{j=i}^n P(X = s_j) \\
&= \sum_{j=1}^n P(X = s_j) \sum_{i=1}^j (s_i - s_{i-1}) \\
&= \sum_{j=1}^n s_j P(X = s_j) \\
&= \mathrm{E}(X)
\end{aligned}
$$

the last line following from the definition of integral for a simple non-negative function.

For $X \ge 0$ which is not necessarily simple or finite, there exist simple functions $0 \le X_n < \infty$ with $X_n \uparrow X$. Then for all x, $[X_n > x] \uparrow [X > x]$ and hence

$$P(X_n > x) \uparrow P(X > x),\ x \ge 0\,.$$

We have already shown

$$\mathrm{E}(X_n) = \int_0^\infty P(X_n > x)\, dx\,.$$

Letting $n \to \infty$ with the help of MCT completes the proof. $\square$

Exercise 3.6.1. *If X is a random variable whose expectation is defined, show that*

$$\mathrm{E}(X) = \int_0^\infty P(X > x)\, dx - \int_{-\infty}^0 P(X < x)\, dx\,.$$

Henceforth, all random variables X are assumed to be "proper", that is,

$$X \ne \pm\infty \text{ a.s.}$$

Here and elsewhere, "a.s." or "almost surely" simply means "with probability 1". In other words, almost everywhere and almost surely mean the same thing, except that the latter is used for a probability space.

Theorem 3.6.2. *For a random variable X and a Borel function $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathrm{E}\left(f(X)\right) = \int_{\mathbb{R}} f(x) P(X \in dx)\,.$$

Proof. Follows from (3.5.11) and the tautology

$$P \circ X^{-1}(A) = P(X \in A),\ A \in \mathcal{B}(\mathbb{R}),$$

a consequence of which is

$$\int_{\mathbb{R}} f(x) P(X \in dx) = \int_{\mathbb{R}} f(x) P \circ X^{-1}(dx)$$

whenever either side is defined. □

Definition. *For a random variable X, a Borel function $f : \mathbb{R} \to [0, \infty]$ is a* density *of X if*

$$P(X \in A) = \int_A f(x)\, dx \text{ for all } A \in \mathcal{B}(\mathbb{R}), \tag{3.6.1}$$

that is, if $f(x) = \frac{P(X \in dx)}{\lambda(dx)}$ where λ is the Lebesgue measure.

Theorem 3.6.3. *If a random variable X has a density f, then for any measurable $g : \mathbb{R} \to \mathbb{R}$,*

$$\mathrm{E}\left(g(X)\right) = \int_{\mathbb{R}} g(x) f(x)\, dx$$

whenever either side is defined.

Proof. Follows from (3.5.10) and Theorem 3.6.2. □

Theorem 3.6.4. *If the CDF F of a random variable X is differentiable, then F' is a density of X.*

Proof. Follows from Theorem 3.5.5. □

Exercise 3.6.2. *If X is a discrete random variable, that is there exists a countable set C such that $P(X \in C) = 1$, then show that X is integrable if and only if*

$$\sum_{x \in C} |x| P(X = x) < \infty,$$

and in that case

$$\mathrm{E}(X) = \sum_{x \in C} x P(X = x)\,.$$

Definition. *For a random variable X with* $\mathrm{E}(X^2) < \infty$*, its* variance *is defined as*

$$\mathrm{Var}(X) = \mathrm{E}\left[\big(X - \mathrm{E}(X)\big)^2\right].$$

The convention is to declare $\mathrm{Var}(X) = \infty$ *whenever* $\mathrm{E}(X^2) = \infty$.

Exercise 3.6.3. *For a random variable X with finite variance, show that*

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2,$$

where $\mu = \mathrm{E}(X)$.

Definition. *For random variables X and Y such that X, Y, XY are integrable, the* covariance *of X and Y is defined by*

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\right].$$

Exercise 3.6.4. *Show that*

$$|\mathrm{Cov}(X, Y)| \le \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}\,.$$

Hint. Use Theorem 3.3.2 (Cauchy-Schwarz inequality).

Exercise 3.6.5. *Show the following.*

1. *If* $\mathrm{Cov}(X, Y)$ *is defined, then*

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\,.$$

2. *If X has a finite variance,* $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

3. *If* $\mathrm{Cov}(X, Y)$ *is defined, then*

$$\mathrm{Cov}(\alpha X + \gamma, \beta Y + \delta) = \alpha\beta \mathrm{Cov}(X, Y)\,, \alpha, \beta, \delta, \gamma \in \mathbb{R}\,.$$

4. *If $X_1, \ldots, X_n$ have finite variances, then*

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j)\,.$$

4 Measures in higher dimensions

4.1 Riemann-Stieltjes measures on $\mathbb{R}^d$

Definition. *The* Borel σ-field on $\mathbb{R}^d$ *is defined by*

$$\mathcal{B}(\mathbb{R}^d) = \sigma\left(\{U \subset \mathbb{R}^d : U \text{ is open}\}\right).$$

Let $\mathcal{H}$ be the collection of all half-open (left open, right closed) bounded rectangles, that is,

$$\mathcal{H} = \{(a_1, b_1] \times \ldots \times (a_d, b_d] : -\infty < a_i < b_i < \infty \text{ for } i = 1, \ldots, d\} .$$

Exercise 4.1.1. *Check that*

$$\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R}^d) .$$

For a function $F : \mathbb{R}^d \to \mathbb{R}$ and $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$, define

$$\Delta_R F = \sum_{(x_1,\ldots,x_d)\in\{a_1,b_1\}\times\ldots\times\{a_d,b_d\}} (-1)^{\#\{i: x_i = a_i\}} F(x_1, \ldots, x_d) . \tag{4.1.1}$$

Definition. *A measure μ on $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$ is* Radon *if $\mu(K) < \infty$ for all compact $K \subset \mathbb{R}^d$.*

Theorem 4.1.1. *If $F : \mathbb{R}^d \to \mathbb{R}$ is a function which is continuous from above and satisfies $\Delta_R F \geq 0$ for all $R \in \mathcal{H}$, then there exists a unique Radon measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that*

$$\mu(R) = \Delta_R F \,, \ \textit{for all } R \in \mathcal{H} . \tag{4.1.2}$$

Proof. Let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy the assumptions, that is,

$$\lim_{y_1\downarrow x_1,\ldots,y_d\downarrow x_d} F(y_1, \ldots, y_d) = F(x_1, \ldots, x_d) \text{ for all } (x_1, \ldots, x_d) \in \mathbb{R}^d , \tag{4.1.3}$$

and

$$\Delta_R F \geq 0 \text{ for all } R \in \mathcal{H} . \tag{4.1.4}$$

Step 1. The function $R \mapsto \Delta_R F$ is a finitely additive set function on $\mathcal{H}$, that is, for disjoint $R_1, \ldots, R_n \in \mathcal{H}$ such that $R = R_1 \cup \ldots \cup R_n \in \mathcal{H}$,

$$\Delta_R F = \sum_{i=1}^n \Delta_{R_i} F .$$

Proof of Step 1. For $R = (a_1, b_1] \times (a_d, b_d] \in \mathcal{H}$, and $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, define

$$\operatorname{sgn}(x, R) = \begin{cases} (-1)^{\#\{i: x_i = a_i\}} , & x \in \{a_1, b_1\} \times \ldots \times \{a_d, b_d\} , \\ 0 , & \text{otherwise} . \end{cases}$$

That is, $\operatorname{sgn}(x, R)$ is zero unless x is a vertex of R.

Rewrite (4.1.1) as

$$\Delta_R F = \sum_{x=(x_1,\ldots,x_d)\in\{a_1,b_1\}\times\ldots\times\{a_d,b_d\}} \operatorname{sgn}(x, R) F(x) .$$

Suppose $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$ and for some $n_1, \ldots, n_d \in \mathbb{N}$,

$$a_i = a_{i,0} < a_{i,1} < \ldots < a_{i,n_i} = b_i \, , i = 1, \ldots, d \, .$$

Let

$$R_{k_1,\ldots,k_d} = \prod_{i=1}^{d} (a_{i,k_i-1}, a_{i,k_i}] \ , 1 \le k_1 \le n_1, \ldots, 1 \le k_d \le n_d \, . \tag{4.1.5}$$

We shall first show that

$$\sum_{k_1=1}^{n_1} \ldots \sum_{k_d=1}^{n_d} \Delta_{R_{k_1,\ldots,k_d}} F = \Delta_R F \, . \tag{4.1.6}$$

The LHS above equals

$$\sum_{x \in A} F(x) \sum_{k_1=1}^{n_1} \ldots \sum_{k_d=1}^{n_d} \operatorname{sgn}(x, R_{k_1,\ldots,k_d}) \ , \tag{4.1.7}$$

where $A = \prod_{i=1}^{d} \{a_{i,0}, a_{i,1}, \ldots, a_{i,n_i}\}$. Let $A_0 = \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}$ and observe that for $x \in A_0$, there exists unique $k_1, \ldots, k_d$ such that

$$\operatorname{sgn}(x, R_{k_1,\ldots,k_d}) \neq 0 \, ,$$

and for this $k_1, \ldots, k_d$,

$$\operatorname{sgn}(x, R_{k_1,\ldots,k_d}) = \operatorname{sgn}(x, R) \, .$$

Thus, the quantity in (4.1.7) equals

$$\sum_{x \in A_0} \operatorname{sgn}(x, R) F(x) + \sum_{x \in A \setminus A_0} F(x) \sum_{k_1=1}^{n_1} \ldots \sum_{k_d=1}^{n_d} \operatorname{sgn}(x, R_{k_1,\ldots,k_d}) \ .$$

Since the first term above is the same as $\Delta_R F$, (4.1.6) would follow once it is shown that

$$\sum_{k_1=1}^{n_1} \ldots \sum_{k_d=1}^{n_d} \operatorname{sgn}(x, R_{k_1,\ldots,k_d}) = 0 \, , x \in A \setminus A_0 \, . \tag{4.1.8}$$

Fix $x = (x_1, \ldots, x_d) \in A \setminus A_0$. Then there exists $i \in \{1, \ldots, d\}$ such that

$$x_i = a_{i,u_i} \text{ for some } 1 \le u_i \le n_i - 1 \, .$$

Thus for $1 \le k_1 \le n_1, \ldots, 1 \le k_d \le n_d$, x is not a vertex of $R_{k_1,\ldots,k_d}$ by (4.1.5), unless k_i equals either u_i or $u_i + 1$, that is,

$$\operatorname{sgn}(x, R_{k_1,\ldots,k_d}) = 0 \text{ if } k_i \notin \{u_i, u_i + 1\} \, .$$

Further,

$$\operatorname{sgn}\left(x, R_{k_1,\ldots,k_{i-1},u_i,k_{i+1},\ldots,k_d}\right) = -\operatorname{sgn}\left(x, R_{k_1,\ldots,k_{i-1},u_i+1,k_{i+1},\ldots,k_d}\right) .$$

Thus (4.1.8) follows which proves (4.1.6).

To complete the proof of Step 1, let $R_1, \ldots, R_n \in \mathcal{H}$ be disjoint such that $R = R_1 \cup \ldots \cup R_n \in \mathcal{H}$. Let $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$ and

$$a_i = a_{i,0} < a_{i,1} < \ldots < a_{i,n_i} = b_i\,, i = 1, \ldots, d\,,$$

be such that vertices of $R_1, \ldots, R_n$ belong to $\prod_{i=1}^d \{a_{i,0}, a_{i,1}, \ldots, a_{i,n_i}\}$. If $R_{k_1,\ldots,k_d}$ is as in (4.1.5), then

$$\text{either } R_{k_1,\ldots,k_d} \subset R_i \text{ or } R_{k_1,\ldots,k_d} \cap R_i = \emptyset\,,$$

for $1 \le k_1 \le n_1, \ldots, 1 \le k_d \le n_d$ and $i = 1, \ldots, n$. Use (4.1.6) to write

$$\begin{aligned}
\Delta_R F &= \sum_{1\le k_1\le n_1,\ldots,1\le k_d\le n_d} \Delta_{R_{k_1,\ldots,k_d}} F \\
&= \sum_{i=1}^n \sum_{1\le k_1\le n_1,\ldots,1\le k_d\le n_d:\, R_{k_1,\ldots,k_d}\subset R_i} \Delta_{R_{k_1,\ldots,k_d}} F \\
&= \sum_{i=1}^n \Delta_{R_i} F\,,
\end{aligned}$$

(4.1.6) being used again in the last line. This completes the proof of Step 1. ☐

Step 2. If $R_1, R_2 \in \mathcal{H}$ and $R_1 \subset R_2$, then $\Delta_{R_1} F \le \Delta_{R_2} F$.

Proof of Step 2. Follows from (4.1.4) and Step 1 by observing that $R_2 \setminus R_1 = S_1 \cup \ldots \cup S_n$ for some disjoint $S_1, \ldots, S_n \in \mathcal{H}$. ☐

Step 3. If $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$ and for $\varepsilon > 0$, $R_\varepsilon = (a_1, b_1 + \varepsilon] \times \ldots \times (a_d, b_d + \varepsilon]$, then

$$\lim_{\varepsilon\downarrow 0} \Delta_{R_\varepsilon} F = \Delta_R F\,.$$

Proof of Step 3. Follows from (4.1.3). ☐

Step 4. If $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$,

$$\lim_{a_1'\downarrow a_1,\ldots,a_d'\downarrow a_d} \Delta_{(a_1',b_1]\times\ldots\times(a_d',b_d]} F = \Delta_R F\,.$$

Proof of Step 4. Follows from (4.1.3). ☐

For the next several steps, fix $n = (n_1, \ldots, n_d) \in \mathbb{Z}^d$ and let

$$\Omega_n = (n_1 - 1, n_1] \times \ldots \times (n_d - 1, n_d] ,$$

and

$$\mathcal{S}_n = \{\emptyset\} \cup \{R \in \mathcal{H} : R \subset \Omega_n\} .$$

Step 5. The collection $\mathcal{S}_n$ is a semi-field on Ω_n and $\mu_n : \mathcal{S}_n \to [0, \infty)$ defined by

$$\mu_n(R) = \Delta_R F \, , \emptyset \neq R \in \mathcal{S}_n \, ,$$

and $\mu_n(\emptyset) = 0$ is a finitely additive set function.

Proof of Step 5. That $\mathcal{S}_n$ is a semi-field is immediate. Finite additivity of μ_n follows from Step 1. □

Step 6. Let $\mathcal{F}_n = \{A_1 \cup \ldots \cup A_k : A_1, \ldots, A_k \in \mathcal{S}_n \text{ are disjoint}\}$. Then $\mathcal{F}_n$ is a field on Ω_n. Extend μ_n to $\mathcal{F}_n$ by

$$\mu_n(A_1 \cup \ldots \cup A_k) = \sum_{i=1}^{k} \mu_n(A_i) \, , A_1, \ldots, A_k \in \mathcal{S}_n \text{ are disjoint} \, .$$

Then μ_n is well defined on $\mathcal{F}_n$, that is, different representations yield the same definition, is finitely additive on $\mathcal{F}_n$, monotone on $\mathcal{F}_n$, finitely sub-additive on $\mathcal{F}_n$ and countably super-additive on $\mathcal{F}_n$.

Proof of Step 6. Follows from Step 5 and Theorems 2.4.2 and 2.4.3. □

Step 7. The set function μ_n is countably additive on $\mathcal{S}_n$.

Proof of Step 7. Let $R_1, R_2, \ldots \in \mathcal{S}_n$ be disjoint such that

$$R = R_1 \cup R_2 \cup \ldots \in \mathcal{S}_n \, .$$

In Step 6, μ_n is shown to be countably super-additive, that is,

$$\mu_n(R) \geq \sum_{i=1}^{\infty} \mu_n(R_i) \, .$$

Thus, countable additivity would follow once it is shown that

$$\mu_n(R) \leq \sum_{i=1}^{\infty} \mu_n(R_i) \, . \tag{4.1.9}$$

Let $R = (a_1, b_1] \times \ldots \times (a_d, b_d]$ and for $i = 1, 2, \ldots$,

$$R_i = (a_{i,1}, b_{i,1}] \times \ldots \times (a_{i,d}, b_{i,d}] \, .$$

Fix $\delta > 0$. Use Step 3 to get $\varepsilon_i > 0$ such that $\Delta_{\tilde{R}_i} F \leq \Delta_{R_i} F + 2^{-i}\delta$ where

$$\tilde{R}_i = (a_{i,1}, b_{i,1} + \varepsilon_i] \times \ldots \times (a_{i,d}, b_{i,d} + \varepsilon_i] \,.$$

Fix $a'_i \in (a_i, b_i)$ for $i = 1, \ldots, d$. Since

$$[a'_1, b_1] \times \ldots \times [a'_d, b_d] \subset R = \bigcup_{i=1}^{\infty} R_i \subset \bigcup_{i=1}^{\infty} (a_{i,1}, b_{i,1} + \varepsilon_i) \times \ldots \times (a_{i,d}, b_{i,d} + \varepsilon_i) \,,$$

the Heine-Borel theorem implies

$$[a'_1, b_1] \times \ldots \times [a'_d, b_d] \subset \bigcup_{i=1}^{k} (a_{i,1}, b_{i,1} + \varepsilon_i) \times \ldots \times (a_{i,d}, b_{i,d} + \varepsilon_i)$$

for some finite k. Letting $R' = (a'_1, b_1] \times \ldots \times (a'_d, b_d]$, it follows that

$$R' \subset \Omega_n \cap \left(\tilde{R}_1 \cup \ldots \cup \tilde{R}_k \right) \,.$$

Monotonicity and finite sub-additivity of μ_n shown in Step 6 implies

$$\begin{aligned} \mu_n(R') &\leq \sum_{i=1}^{k} \mu_n(\tilde{R}_i \cap \Omega_n) \\ &= \sum_{i=1}^{k} \Delta_{\tilde{R}_i \cap \Omega_n} F \\ \text{(Step 2)} &\leq \sum_{i=1}^{k} \Delta_{\tilde{R}_i} F \\ \text{(choice of } \varepsilon_i) &\leq \sum_{i=1}^{\infty} \left(\Delta_{R_i} F + 2^{-i}\delta \right) \\ &= \delta + \sum_{i=1}^{\infty} \mu_n(R_i) \,. \end{aligned}$$

Since δ is arbitrary, it follows that

$$\mu_n(R') \leq \sum_{i=1}^{\infty} \mu_n(R_i) \,.$$

Letting $a'_1 \downarrow a_1, \ldots, a'_d \downarrow a_d$ and using Step 4, (4.1.9) follows. This completes the proof of Step 7. □

Step 8. The set function μ_n can be extended to a measure on $(\Omega_n, \sigma(\mathcal{S}_n))$.

Proof of Step 8. Follows from Step 7 above and Corollary 2.4.1 of the extension theorem of Carathéodory. □

Step 9. If

$$\mu(A) = \sum_{n \in \mathbb{Z}^d} \mu_n(A \cap \Omega_n)\,, A \in \mathcal{B}(\mathbb{R}^d)\,,$$

then μ is a Radon measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying

$$\mu(R) = \Delta_R F\,, R \in \mathcal{H}\,. \tag{4.1.10}$$

Proof of Step 9. As μ_n is a measure on $(\Omega_n, \sigma(\mathcal{S}_n))$ for each $n \in \mathbb{Z}^d$ by Step 8 and $(\Omega_n : n \in \mathbb{Z}^d)$ is a partition of $\mathbb{R}^d$, μ defined above is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For $R \in \mathcal{H}$, as R is bounded and non-empty, there exist $n_1, \ldots, n_k \in \mathbb{Z}^d$ such that $R \cap \Omega_{n_i} \neq \emptyset$ for $i = 1, \ldots, k$ and $R \subset \Omega_{n_1} \cup \ldots \cup \Omega_{n_k}$. Thus,

$$\begin{aligned} \mu(R) &= \sum_{i=1}^{k} \mu_{n_i}(R \cap \Omega_{n_i}) \\ (\emptyset \neq R \cap \Omega_{n_i} \in \mathcal{S}_{n_i}) &= \sum_{i=1}^{k} \Delta_{R \cap \Omega_{n_i}} F \\ (\text{Step 1}) &= \Delta_R F\,, \end{aligned}$$

showing (4.1.10). To see that μ is Radon, for any compact set $K \subset \mathbb{R}^d$, there exists $n \in \mathbb{N}$ such that $R = (-n, n]^d \supset K$. Thus

$$\mu(K) \le \mu(R) = \Delta_R F\,,$$

by (4.1.10). This shows μ is a Radon measure and completes the proof of Step 9. □

Step 10. The measure μ is the only measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying (4.1.10).

Proof of Step 10. Suppose μ' is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that (4.1.10) holds with μ replaced by μ'. Then μ and μ' agree on $\mathcal{H}$, and hence on

$$\mathcal{S} = \left\{ \mathbb{R}^d \cap \prod_{i=1}^{d} (a_i, b_i] : -\infty \le a_i \le b_i \le \infty \right\}\,,$$

because for every set in $\mathcal{S}$ there exist sets in $\mathcal{H}$ increasing to the former. Further, μ and μ' are σ-finite on $\mathcal{H}$ and hence on $\mathcal{S}$ which is a semi-field that generates $\mathcal{B}(\mathbb{R}^d)$. Corollary 2.4.1 shows μ and μ' agree on $\mathcal{B}(\mathbb{R}^d)$, as claimed in Step 10. □

Steps 9 and 10 complete the proof of the fact. □

Definition. *For a function $F : \mathbb{R}^d \to \mathbb{R}$ satisfying the hypotheses of Theorem 4.1.1, the unique measure μ on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *satisfying* (4.1.2) *is the* Riemann-Stieltjes measure *of F.*

Remark 4.1.1. *A function* F *satisfying* (4.1.3) *and* (4.1.4) *is not necessarily monotonic. For example,* $F : \mathbb{R}^2 \to \mathbb{R}$ *defined by*

$$F(x, y) = xy\,,$$

satisfies (4.1.3) *and* (4.1.4)*, though* F *is not monotonic because*

$$F(0, 0) = 0 < F(1, 1) = F(-1, -1) = 1\,.$$

That is, $x_1 \leq x_2$ *and* $y_1 \leq y_2$ *imply neither* $F(x_1, y_1) \leq F(x_2, y_2)$ *nor* $F(x_1, y_1) \geq F(x_2, y_2)$*.*

Theorem 4.1.2. *For* $F : \mathbb{R}^d \to [0, 1]$*, there exists a probability measure* P *on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *such that*

$$P\left(\prod_{i=1}^{d}(-\infty, x_i]\right) = F(x) \text{ for all } x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{4.1.11}$$

if and only if F *is continuous from above,* $\Delta_R F \geq 0$ *for all* $R \in \mathcal{H}$*,*

$$\lim_{x_1 \to \infty, \ldots, x_d \to \infty} F(x_1, \ldots, x_d) = 1 \tag{4.1.12}$$

and

$$\lim_{x_{i_1} \to -\infty, \ldots, x_{i_k} \to -\infty} F(x_1, \ldots, x_d) = 0 \tag{4.1.13}$$

for all $\emptyset \neq \{i_1, \ldots, i_k\} \subset \{1, \ldots, d\}$ *and fixed* $x_j \in \mathbb{R}$ *for all* $j \in \{1, \ldots, d\} \setminus \{i_1, \ldots, i_k\}$*.*

Proof. For the 'if' part, suppose F is continuous from above, $\Delta_R F \geq 0$ for all $R \in \mathcal{H}$ and (4.1.12) and (4.1.13) hold. Theorem 4.1.1 shows there exists a measure P on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$P(R) = \Delta_R F,\ R \in \mathcal{H}\,.$$

Fix $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and put $R_n = (x_1 - n, x_1] \times \ldots \times (x_d - n, x_d]$ for $n = 1, 2, \ldots$ to get

$$P(R_n) = \Delta_{R_n} F\,.$$

Letting $n \to \infty$, (4.1.13) shows the right hand side converges to $F(x)$. Since $R_n \uparrow \prod_{i=1}^{d}(-\infty, x_i]$, (4.1.11) holds. Putting $x_1 = \ldots = x_d = n$ in (4.1.11) and letting $n \to \infty$, (4.1.12) shows

$$P(\mathbb{R}^d) = 1\,.$$

Hence P is a probability measure, that is, the 'if' part follows.

Conversely, suppose a probability measure P satisfying (4.1.11) exists. Observing that for $R \in \mathcal{H}$,

$$\begin{aligned}
P(R) &= \int_{\mathbb{R}^d} \prod_{i=1}^{d} \mathbf{1}_{(a_i,b_i]}(x_i)\, dP(x_1,\ldots,x_d) \\
&= \int_{\mathbb{R}^d} \prod_{i=1}^{d} \left(\mathbf{1}_{(-\infty,b_i]}(x_i) - \mathbf{1}_{(-\infty,a_i]}(x_i)\right)\, dP(x_1,\ldots,x_d) \\
&= \Delta_R F,
\end{aligned}$$

it follows that $\Delta_R F \ge 0$ for all $R \in \mathcal{H}$. For $x^{(n)} = (x_1^{(n)},\ldots,x_d^{(n)}) \in \mathbb{R}^d$ for $n = 1,\ldots,\infty$ such that

$$x_i^{(n)} \downarrow x_i^{(\infty)},\ n \to \infty,\ i = 1,\ldots,d,$$

it holds that

$$\prod_{i=1}^{d} \left(-\infty, x_i^{(n)}\right] \downarrow \prod_{i=1}^{d} \left(-\infty, x_i^{(\infty)}\right] .$$

Since P is a finite measure, it follows that

$$F(x^{(n)}) \downarrow F(x^{(\infty)}) .$$

This shows continuity from above of F. Finally, (4.1.12) and (4.1.13) follow by similar arguments. Hence the proof follows. $\square$

Definition. *For random variables $X_1,\ldots,X_d$ defined on some probability space $(\Omega,\mathcal{A},P)$, their joint C.D.F. is defined by*

$$F(x) = P\left(X_1 \le x_1,\ldots,X_d \le x_d\right),\ x = (x_1,\ldots,x_d) \in \mathbb{R}^d .$$

Theorem 4.1.3. *If $F : \mathbb{R}^d \to [0,1]$ is right continuous satisfying $\Delta_R F \ge 0$ for all $R \in \mathcal{H}$, (4.1.12) and (4.1.13), then there exist random variables $X_1,\ldots,X_d$ defined on some probability space whose joint C.D.F. is F.*

Proof. Theorem 4.1.2 shows the existence of a probability measure P on $\mathbb{R}^d$ satisfying (4.1.11). Let $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$ and define $X_1,\ldots,X_d$ on Ω by

$$X_i(x) = x_i,\ x = (x_1,\ldots,x_d) \in \mathbb{R}^d,$$

for $i = 1,\ldots,d$. It is trivial to check that F is the joint C.D.F. of $X_1,\ldots,X_d$. $\square$

Exercise 4.1.2. *Suppose F is the C.D.F. of $X_1,\ldots,X_d$. Prove or disprove that for $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$,*

$$P(X_1 = x_1,\ldots,X_d = x_d) = 0$$

if and only if F is continuous at x.

4.2 The Lebesgue measure on $\mathbb{R}^d$

Definition. *The Riemann-Stieltjes measure of F, which is defined on $\mathbb{R}^d$ by*

$$F(x_1, \ldots, x_d) = \prod_{i=1}^d x_i,\ (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

is the Lebesgue measure on $\mathbb{R}^d$ *and is usually denoted by λ_d or λ.*

Throughout this subsection, λ denotes the Lebesgue measure on $\mathbb{R}^d$, where d is fixed once and for all, and hence the subscript d in λ_d is unnecessary. The following definition extends the definition of density given by (3.6.1).

Definition. *For an $\mathbb{R}^d$-valued random variable $X = (X_1, \ldots, X_d)$, a Borel measurable function $f : \mathbb{R}^d \to [0, \infty]$ is a* density *of X or a* joint density *of $(X_1, \ldots, X_d)$ if*

$$P(X \in A) = \int_A f(x)\, \lambda(dx) \text{ for all } A \in \mathcal{B}(\mathbb{R}^d)\,. \tag{4.2.1}$$

Exercise 4.2.1. *Given an $\mathbb{R}^d$-valued random variable $X = (X_1, \ldots, X_d)$, show that a Borel function $f : \mathbb{R}^d \to [0, \infty)$ is a density of X if*

$$P\left(X_1 \le x_1, \ldots, X_d \le x_d\right) = \int_{(-\infty, x_1] \times \ldots \times (-\infty, x_d]} f(y)\, \lambda(dy)\,.$$

We shall denote by R^0 and $\overline{R}$ are the interior and closure of R, respectively.

Exercise 4.2.2. *For $R = \prod_{i=1}^d (a_i, b_i] \in \mathcal{H}$, show that*

$$\lambda(R) = \lambda(R^0) = \lambda(\overline{R}) = \prod_{i=1}^d (b_i - a_i)\,.$$

Throughout this subsection and elsewhere, $x \in \mathbb{R}^d$ is to be interpreted as a $d \times 1$ matrix, that is, a column vector of length d. The following is the main result of this subsection.

Theorem 4.2.1. *Suppose $f : \mathbb{R}^d \to \mathbb{R}^d$ is defined by*

$$f(x) = Ax + b$$

where A is a $d \times d$ non-singular matrix and $b \in \mathbb{R}^d$, then

$$\lambda\left(f(B)\right) = |\det(A)| \lambda(B) \text{ for all } B \in \mathcal{B}(\mathbb{R}^d)\,. \tag{4.2.2}$$

Proceeding towards the proof, we start with defining "vertices" of a rectangle.

Definition. *For* $R = \prod_{i=1}^d (a_i, b_i] \in \mathcal{H}$, *the* set of its vertices $v(R)$ *is defined by*

$$v(R) = \prod_{i=1}^d \{a_i, b_i\}\,.$$

The proof of Theorem 4.2.1 relies on the following few lemmas, the first of which is stated in more generality than is immediately needed, for a later requirement.

Lemma 4.2.1. *Suppose* $U \subset \mathbb{R}^d$ *is an open set. If* μ_1, μ_2 *are Radon measures on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *such that* $\mu_1(U^c) = \mu_2(U^c) = 0$ *and*

$$\mu_1(R) \le \mu_2(R) \text{ for all } R \in \mathcal{H} \text{ for which } \overline{R} \subset U \text{ and } v(R) \in \mathbb{Q}^d,$$

then $\mu_1(B) \le \mu_2(B)$ *for all* $B \in \mathcal{B}(\mathbb{R}^d)$.

Proof. We shall first show that the given assumption implies

$$\mu_1(R) \le \mu_2(R) \text{ for all } R \in \mathcal{H} \text{ with } v(R) \in \mathbb{Q}^d\,. \tag{4.2.3}$$

Fix $R = \prod_{i=1}^d (a_i, b_i] \in \mathcal{H}$ with $v(R) \in \mathbb{Q}^d$. Let $R_{1,1}, \ldots R_{1,2^d}$ be the 2^d rectangles in $\mathcal{H}$ obtained by bisecting each side of R_1, that is, for each $j = 1, \ldots, 2^d$,

$$R_{1,j} = \prod_{i=1}^d (\alpha_i, \beta_i],$$

where for every $i = 1, \ldots, d$, either $\alpha_i = a_i$ and $\beta_i = (a_i+b_i)/2$ or $\beta_i = b_i$ and $\alpha_i = (a_i + b_i)/2$. For $n \ge 2$, let $R_{n,1}, \ldots, R_{n,2^{dn}}$ be the rectangles obtained by bisecting the sides of $R_{n-1,1}, \ldots, R_{n-1,2^{(n-1)d}}$. Define

$$B_n = \bigcup_{1 \le j \le 2^{dn} :\, \overline{R_{n,j}} \subset U} R_{n,j},\ n \ge 1\,.$$

The given hypothesis, along with the observation $v(R_{nj}) \in \mathbb{Q}^d$, implies

$$\mu_1(B_n) \le \mu_2(B_n),\ n \ge 1\,.$$

Since U is an open set and the diameter of $\overline{R_{n,j}}$ is 2^{-dn} times the diameter of R, it is immediate that

$$B_n \uparrow R \cap U\,.$$

Continuity of measures from below implies

$$\mu_1(R \cap U) \le \mu_2(R \cap U),$$

which in conjunction with the assumption that μ_1, μ_2 are supported on U shows (4.2.3).

Fix $N \in \mathbb{N}$ and define

$$\mathcal{S} = \left\{ \prod_{i=1}^{d} (a_i, b_i] : -N \le a_i \le b_i \le N \text{ and for all } 1 \le i \le d,\, a_i, b_i \in \mathbb{Q} \right\}.$$

Thinking of $\mathcal{S}$ as a collection of subsets of $\Omega_N := (-N, N]^d$, it is immediate that $\mathcal{S}$ is a semi-field on Ω_N. Since the restrictions of μ_1, μ_2, which are Radon measures, to Ω_N is finite, with the help of standard measure theoretic arguments using the monotone class theorem, (4.2.3) implies

$$\mu_1(B) \le \mu_2(B) \text{ for all } B \in \mathcal{B}(\mathbb{R}^d),\, B \subset (-N, N]^d.$$

Finally, the observation

$$\lim_{N \to \infty} \mu_i \left(B \cap (-N, N]^d \right) = \mu_i(B),\, i = 1, 2,$$

completes the proof. $\square$

An immediate corollary of the above lemma is the following.

Corollary 4.2.1. *If μ_1, μ_2 are Radon measures on $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$ such that*

$$\mu_1(R) = \mu_2(R) \textit{ for all } R \in \mathcal{H} \textit{ with } v(R) \in \mathbb{Q}^d,$$

then μ_1, μ_2 agree on $\mathcal{B}(\mathbb{R}^d)$.

Proceeding towards the proof of Theorem 4.2.1, the following lemma shows that like in 1-dimension, the Lebesgue measure on $\mathbb{R}^d$ is translation invariant.

Lemma 4.2.2. *For $B \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$,*

$$\lambda(B + x) = \lambda(B),$$

where $B + x = \{y + x : y \in B\}$.

Proof. Follows from Exc 4.2.2 and Corollary 4.2.1. $\square$

The next lemma, taking us closer to the proof of Theorem 4.2.1, tells us that it suffices to check (4.2.2) for one Borel set B with $0 < \lambda(B) < \infty$.

Lemma 4.2.3. *If f and A are as in the statement of Theorem 4.2.1, then there exists $\theta \in \mathbb{R}$ such that*

$$\lambda\left(f(B)\right) = \theta \lambda(B) \textit{ for all } B \in \mathcal{B}(\mathbb{R}^d).$$

Proof. Define
$$\theta = \lambda\left(f\left((0,1]^d\right)\right).$$
Setting
$$\mu_1(B) = \lambda(f(B)),\ B \in \mathcal{B}(\mathbb{R}^d),$$
which is a measure because f is a bijection, and
$$\mu_2(B) = \theta\lambda(B),\ B \in \mathcal{B}(\mathbb{R}^d),$$
the claim would follow from Corollary 4.2.1 once it is shown that
$$\lambda\left(f(R)\right) = \theta\lambda(R) \text{ for all } R \in \mathcal{H} \text{ with } v(R) \in \mathbb{Q}^d. \tag{4.2.4}$$

For any $B \in \mathcal{B}(\mathbb{R}^d)$ and $y \in \mathbb{R}^d$,
$$\begin{aligned} f(B+y) &= \{A(x+y)+b : x \in B\} \\ &= \{Ax : x \in B\} + (Ay+b) \\ &= f(B) + Ay. \end{aligned}$$
Lemma 4.2.2 shows
$$\lambda\left(f(B+y)\right) = \lambda\left(f(B)\right). \tag{4.2.5}$$

Since for $n = 1, 2, \ldots$,
$$\left(0, \frac{1}{n}\right]^d + x = \prod_{i=1}^d \left(x_i, x_i + \frac{1}{n}\right],\ x = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$
(4.2.5) implies
$$\lambda\left(f\left(\left(0, \frac{1}{n}\right]^d\right)\right) = \lambda\left(f\left(\prod_{i=1}^d \left(x_i, x_i + \frac{1}{n}\right]\right)\right). \tag{4.2.6}$$
Recalling that f is a bijection,
$$(0,1]^d = \bigcup_{i_1,\ldots,i_d=1}^n \prod_{j=1}^d \left(\frac{i_j}{n}, \frac{i_j}{n} + \frac{1}{n}\right],$$
and the sets on the right hand side above are disjoint, the definition of θ implies
$$\lambda\left(f\left(\left(0, \frac{1}{n}\right]^d\right)\right) = n^{-d}\theta. \tag{4.2.7}$$

Additivity of measure implies that for $m_1, \ldots, m_d \in \mathbb{N}$,

$$\begin{aligned}
\lambda\left(f\left(\prod_{i=1}^d \left(0, \frac{m_i}{n}\right]\right)\right) &= \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \lambda\left(f\left(\prod_{i=1}^d \left(\frac{m_{i_j}-1}{n}, \frac{m_{i_j}}{n}\right]\right)\right) \\
\text{(by (4.2.6) and (4.2.7))} &= \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} n^{-d}\theta \\
&= n^{-d}\theta m_1 \ldots m_d \\
&= \theta \prod_{i=1}^d \frac{m_i}{n}\,.
\end{aligned}$$

In other words,

$$\lambda\left(f\left(\prod_{i=1}^d (0, r_i]\right)\right) = \theta\lambda\left(\prod_{i=1}^d (0, r_i]\right),\ r_1, \ldots, r_d \in \mathbb{Q} \cap (0, \infty)\,.$$

Using (4.2.5) and Lemma 4.2.2 once again establishes (4.2.4), from which, the proof follows. $\square$

Now we are in a position to prove Theorem 4.2.1. The following proof is due to Swapnaneel Bhattacharyya.

Proof of Theorem 4.2.1. Let $GL(d, \mathbb{R})$ be the collection of all $d \times d$ non-singular matrices. Lemma 4.2.3 shows that for all $A \in GL(d, \mathbb{R})$, there exists $\phi(A) \in \mathbb{R}$ such that for all $b \in \mathbb{R}^d$,

$$\lambda\left(\{Ax + b : x \in B\}\right) = \phi(A)\lambda(B) \text{ for all } A \in GL(d, \mathbb{R}),\ B \in \mathcal{B}(\mathbb{R}^d)\,. \tag{4.2.8}$$

To complete the proof, all that needs to be shown is $\phi(A) = |\det(A)|$ for all $A \in GL(d, \mathbb{R})$.

An immediate observation is that

$$\phi(A_1 A_2) = \phi(A_1)\phi(A_2),\ A_1, A_2 \in GL(d, \mathbb{R})\,. \tag{4.2.9}$$

Thus the proof would follow once it can be shown that

$$\phi(D) = \det(D) \text{ for a positive definite diagonal matrix } D, \tag{4.2.10}$$

$$\phi(P) = 1 \text{ for an orthogonal matrix } P, \tag{4.2.11}$$

and that any $A \in GL(d, \mathbb{R})$ can be written as

$$A = P_1 D P_2 \tag{4.2.12}$$

where P_1, P_2 are orthogonal matrices and D is a p.d. diagonal matrix.

If D is a $d \times d$ diagonal matrix whose i-th diagonal entry is $c_i > 0$, then it is immediate that

$$\left\{Dx : x \in (0,1]^d\right\} = \prod_{i=1}^{d}(0, c_i]\,.$$

Putting $B = (0,1]^d$ in (4.2.8) thus shows

$$\phi(D) = c_1 \ldots c_d,$$

(4.2.10) follows from which. If P is a $d \times d$ orthogonal matrix, then

$$\{Px : x \in \mathbb{R}^d,\ \|x\| \le 1\} = \{x \in \mathbb{R}^d : \|x\| \le 1\},$$

$\|\cdot\|$ being the usual L^2-norm on $\mathbb{R}^d$. Taking B to be the set on the right hand side above, putting it in (4.2.8) and using the fact $0 < \lambda(B) < \infty$, (4.2.11) follows.

Finally for (4.2.12), the spectral theorem implies

$$AA' = Q_1 W Q_1'$$

for some orthogonal matrix Q_1 and p.d. diagonal matrix W. Let D be the diagonal matrix whose entries are the positive square root of the corresponding entries of W. Define

$$C = Q_1 D Q_1'$$

which is obviously non-singular, and

$$Q_2 = C^{-1}A\,. \tag{4.2.13}$$

Since C is a symmetric matrix, it follows from the above that

$$\begin{aligned}
Q_2 Q_2' &= C^{-1}AA'C^{-1} \\
\text{(follows from the definition of } C) &= Q_1 D^{-1} Q_1' (AA') Q_1 D^{-1} Q_1' \\
\text{(choice of } Q_1, W) &= Q_1 D^{-1} Q_1' (Q_1 W Q_1') Q_1 D^{-1} Q_1' \\
\left(Q_1' Q_1 = I\right) &= Q_1 D^{-1} W D^{-1} Q_1' \\
\left(Q_1' Q_1 = I,\ W = D^2\right) &= I,
\end{aligned}$$

showing Q_2 is an orthogonal matrix. Rewrite (4.2.13) as

$$\begin{aligned}
A &= CQ_2 \\
&= Q_1 D Q_1' Q_2\,.
\end{aligned}$$

Setting $P_1 = Q_1$ and $P_2 = Q_1' Q_2$, and observing P_1, P_2 are orthogonal matrices and D is diagonal and p.d., (4.2.12) follows.

Combine (4.2.9) and (4.2.12) to write

$$\begin{aligned}\phi(A) &= \phi(P_1)\phi(D)\phi(P_1) \\ \text{(by (4.2.10) and (4.2.11))} &= \det(D) \\ &= |\det(A)|,\end{aligned}$$

the last line following from (4.2.12) and the facts $\det(UV) = \det(U)\det(V)$ and $\det(P) = \pm 1$ if P is orthogonal. Hence the proof follows. $\square$

Exercise 4.2.3. *Show that $\lambda(V + x) = 0$ if V is a subspace of $\mathbb{R}^d$ with dimension at most $d - 1$.*

4.3 The Jacobian Theorem

For stating the result, a Jacobian matrix has to be first defined. Consider an open set $U \subset \mathbb{R}^d$ and a function $F : U \to \mathbb{R}^d$. Denote by $f_1, \ldots, f_d$ the coordinate functions of F, that is,

$$F(x) = \big(f_1(x), \ldots, f_d(x)\big)\,, x \in \mathbb{R}^d\,.$$

If the first partial derivatives of F exist, that is, $\partial f_i(x)/\partial x_j$ exists for all $x \in U$ and $1 \le i, j \le d$, then its Jacobian matrix at x, denoted by $J(x)$, is a $d \times d$ matrix defined by

$$J(x) = \left(\frac{\partial f_i(x)}{\partial x_j}\right)_{1 \le i,j \le d}\,, x \in U\,,$$

that is, the (i, j)-th entry of $J(x)$ is $\partial f_i(x)/\partial x_j$. The statement of the theorem is the following, of which, the one-dimensional change of variables formula is a special case. As in 1-dimension, the integral of a Borel function $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ with respect to λ is denoted by

$$\int f(x)\, dx,$$

whenever it is defined.

Theorem 4.3.1 (The Jacobian theorem)**.** *For open subsets U and V of $\mathbb{R}^d$, let $T : U \to V$ be a bijection which is continuously differentiable, that is, the first partial derivatives of T exist and are continuous. Assume that its Jacobian matrix $J(x)$ is non-singular for all $x \in U$. Then for any measurable function $f : V \to \overline{\mathbb{R}}$,*

$$\int_U f\big(T(x)\big)\, |\det(J(x))|\; dx = \int_V f(y)\, dy\,,$$

$\det(A)$ *denoting the determinant of A for any square matrix A, whenever the integral on either side is defined.*

Putting $d = 1$ in the above theorem yields the change of variable formula in one dimension, which is the following.

Corollary 4.3.1. *Suppose that $-\infty \le a < b \le \infty$, $-\infty \le c < d \le \infty$ and $T : (a, b) \to (c, d)$ is a C^1 bijection whose derivative T' never vanishes. Then*

$$\int_a^b f\left(T(x)\right) |T'(x)| \, dx = \int_c^d f(y) \, dy \,. \tag{4.3.1}$$

The above corollary simply justifies that in the substitution $y = T(x)$, we need to put $dy = |T'(x)| \, dx$. The formula (4.3.1) becomes incorrect if the modulus on $T'(x)$ is removed. Corollary 4.3.1 is slightly weaker than Theorem 3.5.9 because the continuity of the derivative is not assumed in the latter.

The following facts from multivariable analysis are needed for the proof. The first one is the inverse function theorem.

Fact 4.3.1. *Let $U \subset \mathbb{R}^d$ be an open set and $T : U \to \mathbb{R}^d$ be continuously differentiable. Denoting by $J(x)$ the Jacobian matrix of T at $x \in U$, assume that $J(x_0)$ is non-singular for some $x_0 \in U$. Then, there exists an open neighbourhood X of x_0 such that T is one-one on X, the set $T(X)$ is open, T^{-1} is continuously differentiable on $T(X)$ and the Jacobian matrix of T^{-1} at y is $\left(J(T^{-1}y)\right)^{-1}$ for all $y \in T(X)$.*

The following is another fact from multivariable analysis which essentially follows from the one-dimensional mean value theorem.

Fact 4.3.2. *Suppose that $U \subset \mathbb{R}^d$ is open, $R \subset U$ is a compact rectangle and $T : U \to \mathbb{R}^d$ is continuously differentiable such that*

$$|J_{ij}(y) - J_{ij}(x)| \le \alpha \,, x, y \in R \,, 1 \le i, j \le d \,,$$

where $J_{ij}(z)$ is the (i, j)-th entry of the Jacobian matrix $J(z)$ of T at z for all $z \in U$ and $1 \le i, j \le d$. Then,

$$\|T(x) - T(y) - J(x)(x - y)\| \le d\alpha \|x - y\| \,, x, y \in R \,,$$

where $\| \cdot \|$ is the L^∞ norm on $\mathbb{R}^d$ defined by

$$\|x\| = \max_{1 \le i \le d} |x_i| \,, x = (x_1, \ldots, x_d) \in \mathbb{R}^d \,, \tag{4.3.2}$$

$x, y, T(x), T(y)$ are viewed as $d \times 1$ vectors and hence $J(x)(x - y)$ is also a $d \times 1$ vector.

Proof of Fact 4.3.2. Since $\| \cdot \|$ is the L^∞ norm, it suffices to show that the absolute value of each entry of the $d \times 1$ vector $T(x) - T(y) - J(x)(x - y)$

is at most $d\alpha\|x-y\|$. In other words, it suffices to show that if $f:U\to\mathbb{R}$ is continuously differentiable, and

$$|f_i(y)-f_i(x)|\le\alpha\,,x,y\in R,\ i=1,\ldots,d\,,$$

where

$$f_i(x)=\frac{\partial f(x)}{\partial x_i}\,,x\in U,\ i=1,\ldots,d\,,$$

then

$$\left|f(x)-f(y)-\sum_{i=1}^d f_i(x)(x_i-y_i)\right|\le d\alpha\|x-y\|\,,x,y\in R\,.$$

Let f be a function satisfying the hypotheses. Let $x^0=x$, $x^d=y$, and for $1\le i\le d-1$,

$$x^i=(y_1,\ldots,y_i,x_{i+1},\ldots,x_d)\ .$$

Since R is a rectangle, $x^1,\ldots,x^{d-1}\in R$. For a fixed $i=1,\ldots,d$, x^{i-1} and x^i have all entries identical except the i-th one, which are x_i and y_i respectively. The one-dimensional mean value theorem implies there exists ξ_i between x_i and y_i such that

$$f\left(x^{i-1}\right)-f\left(x^i\right)=(x_i-y_i)f_i(y_1,\ldots,y_{i-1},\xi_i,x_{i+1},\ldots,x_d)\,.$$

Since $\tilde{\xi}_i=(y_1,\ldots,y_{i-1},\xi_i,x_{i+1},\ldots,x_d)\in R$ because R is a rectangle, the hypotheses on f imply

$$\left|f_i(\tilde{\xi}_i)-f_i(x)\right|\le\alpha\,,i=1,\ldots,d\,.$$

Therefore,

$$\begin{aligned}
&\left|f(x)-f(y)-\sum_{i=1}^d f_i(x)(x_i-y_i)\right|\\
&=\left|f\left(x^0\right)-f\left(x^d\right)-\sum_{i=1}^d f_i(x)(x_i-y_i)\right|\\
&=\left|\sum_{i=1}^d\left[f\left(x^{i-1}\right)-f\left(x^i\right)\right]-\sum_{i=1}^d f_i(x)(x_i-y_i)\right|\\
&=\left|\sum_{i=1}^d\Big(f_i(\tilde{\xi}_i)-f_i(x)\Big)(x_i-y_i)\right|\\
&\le\sum_{i=1}^d\left|f_i(\tilde{\xi}_i)-f_i(x)\right||x_i-y_i|\\
&\le d\alpha\max_{1\le i\le d}|x_i-y_i|\\
&=d\alpha\|x-y\|\,.
\end{aligned}$$

This completes the proof. □

Proof of Theorem 4.3.1

Without loss of generality, assume f to be non-negative and finite. The proof of Theorem 4.3.1 will be executed by sequentially showing each step below. Step 4. would complete the proof.

Step 1. For any $R \in \mathcal{H}$ with $\overline{R} \subset U$ and $v(R) \in \mathbb{Q}^d$,

$$\lambda\big(T(R)\big) \le \int_R |\det(J(x))| \, dx \, . \tag{4.3.3}$$

Step 2. For all $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda\big(T(A \cap U)\big) \le \int_{A \cap U} |\det(J(x))| \, dx \, . \tag{4.3.4}$$

Step 3. For any non-negative measurable function $f : V \to \mathbb{R}$,

$$\int_U f\big(T(x)\big) \, |\det(J(x))| \, dx \ge \int_V f(y) \, dy \, . \tag{4.3.5}$$

Step 4. The inequality in (4.3.5) is an equality.

The proof of Step 1., which is the main step of the proof, is based on the idea that locally T is like a linear transformation.

Proof of Step 1. Fix $R \in \mathcal{H}$ with $\overline{R} \subset U$ and $v(R) \in \mathbb{Q}^d$; $\overline{R}$ is a compact set as R is bounded. Let $\varepsilon > 0$. Since $\det(J(\cdot))$ is a continuous function, it is uniformly continuous on $\overline{R}$. Choose $\delta_1 > 0$ such that

$$\Big|\det(J(x)) - \det(J(x'))\Big| \le \varepsilon \text{ for all } x, x' \in \overline{R}, \, \|x - x'\| \le \delta_1 \, , \tag{4.3.6}$$

where $\|\cdot\|$ denotes the L^∞ norm as in (4.3.2) throughout.

Recall that the function $A \to A^{-1}$, from the space of $d \times d$ non-singular matrices to itself, is continuous. Since $J(x)$ is non-singular for all $x \in U$, the map $x \mapsto J(x)^{-1}$ is continuous on U. Thus,

$$f : \overline{R} \times \{z \in \mathbb{R}^d : \|z\| = 1\} \to \mathbb{R}^d \, ,$$

defined by

$$f(x, z) = J(x)^{-1} z \, , (x, z) \in \overline{R} \times \{z \in \mathbb{R}^d : \|z\| = 1\} \, ,$$

is a continuous function defined on a compact set; elements of $\mathbb{R}^d$ are viewed as $d \times 1$ vectors by convention. Therefore,

$$c = \max \Big\{ \|f(x, z)\| : \, (x, z) \in \overline{R} \times \{z \in \mathbb{R}^d : \|z\| = 1\} \Big\} < \infty \, .$$

In other words,

$$\big\|J(x)^{-1} z\big\| \le c\|z\| \, , x \in \overline{R}, \, z \in \mathbb{R}^d \, . \tag{4.3.7}$$

Denote by $J_{ij}(x)$ the (i,j)-th entry of $J(x)$ for all $x \in U$ and $1 \le i,j \le d$. Uniform continuity of $J_{ij}(\cdot)$ on $\overline{R}$ ensures the existence of $\delta_2 > 0$ such that

$$\left|J_{ij}(x) - J_{ij}(x')\right| \le \frac{\varepsilon}{cd} \text{ for all } x, x' \in \overline{R},\ \|x - x'\| \le \delta_2 \,. \tag{4.3.8}$$

Let $0 < \delta \le \min\{\delta_1, \delta_2\}$ be such that $\delta^{-1}(b_i - a_i)$ is an integer for every i. Choosing such a δ is possible because $b_i - a_i$ is rational; if p_i, q_i are positive integers with $b_i - a_i = p_i/q_i$, letting

$$\delta = \frac{1}{nq_1 \ldots q_d} \,,$$

works for large n, for example.

Consider the square

$$[a_1 + (i_1 - 1)\delta, a_1 + i_1\delta] \times \ldots \times [a_d + (i_d - 1)\delta, a_d + i_d\delta] \,,$$

where $i_1, \ldots, i_d$ are positive integers with $i_j \le \delta^{-1}(b_j - a_j)$ for $j = 1, \ldots, d$. Denote the collection of all such squares by $\{Q_1, \ldots, Q_k\}$. In other words, $Q_1, \ldots, Q_k$ are compact squares of side-length δ such that

$$\overline{R} = Q_1 \cup \ldots \cup Q_k \,, \tag{4.3.9}$$

and $\lambda(Q_i \cap Q_j) = 0$ for $1 \le i < j \le k$ by Exc 4.2.3. Let x_i be the centre of Q_i (the centre of a square or a rectangle is well defined). Recalling that $\|\cdot\|$ is the L^∞ norm, write

$$Q_i = B_{\delta/2}(x_i)\,, i = 1, \ldots, k \,, \tag{4.3.10}$$

where for $r \ge 0$ and $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$,

$$B_r(z) = \{y \in \mathbb{R}^d : \|y - z\| \le r\} = [z_1 - r, z_1 + r] \times \ldots \times [z_d - r, z_d + r] \,. \tag{4.3.11}$$

The above is precisely the advantage of working with the L^∞ norm.

For $i = 1, \ldots, k$, define

$$\phi_i(z) = J(x_i)(z - x_i) + T(x_i)\,, z \in \mathbb{R}^d \,.$$

Our first claim is that

$$T(Q_i) \subset \phi_i(Q_i^\varepsilon)\,, i = 1, \ldots, k \,, \tag{4.3.12}$$

where

$$Q_i^\varepsilon = B_{(1+\varepsilon)\delta/2}(x_i)\,, i = 1, \ldots, k \,.$$

Proceeding towards proving (4.3.12), fix $i \in \{1, \ldots, k\}$, and use Fact 4.3.2 along with (4.3.8) to claim that for all $z \in Q_i$,

$$\|T(z) - T(x_i) - J(x_i)(z - x_i)\| \le \frac{\varepsilon}{c}\|z - x_i\| \,.$$

Since the left hand side above equals $\|T(z)-\phi_i(z)\|$, it follows that

$$\|T(z)-\phi_i(z)\| \le \frac{\varepsilon}{c}\|z-x_i\|\,, z\in Q_i\,. \tag{4.3.13}$$

Therefore, for $z\in Q_i$,

$$\begin{aligned}\left\|\phi_i^{-1}\circ T(z)-z\right\| &= \left\|\phi_i^{-1}\circ T(z)-\phi_i^{-1}\circ\phi_i(z)\right\| \\ &= \left\|J(x_i)^{-1}\left(T(z)-\phi_i(z)\right)\right\| \\ &\le c\,\|T(z)-\phi_i(z)\| \\ &\le \varepsilon\|z-x_i\|\,,\end{aligned}$$

(4.3.7) and (4.3.13) implying the inequalities in the penultimate line and the last line, respectively. Thus, for $z\in Q_i$,

$$\left\|\phi_i^{-1}\circ T(z)-x_i\right\| \le \left\|\phi_i^{-1}\circ T(z)-z\right\| + \|z-x_i\| \le (1+\varepsilon)\|z-x_i\|\,.$$

Recall (4.3.10) to argue that

$$\phi_i^{-1}\circ T(z)\in Q_i^\varepsilon\,, z\in Q_i\,,$$

which is equivalent to (4.3.12).

An immediate implication of (4.3.12) is that for fixed $i=1,\ldots,k$,

$$\begin{aligned}\lambda\left(T(Q_i)\right) &\le \lambda\left(\{J(x_i)z+T(x_i)-J(x_i)x_i : z\in Q_i^\varepsilon\}\right) \\ &= |\det(J(x_i))|\lambda(Q_i^\varepsilon)\,,\end{aligned}$$

the second line following from Theorem 4.2.1. This is the crux of the proof in that it shows how the modulus of the determinant of the Jacobian appears. Further, (4.3.11) shows Q_i^ε is a square of side-length $(1+\varepsilon)\delta$. Therefore,

$$\lambda(Q_i^\varepsilon) = (1+\varepsilon)^d\delta^d = (1+\varepsilon)^d\lambda(Q_i)\,,$$

(4.3.10) implying the second equality. Put everything together to get

$$\lambda\left(T(Q_i)\right) \le |\det(J(x_i))|(1+\varepsilon)^d\lambda(Q_i)\,.$$

Thus,

$$\begin{aligned}\lambda\left(T(R)\right) &\le \sum_{i=1}^k \lambda(T(Q_i)) \\ &\le (1+\varepsilon)^d\sum_{i=1}^k |\det(J(x_i))|\lambda(Q_i) \\ \text{(by (4.3.6) and } \delta\le\delta_1) &\le (1+\varepsilon)^d\sum_{i=1}^k\left(\varepsilon+\min_{z\in Q_i}|\det(J(z))|\right)\lambda(Q_i) \\ (\lambda(Q_i\cap Q_j)=0,\ i\ne j) &\le (1+\varepsilon)^d\left(\varepsilon\lambda(R)+\int_{Q_1\cup\ldots\cup Q_k}|\det(J(x))|\,dx\right) \\ &= (1+\varepsilon)^d\left(\varepsilon\lambda(R)+\int_R|\det(J(x))|dx\right),\end{aligned}$$

the last line following from (4.3.9) and Exc 4.2.2. Since the above holds for all $\varepsilon > 0$, letting $\varepsilon \downarrow 0$ completes the proof of Step 1. $\square$

While Step 1. was mostly based on analysis and linear algebra, the proof of Step 2. is standard in measure theory and follows from Lemma 4.2.1.

Proof of Step 2. Define measures μ and ν on $\mathbb{R}^d$ by

$$\mu(A) = \lambda(T(A \cap U))\,, A \in \mathcal{B}(\mathbb{R}^d)\,,$$

and

$$\nu(B) = \int_{B\cap U} |\det(J(x))|\; dx\,, B \in \mathcal{B}(\mathbb{R}^d)\,.$$

The claim (4.3.4) is equivalent to

$$\mu(A) \le \nu(A)\,, A \in \mathcal{B}(\mathbb{R}^d)\,. \tag{4.3.14}$$

In view of Lemma 4.2.1, it suffices to show that the claim holds for any $R \in \mathcal{H}$ with $\overline{R} \subset U$ and $v(R) \in \mathbb{Q}^d$, that is,

$$\mu(R) \le \nu(R)\,. \tag{4.3.15}$$

This is precisely what has been shown in Step 1. $\square$

The proof of Step 3., which is also standard, is based on approximating a non-negative measurable function by simple functions from below.

Proof of Step 3. First let $f : V \to \mathbb{R}$ be a non-negative simple function, that is,

$$f = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{A_i}\,,$$

for some $\alpha_1, \ldots, \alpha_k \in [0, \infty]$ and $A_1, \ldots, A_k \in \mathcal{B}(\mathbb{R}^d)$ with $A_i \subset V$ for all i.

Then,

$$\begin{aligned}
\int_V f(y)\,dy &= \sum_{i=1}^k \alpha_i \lambda(A_i) \\
&= \sum_{i=1}^k \alpha_i \lambda(T(T^{-1}A_i)) \\
&\le \sum_{i=1}^k \alpha_i \int_{T^{-1}A_i} |\det(J(x))|\,dx \\
&= \int_U |\det(J(x))| \sum_{i=1}^k \alpha_i \mathbf{1}_{T^{-1}A_i}(x)\,dx \\
&= \int_U |\det(J(x))| \sum_{i=1}^k \alpha_i \mathbf{1}_{A_i}(T(x))\,dx \\
&= \int_U |\det(J(x))|\, f(T(x))\,dx\,,
\end{aligned}$$

the inequality in the third line following from Step 2. Thus,

$$\int_V f(y)\,dy \le \int_U |\det(J(x))|\, f(T(x))\,dx\,. \tag{4.3.16}$$

For a measurable function $f : V \to [0, \infty)$, there exist non-negative simple functions f_n such that $f_n \uparrow f$. The desired inequality (4.3.16) holds with f replaced by f_n therein. Letting $n \to \infty$ with the help of MCT, the proof of Step 3. follows. □

Step 4. is a consequence of the inverse function theorem.

Proof of Step 4. Fact 4.3.1 and the assumption that $J(x)$ is non-singular for all $x \in U$ imply that $T^{-1} : V \to U$ is a continuously differentiable bijection whose Jacobian matrix is $J(T^{-1}y)^{-1}$ for all $y \in V$. Using Step 3. with U, V, T replaced by V, U, T^{-1} implies

$$\int_U g(x)\,dx \le \int_V g \circ T^{-1}(y) \left|\det\left(J(T^{-1}y)^{-1}\right)\right|\,dy\,, \tag{4.3.17}$$

for any measurable $g : U \to [0, \infty)$.

Fix a measurable $f : V \to [0, \infty)$. Define

$$g(x) = f \circ T(x) |\det(J(x))|\,, x \in U\,.$$

Apply (4.3.17) to this g to get

$$
\begin{aligned}
\int_U f \circ T(x) |\det(J(x))|\, dx &\le \int_V g \circ T^{-1}(y) \left|\det\left(J(T^{-1}y)^{-1}\right)\right|\, dy \\
&= \int_V f(y) \left|\det\left(J(T^{-1}y)\right)\right| \left|\det\left(J(T^{-1}y)^{-1}\right)\right|\, dy \\
&= \int_V f(y)\, dy\,.
\end{aligned}
$$

Compare this with (4.3.5) obtained in Step 3. to get

$$\int_U f \circ T(x) |\det(J(x))|\, dx = \int_V f(y)\, dy\,.$$

This completes the proof of Step 4. and that of Theorem 4.3.1 as well. □

The following special case of Theorem 4.3.1 deserves special mention.

Corollary 4.3.2 (Transformation to polar coordinates)**.** *For a Borel measurable $f : \mathbb{R}^2 \to \overline{\mathbb{R}}$,*

$$\int_{\mathbb{R}^2} f(x,y)\, \lambda(dx, dy) = \int_{(0,2\pi)\times(0,\infty)} f\left(r\cos\theta, r\sin\theta\right) r\, \lambda(dr,\, d\theta), \quad (4.3.18)$$

whenever the integral on either side makes sense.

Proof. Use Theorem 4.3.1 with $U = (0,2\pi)\times(0,\infty)$, $V = \mathbb{R}^2\backslash\{(x,0) : x \ge 0\}$ and $T : U \to V$ defined by

$$T(\theta, r) = \left(r\cos\theta, r\sin\theta\right), (\theta, r) \in U\,.$$

The Jacobian matrix of T is

$$J(\theta, r) = \begin{bmatrix} -r\sin\theta & r\cos\theta \\ \cos\theta & \sin\theta \end{bmatrix}, (\theta, r) \in U\,,$$

and hence

$$|\det J(\theta, r)| = r\,.$$

Since T is a continuously differentiable bijection from U to V whose Jacobian matrix is always non-singular, (4.3.18) follows from Theorem 4.3.1 whenever either side of it makes sense. □

An important probabilistic consequence of Theorem 4.3.1 is the following.

Theorem 4.3.2. *Suppose $X = (X_1, \ldots, X_d)$ is a random vector with $P(X \in U) = 1$ for some open set $U \subset \mathbb{R}^d$. Let $\psi : U \to V$ be a bijection for some open set $V \subset \mathbb{R}^d$. Let $T : V \to U$ be the inverse of ψ. Assume T is continuously differentiable and its Jacobian matrix $J(y)$ at $y \in V$, defined by*

$$J(y) = \frac{\partial T(y)}{\partial y},$$

is non-singular for all $y \in V$. Then the joint density of $Y = (Y_1, \ldots, Y_d) = \psi(X)$ is

$$g(y) = \begin{cases} f \circ T(y) |\det(J(y))|, & y \in V, \\ 0, & y \notin V. \end{cases}$$

Proof. Since $(Y_1, \ldots, Y_d) \in V$ a.s., for $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} P\big((Y_1, \ldots, Y_d) \in B\big) &= P\big((Y_1, \ldots, Y_d) \in B \cap V\big) \\ &= P\big((X_1, \ldots, X_d) \in T(B \cap V)\big) \\ &= \int_{T(B \cap V)} f(x)\, dx \\ &= \int_{B \cap V} f \circ T(y) |\det(J(y))|\, dy \\ &= \int_B g(y)\, dy, \end{aligned}$$

the penultimate line following from Theorem 4.3.1. Hence the proof follows. □

5 Product measures

5.1 Product of two measure spaces

Suppose $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ are finite measure spaces. Define

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}, \tag{5.1.1}$$

that is, $\Omega_1 \times \Omega_2$ is the usual Cartesian product of Ω_1 and Ω_2, and let

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma\left(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}\right). \tag{5.1.2}$$

For all $E \subset \Omega_1 \times \Omega_2$, define

$$\begin{aligned} E_{\omega_1} &= \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in E\}, \, \omega_1 \in \Omega_1, \\ E^{\omega_2} &= \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in E\}, \, \omega_2 \in \Omega_2. \end{aligned}$$

The following result is the first step towards constructing product measures.

Theorem 5.1.1. *For all fixed $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$, the following holds:*

1. *$E_{\omega_1} \in \mathcal{A}_2$ for all $\omega_1 \in \Omega_1$,*
2. *$\omega_1 \mapsto \mu_2(E_{\omega_1})$ is an $\mathcal{A}_1$-measurable function on Ω_1 and*
3. *μ defined on $\mathcal{A}_1 \otimes \mathcal{A}_2$ by*

$$\mu(E) = \int_{\Omega_1} \mu_2(E_{\omega_1})\, \mu_1(d\omega_1),\ E \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

   is a measure on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.

Similarly, the following holds for all $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$:

4. *$E^{\omega_2} \in \mathcal{A}_1$ for all $\omega_2 \in \Omega_2$,*
5. *$\omega_2 \mapsto \mu_1(E^{\omega_2})$ is an $\mathcal{A}_2$-measurable function on Ω_2 and*
6. *μ' defined on $\mathcal{A}_1 \otimes \mathcal{A}_2$ by*

$$\mu'(E) = \int_{\Omega_2} \mu_1(E^{\omega_2})\, \mu_2(d\omega_2),\ E \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

   is a measure on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.

Finally,

$$\mu(E) = \mu'(E) \text{ for all } E \in \mathcal{A}_1 \otimes \mathcal{A}_2\,. \tag{5.1.3}$$

Proof. The proof follows by routine verifications, as sketched below. Let

$$\mathcal{G} = \{A \subset \Omega_1 \times \Omega_2 :\ \text{each of 1,2, 4 and 5 above holds for } A\}\,.$$

Define

$$\mathcal{S} = \{A_1 \times A_2 :\ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\},$$

which is a semi-field because it is trivially closed under finite intersections and for $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,

$$(A_1 \times A_2)^c = (A_1 \times A_2^c) \cup (A_1^c \times \Omega_2)\ .$$

It is immediate that $\mathcal{S} \subset \mathcal{G}$ because for $A_1 \times A_2 \in \mathcal{S}$,

$$(A_1 \times A_2)_{\omega_1} = \begin{cases} A_2, & \omega_1 \in A_1, \\ \emptyset, & \text{else}, \end{cases}$$

and hence

$$\mu_2\left((A_1 \times A_2)_{\omega_1}\right) = \mu_2(A_2)\mathbf{1}_{A_1}(\omega_1) \tag{5.1.4}$$

is an $\mathcal{A}_1$-measurable function of ω_1, showing 1 and 2 hold; 4 and 5 hold for $A_1 \times A_2$ by similar arguments. Routine arguments show that the field generated by $\mathcal{S}$ is contained in $\mathcal{G}$. Finally $\mathcal{G}$ can be shown to be a monotone class with the help of DCT, using the fact that μ_1, μ_2 are finite measures, by standard arguments. The monotone class theorem shows $\mathcal{G} \supset \sigma(\mathcal{S}) = \mathcal{A}_1 \otimes \mathcal{A}_2$, that is, 1, 2, 4 and 5 hold for all $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$.

Once the RHS of 3 is defined, due to 2, that μ is a measure is immediate. Similarly, 6 also follows. For (5.1.3), observe that (5.1.4) implies for all $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$,

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) .$$

Similarly,

$$\mu'(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),\ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 .$$

Thus μ and μ' are finite measures which agree on the semi-field $\mathcal{S}$. Hence (5.1.3) follows. ☐

The jump from finite to σ-finite is easy and is left as an exercise.

Theorem 5.1.2. *The claims of Theorem 5.1.1 hold when μ_1 and μ_2 are σ-finite measures.*

Proof. Exercise. ☐

Now we are in a position to define the product of two σ-finite measure spaces.

Definition. *Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Let $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ be as in* (5.1.1) *and* (5.1.2)*, respectively. The unique σ-finite measure μ on $(\Omega, \mathcal{A})$ satisfying*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \text{ for all } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$$

is the product measure *of μ_1 and μ_2. The measure space $(\Omega, \mathcal{A}, \mu)$ is the* product of the measure spaces $(\Omega_1, \mathcal{A}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{A}_2, \mu_2)$*, which is written as*

$$(\Omega_1, \mathcal{A}_1, \mu_1) \otimes (\Omega_2, \mathcal{A}_2, \mu_2) = (\Omega, \mathcal{A}, \mu) .$$

Usually we also write $\mu = \mu_1 \otimes \mu_2$.

5.2 Tonelli and Fubini

Throughout this subsection, $(\Omega_i, \mathcal{A}_i, \mu_i)$ is a σ-finite measure space for $i = 1, 2$.

Theorem 5.2.1 (Tonelli)**.** *Suppose $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ is $\mathcal{A}_1 \otimes \mathcal{A}_2$-measurable. Then*

1. *for all fixed $\omega_1 \in \Omega_1$, $f(\omega_1, \cdot)$ is an $\mathcal{A}_2$-measurable function from Ω_2 to $[0, \infty]$,*

2. $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\mu_2(d\omega_2)$ *is an* $\mathcal{A}_1$*-measurable function of* ω_1 *on* Ω_1*,*
3. *for all fixed* $\omega_2 \in \Omega_2$*,* $f(\cdot, \omega_2)$ *is an* $\mathcal{A}_1$*-measurable function from* Ω_1 *to* $[0, \infty]$*,*
4. $\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\mu_1(d\omega_1)$ *is an* $\mathcal{A}_2$*-measurable function of* ω_2 *on* Ω_2*,*
5. *and*

$$\begin{aligned}\int_{\Omega_1}\int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2)\, \mu_1(d\omega_1) &= \int_{\Omega_1 \times \Omega_2} f \, d\,(\mu_1 \otimes \mu_2) \\ &= \int_{\Omega_2}\int_{\Omega_1} f(\omega_1, \omega_2)\, \mu_1(d\omega_1)\, \mu_2(d\omega_2)\end{aligned}$$

where

$$\int_{\Omega_1}\int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2)\, \mu_1(d\omega_1)$$

means

$$\int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2)\right)\, \mu_1(d\omega_1)$$

by convention.

Proof. For $f = \mathbf{1}_E$ where $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$, 1, 2, 4, 5 and (5.1.3) of Theorem 5.1.1, which hold for σ-finite measure spaces by Theorem 5.1.2, prove 1-5 above. Routine arguments via simple functions and MCT complete the proof. □

Theorem 5.2.2 (Fubini)**.** *Suppose* $(\Omega, \mathcal{A}, \mu) = (\Omega_1, \mathcal{A}_1, \mu_1) \otimes (\Omega_2, \mathcal{A}_2, \mu_2)$ *and*

$$f \in L^1\left(\Omega, \mathcal{A}, \mu\right) .$$

Then the following holds.

1. *For all* $\omega_1 \in \Omega_1$*,* $f(\omega_1, \cdot)$ *is an* $\mathcal{A}_2$*-measurable function form* Ω_2 *to* $\overline{\mathbb{R}}$*.*
2. *For* (μ_1) *almost all* $\omega_1 \in \Omega_1$*,*

$$\int_{\Omega_2} |f(\omega_1, \omega_2)|\, \mu_2(d\omega_2) < \infty .$$

3. *If* $g : \Omega_1 \to \mathbb{R}$ *is defined by*

$$g(\omega_1) = \begin{cases} \int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2), & \text{if } \int_{\Omega_2} |f(\omega_1, \omega_2)|\, \mu_2(d\omega_2) < \infty, \\ 0, & \text{else,} \end{cases}$$

then g *is measurable,* μ_1*-integrable and*

$$\int_{\Omega_1} g \, d\mu_1 = \int_{\Omega} f \, d\mu .$$

4. For all $\omega_2 \in \Omega_2$, $f(\cdot, \omega_2)$ is an $\mathcal{A}_1$-measurable function form Ω_1 to $\overline{\mathbb{R}}$.

5. For (μ_2) almost all $\omega_2 \in \Omega_2$,

$$\int_{\Omega_1} |f(\omega_1, \omega_2)| \, mu_1(d\omega_1) < \infty \, .$$

6. If $h : \Omega_2 \to \mathbb{R}$ is defined by

$$h(\omega_2) = \begin{cases} \int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1), & \text{if } \int_{\Omega_1} |f(\omega_1, \omega_2)| \, \mu_1(d\omega_1) < \infty, \\ 0, & \text{else,} \end{cases}$$

then h is measurable, μ_2-integrable and

$$\int_{\Omega_2} h \, d\mu_2 = \int_{\Omega} f \, d\mu \, .$$

Proof. Follows by applying Theorem 5.2.1 to f^+ and f^-. □

The following corollary of Theorems 5.2.1 and 5.2.2 is often referred to jointly as the Tonelli-Fubini theorem.

Corollary 5.2.1. *For a measurable $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$, it holds that*

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, \mu_2(d\omega_2) \, \mu_1(d\omega_1) = \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1) \, \mu_2(d\omega_2),$$

whenever either $f \geq 0$ or

$$\int_{\Omega_1 \times \Omega_2} |f| \, d(\mu_1 \otimes \mu_2) < \infty \, .$$

5.3 Infinite Product of probability spaces

Applying mathematical induction to Theorem 5.1.1 and Corollary 5.2.1 yields the following.

Theorem 5.3.1. *If $(\Omega_i, \mathcal{A}_i, \mu_i)$ is a σ-finite measure space for $i = 1, \dots, n$, then there exists a unique σ-finite measure μ on $(\Omega, \mathcal{A})$, where*

$$\Omega = \prod_{i=1}^{n} \Omega_i \text{ and } \mathcal{A} = \sigma\left(\{A_1 \times \ldots \times A_n : \, A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n\}\right), \quad (5.3.1)$$

satisfying

$$\mu(A_1 \times \ldots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i), \; A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n \, . \quad (5.3.2)$$

Then for any permutation π of $\{1,\ldots,n\}$ and a $\mathcal{A}$-measurable $f:\Omega\to\overline{\mathbb{R}}$,

$$\int_{\Omega} f\,d\mu = \int_{\Omega_{\pi_1}} \ldots \int_{\Omega_{\pi_n}} f\left(\omega_1,\ldots,\omega_n\right)\,\mu_{\pi_n}(d\omega_{\pi_n})\,\ldots\,\mu_{\pi_1}(d\omega_{\pi_1}), \quad (5.3.3)$$

whenever either $f\ge 0$ or $f\in L^1\left(\Omega,\mathcal{A},\mu\right)$.

Definition. *If $(\Omega_i,\mathcal{A}_i,\mu_i)$ is a σ-finite measure space for $i=1,\ldots,n$, then their* product *is $(\Omega,\mathcal{A},\mu)$ where Ω and $\mathcal{A}$ are as in* (5.3.1) *and μ is as in* (5.3.2)*. We write*

$$\bigotimes_{i=1}^{n}(\Omega_i,\mathcal{A}_i,\mu_i) = (\Omega,\mathcal{A},\mu)\,.$$

Exercise 5.3.1. *1. Show that for $n=2,3,\ldots$,*

$$\bigotimes_{i=1}^{n}\left(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda\right) = \left(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\lambda_n\right)$$

where λ and λ_n are the Lebesgue measures on $\mathbb{R}$ and $\mathbb{R}^n$, respectively.

2. Prove that

$$\int_{\mathbb{R}} e^{-x^2/2}\,dx = \sqrt{2\pi}\,.$$

Hint. *Use 1 above with $n=2$ and Corollary 4.3.2.*

We now proceed to show that a countable product or probability spaces exist. A moment's thought makes it clear that such an infinite product would not have been possible if the measure of the whole space were not 1. This justifies why the measure of whole space is taken to be 1 in probability theory. It is instructive to notice that the hypothesis that the underlying measures are probability measures has been used several times in the proof of the following theorem.

Theorem 5.3.2. *If $(\Omega_n,\mathcal{A}_n,P_n)$ is a probability space for $n=1,2,\ldots$, then there exists a unique probability measure P on $(\prod_{n=1}^{\infty}\Omega_n,\bigotimes_{n=1}^{\infty}\mathcal{A}_n)$ such that for all $n=1,2,\ldots$,*

$$P\left(\left(\prod_{i=1}^{n}A_i\right)\times\prod_{i=n+1}^{\infty}\Omega_i\right) = \prod_{i=1}^{n}P_i(A_i) \text{ for all } A_1\in\mathcal{A}_1,\ldots,A_n\in\mathcal{A}_n, \quad (5.3.4)$$

where

$$\bigotimes_{n=1}^{\infty}\mathcal{A}_n = \sigma\left(\left\{\left(\prod_{i=1}^{n}A_i\right)\times\prod_{i=n+1}^{\infty}\Omega_i : A_i\in\mathcal{A}_i, i=1,\ldots,n,\, n\ge 1\right\}\right)\,. \quad (5.3.5)$$

Proof. For $n = 1, 2, \ldots$, define

$$\mathcal{F}_n = \left\{ A \times \Omega_{n+1} \times \Omega_{n+2} \ldots : A \in \bigotimes_{i=1}^{n} \mathcal{A}_i \right\}. \tag{5.3.6}$$

It is easy to see that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ and each $\mathcal{F}_n$ is a σ-field. Hence,

$$\mathcal{F}_\infty = \bigcup_{n=1}^{\infty} \mathcal{F}_n$$

is a field. Further,

$$\sigma(\mathcal{F}_\infty) = \bigotimes_{n=1}^{\infty} \mathcal{A}_n .$$

Let $P^{(n)} = \bigotimes_{i=1}^n P_i$, that is, it is the unique measure on $\bigotimes_{i=1}^n \mathcal{A}_i$ satisfying

$$P^{(n)}(A_1 \times \ldots \times A_n) = \prod_{i=1}^{n} P_i(A_i),\ A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n, \tag{5.3.7}$$

which exists by Theorem 5.3.1. It is easy to see that for $1 \le n_1 < n_2$,

$$P^{(n_2)}\left(A \times \prod_{i=n_1+1}^{n_2} \Omega_i \right) = P^{(n_1)}(A). \tag{5.3.8}$$

Therefore, the following definition of P on $\mathcal{F}_\infty$ is unambiguous:

$$P\left(A \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots\right) = P^{(n)}(A) \text{ for all } A \in \mathcal{F}_n,\ n \ge 1. \tag{5.3.9}$$

It is immediate that P is finitely additive on $\mathcal{F}_\infty$ because so is each $P^{(n)}$. In view of the Carathéodory extension theorem, it needs to be shown that P is countably additive. As P is finitely additive, it suffices to prove that if $A_n \in \mathcal{F}_\infty$ are such that

$$A_n \downarrow \emptyset, \tag{5.3.10}$$

then

$$P(A_n) \downarrow 0 .$$

Fix $A_n \in \mathcal{F}_\infty$ satisfying (5.3.10). Finite additivity implies that $P(A_n) \downarrow$. Thus the above can fail only if there exists $\varepsilon > 0$ such that

$$P(A_n) \ge \varepsilon \text{ for all } n . \tag{5.3.11}$$

Assume (5.3.11) for the sake of contradiction.

In a way very similar to (5.3.6)-(5.3.9), define for all $1 \le k \le n$,

$$\mathcal{F}_{k:n} = \left\{ A \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots :\ A \in \bigotimes_{i=k}^{n} \mathcal{A}_i \right\},$$

which is a a σ-field on $\Omega_k \times \Omega_{k+1} \times \ldots$,

$$\mathcal{F}_{k:\infty} = \bigcup_{n=k}^{\infty} \mathcal{F}_{k:n}, \text{ which is a field on } \Omega_k \times \Omega_{k+1} \times \ldots,$$

and a finitely additive function $P^{(k:\infty)} : \mathcal{F}_{k:\infty} \to [0,1]$ satisfying

$$P^{(k:\infty)}\left(A \times \prod_{i=n+1}^{\infty} \Omega_i\right) = \left(\bigotimes_{i=k}^{n} P_i\right)(A) \text{ for all } A \in \bigotimes_{i=k}^{n} \mathcal{A}_i,\ n \geq k\,. \tag{5.3.12}$$

Define for all $E \in \mathcal{F}_{k:\infty}$, $n = k, k+1, k+2, \ldots$ and $\omega_k \in \Omega_k, \ldots, \omega_n \in \Omega_n$,

$$E^{(\omega_k,\ldots,\omega_n)} = \left\{(\omega_{n+1}, \omega_{n+2}, \ldots) \in \prod_{i=n+1}^{\infty} \Omega_i \,:\, (\omega_k, \omega_{k+1}, \ldots) \in E\right\}. \tag{5.3.13}$$

Fix $n \geq 1$. Recall that $A_n \in \mathcal{F}_\infty$ implies $A_n \in \mathcal{F}_{k_n}$ for some k_n, that is,

$$A_n = \tilde{A}_n \times \Omega_{k_n+1} \times \Omega_{k_n+2} \times \ldots, \text{ for some } \tilde{A}_n \in \bigotimes_{i=1}^{k_n} \mathcal{A}_i\,. \tag{5.3.14}$$

Assuming without loss of generality that $k_n \geq 2$, (5.3.12) implies

$$P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) = \left(\bigotimes_{i=2}^{k_n} P_i\right)\left(\{(\omega_2, \ldots, \omega_{k_n}) \,:\, (\omega_1, \ldots, \omega_{k_n}) \in \tilde{A}_n\}\right). \tag{5.3.15}$$

It follows from (5.3.9) that

$$\begin{aligned}
P(A_n) &= P^{(k_n)}(\tilde{A}_n) \\
&= \int_{\Omega_1 \times \ldots \times \Omega_{k_n}} \mathbf{1}_{\tilde{A}_n}\, d\left(\bigotimes_{i=1}^{k_n} P_i\right) \\
\text{(by (5.3.3))} &= \int_{\Omega_1}\left(\int_{\Omega_2 \times \ldots \times \Omega_{k_n}} \mathbf{1}_{\tilde{A}_n}(\omega_1, \ldots, \omega_{k_n}) \bigotimes_{i=2}^{k_n} P_i(d\omega_i)\right) P_1(d\omega_1) \\
&= \int_{\Omega_1}\left(\bigotimes_{i=2}^{k_n} P_i\right)\left(\{(\omega_2, \ldots, \omega_{k_n}) \,:\, (\omega_1, \ldots, \omega_{k_n}) \in \tilde{A}_n\}\right) P_1(d\omega_1) \\
&= \int_{\Omega_1} P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) P_1(d\omega_1),
\end{aligned} \tag{5.3.16}$$

(5.3.15) implying the last line. Setting

$$F_n^{(1)} = \left\{\omega_1 \in \Omega_1 : P^{(2:\infty)}(A_n^{(\omega_1)}) > \frac{\varepsilon}{2}\right\}, \tag{5.3.17}$$

it thus follows from (5.3.11) and (5.3.16) that

$$\begin{aligned}\varepsilon \le P(A_n) &= \int_{\Omega_1} P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) P_1(d\omega_1)\\ &= \int_{F_n^{(1)}} P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) P_1(d\omega_1)\\ &\quad + \int_{\Omega_1 \setminus F_n^{(1)}} P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) P_1(d\omega_1)\\ &\le \int_{F_n^{(1)}} 1\, P_1(d\omega_1) + \int_{\Omega_1\setminus F_n^{(1)}} \frac{\varepsilon}{2}\, P_1(d\omega_1)\\ &\le P_1\left(F_n^{(1)}\right) + \frac{\varepsilon}{2},\end{aligned}$$

the inequality in the penultimate line following from the fact

$$P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) \le 1 \text{ for all } \omega_1 \in \Omega_1$$

and the definition of $F_n^{(1)}$, while the last line uses the fact P_1 is a probability measure.

Thus

$$P_1\left(F_n^{(1)}\right) \ge \frac{\varepsilon}{2},\ n \ge 1\,.$$

Since $A_n \supset A_{n+1}$ and $P^{(2:\infty)}$ is finitely additive, (5.3.17) implies $F_n^{(1)} \supset F_{n+1}^{(1)}$. Denoting

$$F^{(1)} = \bigcap_{n=1}^{\infty} F_n^{(1)},$$

it is easy to see that $F_n^{(1)} \downarrow F^{(1)}$ and hence

$$P_1\left(F^{(1)}\right) = \lim_{n\to\infty} P_1\left(F_n^{(1)}\right) \ge \frac{\varepsilon}{2},$$

because P_1 is a probability measure. Thus $F^{(1)} \ne \emptyset$. Fix $\omega_1 \in F^{(1)}$. In other words, (5.3.17) implies that $\omega_1 \in \Omega_1$ satisfies

$$P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) > \frac{\varepsilon}{2},\ n \ge 1\,.$$

An argument similar to (5.3.16), using the observation

$$\left(A_n^{(\omega_1)}\right)^{(\omega_2)} = A_n^{(\omega_1,\omega_2)},\ \omega_2 \in \Omega_2,$$

the above notations being as in (5.3.13), shows

$$P^{(2:\infty)}\left(A_n^{(\omega_1)}\right) = \int_{\Omega_2} P^{(3:\infty)}\left(A_n^{(\omega_1,\omega_2)}\right) P_2(d\omega_2)\,.$$

It can thus be shown that there exists $\omega_2 \in \Omega_2$ such that

$$P^{(3:\infty)}\left(A_n^{(\omega_1,\omega_2)}\right) > \frac{\varepsilon}{4},\, n \ge 1\,.$$

Proceeding inductively, $\omega_k \in \Omega_k$ can be chosen such that

$$P^{(k+1:\infty)}\left(A_n^{(\omega_1,\ldots,\omega_k)}\right) > \varepsilon 2^{-k},\ \text{for all } n, k \ge 1\,.$$

Define $\omega = (\omega_1, \omega_2, \ldots)$. Recalling (5.3.14), it is immediate from the above that

$$\varepsilon 2^{-k_n} < P^{(k_n+1:\infty)}\left(A_n^{(\omega_1,\ldots,\omega_{k_n})}\right),$$

a consequence of which is

$$(\omega_1, \ldots, \omega_{k_n}) \in \tilde{A}_n,$$

and hence $\omega \in A_n$. Since this is true for all n,

$$\omega \in \bigcap_{n=1}^{\infty} A_n,$$

which contradicts (5.3.10). Thus (5.3.11) is not possible, that is, $P(A_n) \downarrow 0$.

In other words, P is countably additive on $\mathcal{F}_\infty$. Theorem 2.2.1, which is the Carathéodory extension theorem, shows that P can be extended to a measure on the space $(\prod_{n=1}^{\infty} \Omega_n, \bigotimes_{n=1}^{\infty} \mathcal{A}_n)$. Since

$$P\left(\prod_{n=1}^{\infty} \Omega_n\right) = 1$$

by definition and $\mathcal{F}_\infty$ is a field, the extension is unique and yields a probability measure. This completes the proof. $\square$

Definition. *If* $(\Omega_n, \mathcal{A}_n, P_n)$ *is a probability space for* $n = 1, 2, \ldots$*, then the unique probability measure* P *on* $(\prod_{n=1}^{\infty} \Omega_n \bigotimes_{n=1}^{\infty} \mathcal{A}_n)$ *satisfying* (5.3.4)*, where* $\bigotimes_{n=1}^{\infty} \mathcal{A}_n$ *is as in* (5.3.5)*, is called the* product *of* $P_1, P_2, \ldots$ *and we write*

$$P = \bigotimes_{n=1}^{\infty} P_n\,.$$

6 Independence and its consequences

As usual, $(\Omega, \mathcal{A}, P)$ is the probability space underlying everything talked about and all random variables are $\mathbb{R}$-valued, unless specified otherwise.

6.1 Independence

The concept of independence is defined for σ-fields as follows.

Definition. *Suppose $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are sub-σ-fields of $\mathcal{A}$, that is, $\mathcal{A}_i \subset \mathcal{A}$ and $\mathcal{A}_i$ is a σ-field for $i = 1, \ldots, n$. We say $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are* independent *if*

$$P\left(A_1 \cap \ldots \cap A_n\right) = P(A_1) \ldots P(A_n) \text{ for all } A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n\,.$$

A possibly infinite collection $\{\mathcal{A}_\alpha\}_{\alpha \in I}$ of sub-σ-fields of $\mathcal{A}$ is independent *if $A_{\alpha_1}, \ldots, A_{\alpha_n}$ are independent for all distinct $\alpha_1, \ldots, \alpha_n \in I$.*

The following exercise connects that the usually given definition of independence of events with the above one.

Exercise 6.1.1. *For $A_1, \ldots, A_n \in \mathcal{A}$, show that $\sigma(\{A_1\}), \ldots, \sigma(\{A_n\})$ are independent if and only if*

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}) \text{ for all } 1 \leq i_1 < \ldots < i_k \leq n,\ k = 2, \ldots, n\,.$$

We next define independence of random variables.

Definition. *Suppose for each $\alpha \in I$, χ_α is a collection of random variables defined on $(\Omega, \mathcal{A}, P)$. We say $\{\chi_\alpha\}_{\alpha \in I}$ is an independent collection if the collection of σ-fields $\{\sigma(\chi_\alpha)\}_{\alpha \in I}$ are independent, where $\sigma(\chi_\alpha)$ is the smallest σ-field with respect to which each random variable in χ_α is measurable.*

Recalling that for a random variable X,

$$\sigma(X) = \{X^{-1}B :\ B \in \mathcal{B}(\overline{\mathbb{R}})\},$$

an interpretation of the above definition is that $X_1, \ldots, X_n$ are said to be independent if

$$P\left(X_1 \in B_1, \ldots, X_n \in B_n\right) = \prod_{i=1}^{n} P(X_i \in B_i) \text{ for all } B_1, \ldots, B_n \in \mathcal{B}(\overline{\mathbb{R}})\,. \tag{6.1.1}$$

Exercise 6.1.2. *If $\{\mathcal{A}_\alpha\}_{\alpha \in I}$ is a collection of independent σ-fields and X_α is an $\mathcal{A}_\alpha$-measurable random variable for all $\alpha \in I$, show that $\{X_\alpha\}_{\alpha \in I}$ is an independent collection of random variables.*

The following theorem connects independence with product measures, and in particular, Theorem 5.3.2. It formalizes the concept of coupling several (countably many) random experiments on one probability space on which these become independent.

Theorem 6.1.1. *Suppose for* $n = 1, 2, \ldots$, $(\Omega_n, \mathcal{A}_n, P_n)$ *is a probability space on which a random variable* X_n *is defined. Let*

$$(\Omega, \mathcal{A}, P) = \bigotimes_{n=1}^{\infty} (\Omega_n, \mathcal{A}_n, P_n),$$

and

$$\tilde{\mathcal{A}}_n = \left\{ \left(\prod_{i=1}^{n-1} \Omega_i \right) \times A \times \prod_{i=n+1}^{\infty} \Omega_i : A \in \mathcal{A}_n \right\}, \, n \geq 1 \, .$$

Then $\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots$ *are independent* σ*-fields in the probability space* $(\Omega, \mathcal{A}, P)$. *Define for all* $n \geq 1$, $\tilde{X}_n : \Omega \to \overline{\mathbb{R}}$ *by*

$$\tilde{X}_n(\omega) = X_n(\omega_n) \text{ for all } \omega = (\omega_1, \omega_2, \ldots) \in \Omega \, .$$

Then $\tilde{X}_n$ *has the same distribution as* X_n *for all* n, *that is,*

$$P_n \circ X_n^{-1}(B) = P \circ \tilde{X}_n^{-1}(B) \text{ for all } B \in \mathcal{B}(\overline{\mathbb{R}}),$$

and $\tilde{X}_1, \tilde{X}_2, \ldots$ *are independent.*

Proof. The independence of $\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots$ would follow once it is shown that for all $n = 2, 3, \ldots$, $\tilde{\mathcal{A}}_1, \ldots, \tilde{\mathcal{A}}_n$ are independent. For that, fix n and $\tilde{A}_i \in \tilde{\mathcal{A}}_i$, that is,

$$\tilde{A}_i = \left(\prod_{j=1}^{i-1} \Omega_j \right) \times A_i \times \prod_{j=i+1}^{\infty} \Omega_j \text{ for some } A_i \in \mathcal{A}_i \, .$$

Then

$$P\left(\bigcap_{i=1}^{n} \tilde{A}_i \right) = P\left(\left(\prod_{i=1}^{n} A_i \right) \times \prod_{i=n+1}^{\infty} \Omega_i \right) = \prod_{i=1}^{n} P_i(A_i) = \prod_{i=1}^{n} P(\tilde{A}_i),$$

where the second and third equalities follow from the definition of product measures. Thus $\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots$ are independent. Since $\tilde{X}_n$ is $\tilde{\mathcal{A}}_n$-measurable for all $n = 1, 2, \ldots$, Exc 6.1.2 shows $\tilde{X}_1, \tilde{X}_2, \ldots$ are independent. Finally,

$$P_n \circ X_n^{-1} = P \circ \tilde{X}_n^{-1} \text{ for all } n = 1, 2, \ldots$$

also follows trivially from the definition of product measure. □

The following is an immediate consequence of the above theorem and Theorem 2.6.4.

Corollary 6.1.1. *Given functions* $F_1, F_2, \ldots$ *from* $\mathbb{R}$ *to* $[0, 1]$, *which are non-decreasing right continuous and have limits* 0 *and* 1 *at* $-\infty$ *and* ∞, *respectively, there exist independent random variables* $X_1, X_2, \ldots$ *defined on some probability space whose CDFs are* $F_1, F_2, \ldots$, *respectively.*

Definition. *Random variables* $X_1, X_2, \ldots$ *defined on the same probability space* $(\Omega, \mathcal{A}, P)$ *are* independent and identically distributed or i.i.d. *if they are independent and* $X_i \stackrel{d}{=} X_j$ *for all* i, j*, that is,*

$$P \circ X_i^{-1} = P \circ X_j^{-1} .$$

The following result is often very useful in proving independence.

Theorem 6.1.2. *Suppose* $\mathcal{S}_i \subset \mathcal{A}$ *and* $\mathcal{S}_i$ *is a semi-field for* $i = 1, \ldots, n$*. If*

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \ldots P(A_n) \text{ for all } A_1 \in \mathcal{S}_1, \ldots, A_n \in \mathcal{S}_n,$$

then $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_n)$ *are independent.*

Proof. Fix $A_2 \in \mathcal{S}_2, \ldots, A_n \in \mathcal{S}_n$ and define for all $A \in \mathcal{A}$

$$\mu(A) = P(A \cap A_2 \cap \ldots \cap A_n),$$
$$\nu(A) = P(A) \prod_{i=2}^{n} P(A_i) .$$

Clearly μ and ν are finite measures on $(\Omega, \mathcal{A})$ and the hypothesis implies they agree on the σ-field $\mathcal{S}_1$. Hence

$$\mu(A) = \nu(A) \text{ for all } A \in \sigma(\mathcal{S}_1) .$$

In other words,

$$P(A_1 \cap \ldots \cap A_n) = \prod_{i=1}^{n} P(A_i) \text{ for all } A_1 \in \sigma(\mathcal{S}_1), A_2 \in \mathcal{S}_2, \ldots, A_n \in \mathcal{S}_n .$$

Proceeding inductively, it can be shown that for $i = 1, \ldots, n$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i),$$

for all $A_1 \in \sigma(\mathcal{S}_1), \ldots, A_i \in \sigma(\mathcal{S}_i), A_{i+1} \in \mathcal{S}_{i+1}, \ldots, A_n \in \mathcal{S}_n$. The proof follows from the above when $i = n$. □

Definition. *For a collection* $\{\mathcal{A}_\alpha\}_{\alpha \in I}$ *of* σ*-fields, define*

$$\bigvee_{\alpha \in I} \mathcal{A}_\alpha = \sigma\left(\bigcup_{\alpha \in I} \mathcal{A}_\alpha\right) .$$

For independent random variables $X_1, \ldots, X_n$, Theorem 6.1.3 below simply means that $(X_1, \ldots, X_m)$ is independent of $(X_{m+1}, \ldots, X_n)$.

Theorem 6.1.3. *If $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent σ-fields, then for all $1 \le m \le n-1$, $\bigvee_{i=1}^m \mathcal{A}_i$ and $\bigvee_{i=m+1}^n \mathcal{A}_i$ are independent.*

Proof. Define

$$\begin{aligned} \mathcal{S}_1 &= \{A_1 \cap \ldots \cap A_m : A_1 \in \mathcal{A}_1, \ldots, A_m \in \mathcal{A}_m\}, \\ \mathcal{S}_2 &= \{A_{m+1} \cap \ldots \cap A_n : A_{m+1} \in \mathcal{A}_{m+1}, \ldots, A_n \in \mathcal{A}_n\}. \end{aligned}$$

Then $\mathcal{S}_1$ is a semi-field because it is closed under finite intersections and for $A_1 \in \mathcal{A}_1, \ldots, A_m \in \mathcal{A}_m$,

$$(A_1 \cap \ldots \cap A_m)^c = A_1^c \cup (A_1 \cap A_2^c) \cup \ldots \cup (A_1 \cap \ldots \cap A_{n-1} \cap A_n^c),$$

and similarly so is $\mathcal{S}_2$. Further

$$\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_m \subset \mathcal{S}_1 \subset \bigvee_{i=1}^m \mathcal{A}_i,$$

implying

$$\sigma(\mathcal{S}_1) = \bigvee_{i=1}^m \mathcal{A}_i,$$

and similarly

$$\sigma(\mathcal{S}_2) = \bigvee_{i=m+1}^n \mathcal{A}_i.$$

Thus in view of Theorem 6.1.2, the proof would follow once it is shown that

$$P(A \cap B) = P(A)P(B) \text{ for all } A \in \mathcal{S}_1, B \in \mathcal{S}_2.$$

The above however follows directly from the given hypothesis, which completes the proof. $\square$

The rest of this subsection focuses on independence of random variables.

Theorem 6.1.4. *If X and Y are independent non-negative random variables, then*

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y).$$

If X and Y are independent integrable random variables, then XY is integrable and

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y).$$

Proof. The independence of X and Y implies that

$$P\big((X,Y) \in \cdot\big) = P(X \in \cdot) \otimes P(Y \in \cdot)$$

as measures on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ (recall that $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$ from Exc 5.3.1). Since X and Y are non-negative,

$$\begin{aligned}
\mathrm{E}(XY) &= \int_{[0,\infty)\times[0,\infty)} xy\, P\big((X,Y) \in (dx, dy)\big) \\
&= \int_{[0,\infty)\times[0,\infty)} xy\, P(X \in dx) \otimes P(Y \in dy) \\
&= \int_{[0,\infty)} y \int_{[0,\infty)} x\, P(X \in dx)\, P(Y \in dy) \\
&= \mathrm{E}(X)\mathrm{E}(Y),
\end{aligned}$$

the penultimate line being implied by Theorem 5.2.1, which is Tonelli's theorem.

If X and Y are independent integrable random variables, then $|X|$ and $|Y|$ are independent as they are measurable with respect to $\sigma(X)$ and $\sigma(Y)$, respectively. By the already proven result for non-negative random variables, we get

$$\mathrm{E}\left(|XY|\right) = \mathrm{E}(|X|)\mathrm{E}(|Y|) < \infty\,.$$

Thus XY is integrable. Splitting $X = X^+ - X^-$ and likewise for Y, and using that $X^\pm$ is independent of $Y^\pm$, the final claim follows. $\square$

The following result gives an easy way of checking independence of random variables.

Theorem 6.1.5. *Random variables $X_1, \ldots, X_n$ are independent if and only if*

$$P(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{i=1}^{n} P(X_i \le x_i) \text{ for all } x_1, \ldots, x_n \in \mathbb{R}\,. \tag{6.1.2}$$

Proof. The 'only if' part follows trivially from (6.1.1), which is essentially the definition of independence, by taking $B_i = (-\infty, x_i]$ for $i = 1, \ldots, n$.

Conversely, assume (6.1.2). Define

$$\mathcal{S}_i = \left\{ X_i^{-1}\left((a,b] \cap \mathbb{R}\right) :\, -\infty \le a \le b \le \infty \right\},\ i = 1, \ldots, n\,.$$

Since $\sigma(\mathcal{S}_i) = \{X_i^{-1}B : B \in \mathcal{B}(\mathbb{R})\} = \sigma(X_i)$ and $\mathcal{S}_i$ is a semi-field for $i = 1, \ldots, n$, Theorem 6.1.2 would complete the proof once it is shown that

$$P\left(a_i < X_i \le b_i,\, i = 1, \ldots, n\right) = \prod_{i=1}^{n} P(a_i < X_i \le b_i), \tag{6.1.3}$$

whenever $-\infty \le a_i \le b_i \le \infty$ for $i = 1, \ldots, n$.

Fix a_i, b_i as above and write

$$
\begin{aligned}
&P\left(a_i < X_i \le b_i,\, i = 1, \ldots, n\right) \\
&= \mathrm{E}\big(\mathbf{1}\left(a_i < X_i \le b_i,\, i = 1, \ldots, n\right)\big) \\
&= \mathrm{E}\left(\prod_{i=1}^{n} \mathbf{1}(a_i < X_i \le b_i)\right) \\
&= \mathrm{E}\left(\prod_{i=1}^{n} \big(\mathbf{1}(X_i \le b_i) - \mathbf{1}(X_i \le a_i)\big)\right) \\
&= \mathrm{E}\left(\sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} (-1)^{\#\{i:x_i=a_i\}} \prod_{j=1}^{n} \mathbf{1}(X_i \le x_i)\right) \\
&= \mathrm{E}\left(\sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} (-1)^{\#\{i:x_i=a_i\}} \mathbf{1}(X_1 \le x_1, \ldots, X_n \le x_n)\right) \\
&= \sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} (-1)^{\#\{i:x_i=a_i\}} P\left(X_1 \le x_1, \ldots, X_n \le x_n\right) \\
&= \sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} (-1)^{\#\{i:x_i=a_i\}} \prod_{i=1}^{n} P(X_i \le x_i) \\
&= \prod_{i=1}^{n} \big(P(X_i \le b_i) - P(X_i \le a_i)\big) = \prod_{i=1}^{n} P(a_i < X_i \le b_i),
\end{aligned}
$$

(6.1.2) being used for the equality in the penultimate line. Thus (6.1.3) follows, which proves the 'if' part and hence completes the proof. □

Exercise 6.1.3. *Suppose $X_1, \ldots, X_n$ are independent random variables.*

1. *Show that*

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i),\; x_1, \ldots, x_n \in \mathbb{R}\,.$$

2. *Assuming that $f_1, \ldots, f_n$ are the respective densities, as defined in* (3.6.1)*, of $X_1, \ldots, X_n$, show that the joint density f of $(X_1, \ldots, X_n)$, as in* (4.2.1)*, is*

$$f(x) = \prod_{i=1}^{n} f_i(x_i),\; x = (x_1, \ldots, x_n) \in \mathbb{R}^n\,.$$

Exercise 6.1.4. 1. *If $X_1, \ldots, X_n$ are random variables, such that there exist countable sets $C_1, \ldots, C_n$ satisfying*

$$P(X_i \in C_i) = 1,\; i = 1, \ldots, n,$$

show that $X_1, \ldots, X_n$ are independent if and only if

$$P\left(\bigcap_{i=1}^{n}[X_i = x_i]\right) = \prod_{i=1}^{n} P(X_i = x_i) \text{ for all } x_1 \in C_1, \ldots, x_n \in C_n\,.$$

2. *Suppose $(X_1, \ldots, X_n)$ has a joint density f, which can be written as*

$$f(x) = c\prod_{i=1}^{n} g_i(x_i),\ x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

for some $c \in \mathbb{R}$ and one-dimensional densities $g_1, \ldots, g_n$, that is, g_i is a non-negative Borel function on $\mathbb{R}$ satisfying $\int_{\mathbb{R}} g_i(x)\,dx = 1$ for $i = 1, \ldots, n$. Show that $c = 1$ and $X_1, \ldots, X_n$ are independent with respective densities $g_1, \ldots, g_n$.

Example 6.1.1. *Let $X \sim$ Gamma(α) and $Y \sim$ Gamma(β) independently of each other, that is, each of them has density r defined by*

$$r(x) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{-\alpha-1} \mathbf{1}(x > 0),$$

where

$$\Gamma(z) = \int_0^\infty e^{-x} x^{-z-1}\,dx,\ z > 0\,.$$

We want to find the distribution of $W = X/(X+Y)$.

Theorem 4.3.2 is the only tool at our disposal, which is valid for one-one functions from an open subset of $\mathbb{R}^2$ to $\mathbb{R}^2$. Therefore, we define an auxiliary random variable $Z = X + Y$. Thus, $(W, Z) = \psi(X, Y)$ where $\psi : U \to V$ is a bijection defined by

$$\psi(x, y) = \left(\frac{x}{x+y}, x + y\right), (x, y) \in U\,,$$

and $U = (0, \infty)^2$ and $V = (0, 1) \times (0, \infty)$ are open sets. The inverse of ψ is $T : V \to U$ defined by

$$T(w, z) = (wz, z - wz)\,, (w, z) \in V\,.$$

The Jacobian matrix of T is

$$J(w, z) = \begin{bmatrix} z & w \\ -z & 1 - w \end{bmatrix},$$

showing $|\det J(w, z)| = z$ for $(w, z) \in V$. The joint density of (X, Y) is

$$f(x, y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} e^{-x-y} x^{\alpha-1} y^{\beta-1}\,, (x, y) \in U\,.$$

Theorem 4.3.2 shows that the joint density g of (W, Z) at $(w, z) \in V$ is

$$\begin{aligned}
g(w, z) &= f \circ T(w, z) |\det J(w, z)| \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} e^{-z} (wz)^{\alpha-1} (z - wz)^{\beta-1} z \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1} (1 - w)^{\beta-1} e^{-z} z^{\alpha+\beta-1} ,
\end{aligned}$$

and $g(w, z) = 0$ for $(w, z) \notin V$. In other words,

$$g(w, z) = c\, h_1(w) h_2(z) \, , (w, z) \in \mathbb{R}^2 \, ,$$

where

$$h_1(w) = \frac{1}{B(\alpha, \beta)} w^{\alpha-1} (1 - w)^{\beta-1} \mathbf{1}(0 < w < 1) \, , w \in \mathbb{R} \, ,$$

$$h_2(z) = \frac{1}{\Gamma(\alpha + \beta)} e^{-z} z^{\alpha+\beta-1} \mathbf{1}(z > 0) \, , z \in \mathbb{R} \, ,$$

and

$$c = \frac{B(\alpha, \beta)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \, .$$

Since h_1 and h_2 are densities of Beta(α, β) and Gamma$(\alpha+\beta)$, respectively, it follows that $c = 1$ and W and Z are independent with respective densities h_1 and h_2. In particular, this means $X/(X + Y)$ follows Beta(α, β). Furthermore, $c = 1$ proves that

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \, , \alpha, \beta > 0 \, .$$

6.2 Kolmogorov's zero-one law

Theorem 6.2.1 (Kolmogorov's zero-one law). *If $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are independent σ-fields, then the tail σ-field*

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \bigvee_{m=n}^{\infty} \mathcal{A}_m$$

is trivial, that is, for all $A \in \mathcal{T}$, $P(A)$ is either 0 or 1.

Proof. The first step in the proof is the following.
Step 1. For all $n \geq 1$, $\mathcal{T}$ is independent of $\bigvee_{i=1}^{n} \mathcal{A}_i$.

Proof of Step 1. Fix $n \geq 1$. For any $m > n$, Theorem 6.1.3 implies $\bigvee_{i=1}^{n} \mathcal{A}_i$ is independent of $\bigvee_{i=n+1}^{m} \mathcal{A}_i$. Since the above holds for all $m > n$, it follows that

$$P(A \cap B) = P(A)P(B) \text{ for all } A \in \bigvee_{i=1}^{n} \mathcal{A}_i, \, B \in \bigcup_{m=n+1}^{\infty} \bigvee_{i=n+1}^{m} \mathcal{A}_i \, .$$

Since $\bigcup_{m=n+1}^{\infty} \bigvee_{i=n+1}^{m} \mathcal{A}_i$ is a field (and $\bigvee_{i=1}^{n} \mathcal{A}_i$ is a σ-field), Theorem 6.1.2 implies that

$$\bigvee_{i=1}^{n} \mathcal{A}_i \text{ and } \sigma\left(\bigcup_{m=n+1}^{\infty} \bigvee_{i=n+1}^{m} \mathcal{A}_i\right)$$

are independent. Since $\sigma\left(\bigcup_{m=n+1}^{\infty} \bigvee_{i=n+1}^{m} \mathcal{A}_i\right) = \bigvee_{i=n+1}^{\infty} \mathcal{A}_i \supset \mathcal{T}$, Step 1 follows. □

Step 2. The tail σ-field $\mathcal{T}$ is independent of $\bigvee_{i=1}^{\infty} \mathcal{A}_i$.

Proof of Step 2. A consequence of Step 1 is that

$$P(A \cap B) = P(A)P(B) \text{ for all } A \in \bigcup_{n=1}^{\infty} \bigvee_{i=1}^{n} \mathcal{A}_i,\ B \in \mathcal{T}.$$

Since $\bigcup_{n=1}^{\infty} \bigvee_{i=1}^{n} \mathcal{A}_i$ is a field, the σ-field generated by which is $\bigvee_{i=1}^{\infty} \mathcal{A}_i$, Step 2 follows by using Theorem 6.1.2 once again. □

Since $\bigvee_{i=1}^{\infty} \mathcal{A}_i \supset \mathcal{T}$, Step 2 implies that $\mathcal{T}$ is independent of itself. In other words,

$$P(A) = P(A \cap A) = P(A)^2 \text{ for all } A \in \mathcal{T}.$$

Thus $P(A)$ equals either 0 or 1 for all $A \in \mathcal{T}$, as claimed in the statement. This completes the proof. □

6.3 The strong law of large numbers

Definition. *Random variables X_n* converge almost surely (or a.s.) *to X if $X_n \to X$ a.e., that is,*

$$P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

The following is a trivial consequence of Kolmogorov's zero-one law.

Exercise 6.3.1. *Suppose $X_1, X_2, \ldots$ are independent random variable such that as $n \to \infty$,*

$$n^{-1} \sum_{i=1}^{n} X_i \to X \text{ a.s.}$$

Show that X is a degenerate random variable.

The following theorem, known as the strong law of large numbers or SLLN, shows that if $X_1, X_2, \ldots$ are i.i.d. with finite mean μ, then the above holds with $X = \mu$.

Theorem 6.3.1 (SLLN)**.** *For i.i.d. random variables $X_1, X_2, \ldots$ with finite mean μ,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \ a.s.,$$

as $n \to \infty$.

For proving the SLLN, the following inequality will be used.

Theorem 6.3.2 (Kolmogorov maximal inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables with finite variance. Then, for any $\alpha > 0$,*

$$P\left(\max_{1\le k\le n} |S_k - \mathrm{E}(S_k)| \ge \alpha\right) \le \alpha^{-2}\mathrm{Var}(S_n)\,,$$

where

$$S_k = \sum_{i=1}^{k} X_i,\ 1 \le k \le n\,.$$

The following inequality obtained by putting $n = 1$ above is known as Chebyshev's inequality in probability theory. This follows directly from Markov's inequality (Theorem 3.3.4) as well.

Corollary 6.3.1 (Chebyshev's inequality)**.** *If X has mean μ and finite variance σ^2, then*

$$P\left(|X - \mu| \ge \alpha\right) \le \alpha^{-2}\sigma^2\,, \alpha > 0\,.$$

Proof of Theorem 6.3.2. WLOG, assume that $X_1, \ldots, X_n$ are zero mean. We start with the observation that

$$\left[\max_{1\le k\le n} |S_k| \ge \alpha\right] = \bigcup_{k=1}^{n} A_k\,,$$

where

$$A_k = \left[|S_k| \ge \alpha > |S_j| \text{ for all } 1 \le j \le k-1\right],\ k = 1, \ldots, n\,.$$

Since $A_1, \ldots, A_n$ are disjoint, it follows that

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{E}(S_n^2) \\
&= \sum_{k=1}^{n} \mathrm{E}\left(S_n^2 \mathbf{1}_{A_k}\right) \\
&= \sum_{k=1}^{n} \left[\mathrm{E}\left((S_n - S_k)^2 \mathbf{1}_{A_k}\right) + \mathrm{E}\left(S_k^2 \mathbf{1}_{A_k}\right) + 2\mathrm{E}\left((S_n - S_k)S_k \mathbf{1}_{A_k}\right)\right] \\
&\ge \sum_{k=1}^{n} \left[\mathrm{E}\left(S_k^2 \mathbf{1}_{A_k}\right) + 2\mathrm{E}\left((S_n - S_k)S_k \mathbf{1}_{A_k}\right)\right]\,.
\end{aligned}$$

Since $S_n - S_k$ and $S_k \mathbf{1}_{A_k}$ are independent and the former has zero mean, it follows that

$$\mathrm{E}\left((S_n - S_k) S_k \mathbf{1}_{A_k}\right) = 0\,,$$

and hence

$$\begin{aligned}
\mathrm{Var}(S_n) &\geq \sum_{k=1}^{n} \mathrm{E}\left(S_k^2 \mathbf{1}_{A_k}\right) \\
&\geq \sum_{k=1}^{n} \mathrm{E}\left(\alpha^2 \mathbf{1}_{A_k}\right) \\
&= \alpha^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq \alpha\right).
\end{aligned}$$

This completes the proof. □

Exercise 6.3.2. *Prove the following.*

1. *If X and Y are independent and integrable random variables, then* $\mathrm{Cov}(X, Y)$ *exists and equals zero.*

2. *If $X_1, \ldots, X_n$ are independent random variables each having a finite variance, then*

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\,.$$

Proof of Theorem 6.3.1. WLOG, assume that $E(X_1) = 0$. For $n \geq 1$, define

$$S_n := \sum_{i=1}^{n} X_i\,,$$

$$X_n' := X_n \mathbf{1}(|X_n| \leq n)\,,$$

and

$$S_n' := \sum_{k=1}^{n} X_k'\,.$$

Notice that

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(X_n \neq X_n'\right) &= \sum_{n=1}^{\infty} P(|X_1| > n) \\
&\leq \sum_{n=1}^{\infty} \int_{n-1}^{n} P(|X_1| > s) ds \qquad (6.3.1) \\
&= \int_0^{\infty} P(|X_1| > s) ds \\
&= E(|X_1|) \\
&< \infty\,,
\end{aligned}$$

(6.3.1) following from the observation that $P(|X_1| > n) \le P(|X_1| > s)$ for $s \le n$. From the Borel-Cantelli lemma (Theorem 3.4.3), it follows that

$$\limsup_{n\to\infty} \left|S_n - S'_n\right| < \infty$$

almost surely, and hence

$$n^{-1}S_n - n^{-1}S'_n \to 0$$

almost surely. So it suffices to show that

$$n^{-1}S'_n \to 0 \text{ a.s.} \tag{6.3.2}$$

Notice that by DCT,

$$E(X'_n) = E\left[X_1 \mathbf{1}(|X_1| \le n)\right] \to E(X_1) = 0\,,$$

and hence,

$$\lim_{n\to\infty} n^{-1}\mathrm{E}(S'_n) = 0\,.$$

Therefore, (6.3.2) will follow if we can show that

$$n^{-1}\left[S'_n - E(S'_n)\right] \to 0 \text{ a.s.}$$

For $r \ge 1$, set

$$Z_r := \max_{2^{r-1} \le k < 2^r} |S'_k - E(S'_k)|\,.$$

Since

$$\frac{1}{k}|S'_k - E(S'_k)| \le 2^{-(r-1)} Z_r \text{ for all } 2^{r-1} \le k \le 2^r\,,$$

it suffices to show that

$$2^{-r} Z_r \to 0 \text{ a.s.}$$

The above would follow from Theorems 3.4.3 and 3.4.4 if it can be shown that

$$\sum_{r=1}^{\infty} P\left[|Z_r| > 2^r \varepsilon\right] < \infty\,.$$

for any $\varepsilon > 0$. Kolmogorov's inequality implies that

$$\begin{aligned}
\sum_{r=1}^{\infty} P\left[|Z_r| > 2^r\varepsilon\right] &\leq \sum_{r=1}^{\infty} P\left(\max_{1\leq k\leq 2^r} \left|S'_k - \mathrm{E}(S'_k)\right| > 2^r\varepsilon\right) \\
&\leq \sum_{r=1}^{\infty} \varepsilon^{-2}4^{-r}\mathrm{Var}(S'_{2^r}) \\
\text{(Exc 6.3.2)} \quad &= \varepsilon^{-2}\sum_{r=1}^{\infty} 4^{-r}\sum_{j=1}^{2^r}\mathrm{Var}(X'_j) \\
&= \varepsilon^{-2}\sum_{j=1}^{\infty}\mathrm{Var}(X'_j)\sum_{r=\lceil \log_2 j\rceil}^{\infty} 4^{-r} \\
&\leq K\sum_{j=1}^{\infty} j^{-2}\mathrm{Var}(X'_j)\,,
\end{aligned}$$

the last line following from the calculation that

$$\begin{aligned}
\sum_{r=\lceil \log_2 j\rceil}^{\infty} 4^{-r} &= \frac{4}{3}4^{-\lceil \log_2 j\rceil} \\
&\leq \frac{4}{3}4^{-\log_2 j} \\
&= \frac{4}{3}j^{-2}\,.
\end{aligned}$$

Thus, in order to complete the proof, all that needs to be shown is that

$$\sum_{j=1}^{\infty} j^{-2}\mathrm{Var}(X'_j) < \infty\,.$$

To that end, observe that

$$\begin{aligned}
\sum_{j=1}^{\infty} j^{-2}\mathrm{Var}(X'_j) &\leq \sum_{j=1}^{\infty} j^{-2}E({X'_j}^2) \\
&= \sum_{j=1}^{\infty} j^{-2}E(X_1^2\mathbf{1}(|X_1|\leq j)) \\
&= \sum_{k=1}^{\infty} E(X_1^2\mathbf{1}(k-1<|X_1|\leq k))\sum_{j=k}^{\infty} j^{-2} \\
&\leq \sum_{k=1}^{\infty} E(X_1^2\mathbf{1}(k-1<|X_1|\leq k))2/k \qquad (6.3.3) \\
&\leq 2\sum_{k=1}^{\infty} E(|X_1|\mathbf{1}(k-1<|X_1|\leq k)) \\
&= 2E|X_1| < \infty\,,
\end{aligned}$$

(6.3.3) following from the fact that for $k \geq 2$,

$$\sum_{j=k}^{\infty} j^{-2} \leq \sum_{j=k}^{\infty} \int_{j-1}^{j} x^{-2}\, dx = \frac{1}{k-1} \leq \frac{2}{k}\,,$$

and

$$\sum_{j=1}^{\infty} j^{-2} \leq 1 + \int_{1}^{\infty} x^{-2}\, dx = 2\,,$$

which together imply

$$\sum_{j=k}^{\infty} j^{-2} \leq \frac{2}{k}\,, k \in \mathbb{N}\,.$$

Hence, the proof follows. $\square$

6.4 The second Borel-Cantelli lemma

Definition. *Events* $A_1, A_2, \ldots$ *are* independent *if* $\sigma(\{A_1\}), \sigma(\{A_2\}), \ldots$ *are independent* σ*-fields.*

The following theorem is a converse of Theorem 3.4.3, the first Borel-Cantelli lemma, when the events are independent.

Theorem 6.4.1 (Second Borel-Cantelli lemma)**.** *If* $A_1, A_2, \ldots$ *are independent events such that*

$$\sum_{n=1}^{\infty} P(A_n) = \infty\,,$$

then

$$P\left(A_n \text{ occurs for infinitely many } n\right) = 1\,.$$

Proof. Since

$$E = [A_n \text{ occurs for infinitely many } n] \in \bigcap_{n=1}^{\infty} \bigvee_{k=n}^{\infty} \{A_k, A_k^c, \emptyset, \Omega\}\,,$$

Kolmogorov's zero-one law (Theorem 6.2.1) shows $P(E)$ is either 0 or 1. Thus it suffices to show

$$P(E) > 0\,. \tag{6.4.1}$$

Recall that for any $n \geq 1$,

$$
\begin{aligned}
P\left(\bigcup_{i=1}^n A_i\right) &= 1 - P\left(\bigcap_{i=1}^n A_i^c\right) \\
&= 1 - \prod_{i=1}^n (1 - P(A_i)) \\
\big(1 - x \leq e^{-x} \text{ for all } x \in \mathbb{R}\big) \quad &\geq 1 - \prod_{i=1}^n e^{-P(A_i)} \\
&= 1 - \exp\left(-\sum_{i=1}^n P(A_i)\right) \\
&\to 1\,,
\end{aligned}
$$

as $n \to \infty$ because $\sum_{i=1}^\infty P(A_i) = \infty$.

Let $\alpha_1, \alpha_2, \ldots \in (0,1)$ be such that

$$\prod_{i=1}^\infty \alpha_i > 0\,.$$

For example, $\alpha_i = e^{-1/i^2}$ for $i = 1, 2, \ldots$ satisfies the above. The above calculations show there exists n_1 such that

$$P\left(\bigcup_{i=1}^{n_1} A_i\right) \geq \alpha_1\,.$$

Since $\sum_{i=n_1+1}^\infty P(A_i) = \infty$, a similar calculation shows there exists $n_2 > n_1$ such that

$$P\left(\bigcup_{i=n_1+1}^{n_2} A_i\right) \geq \alpha_2\,.$$

Proceeding inductively, get integers $0 = n_0 < n_1 < n_2 < \ldots$ such that

$$P\left(\bigcup_{i=n_{k-1}+1}^{n_k} A_i\right) \geq \alpha_k\,, k \in \mathbb{N}\,.$$

Clearly,

$$E \supset \bigcap_{k=1}^\infty \bigcup_{i=n_{k-1}+1}^{n_k} A_i\,.$$

Therefore,

$$\begin{aligned}
P(E) &\geq P\left(\bigcap_{k=1}^{\infty} \bigcup_{i=n_{k-1}+1}^{n_k} A_i\right) \\
&= \prod_{k=1}^{\infty} P\left(\bigcup_{i=n_{k-1}+1}^{n_k} A_i\right) \\
&\geq \prod_{k=1}^{\infty} \alpha_k > 0 \,.
\end{aligned}$$

Thus (6.4.1) holds, from which the proof follows. □

An immediate consequence of the second Borel-Cantelli lemma is the following, which should be compared with Theorem 3.4.4.

Exercise 6.4.1. *If $X_1, X_2, X_3, \ldots$ are independent random variables, then show that*

$$X_n \to X \ \textit{a.s.} \iff \sum_{n=1}^{\infty} P\left(|X_n - X| > \varepsilon\right) < \infty \text{ for all } \varepsilon > 0 \,.$$

Show that if the above holds, then X is a degenerate random variable.

Exercise 6.4.2. *Suppose $X_1, X_2, \ldots$ are random variables such that*

$$\sum_{n=1}^{\infty} \mathrm{E}(X_n^2) < \infty \,.$$

If $Y_1, Y_2, \ldots$ are such that $\sigma(X_n : n \geq 1), \sigma(Y_1), \sigma(Y_2), \ldots$ are independent and Y_n takes values 1 and -1, each with probability $1/2$ for $n = 1, 2, \ldots$, show that

$$\sum_{i=1}^{n} X_i Y_i \to Z \,, \ \text{as } n \to \infty \,,$$

in L^2, for some $Z \in L^2(\Omega)$.

Soln.: Let

$$Z_n = \sum_{i=1}^{n} X_i Y_i \,, n \geq 1 \,.$$

Theorem 3.4.6 implies $L^2(\Omega)$ is a complete metric space, it suffices to show that $\{Z_n : n \geq 1\}$ is a Cauchy sequence in $L^2(\Omega)$. For $1 \leq m < n$,

$$\begin{aligned}
\mathrm{E}\left[(Z_n - Z_m)^2\right] &= \mathrm{E}\left[\left(\sum_{i=m+1}^{n} X_i Y_i\right)^2\right] \\
\left(Y_i^2 = 1\right) &= \sum_{i=m+1}^{n} \mathrm{E}(X_i^2) + 2 \sum_{m+1 \leq i < j \leq n} \mathrm{E}\left(X_i X_j Y_i Y_j\right) \,.
\end{aligned}$$

For $m+1 \le i < j \le n$, independence of $\sigma(X_1, X_2, \ldots), \sigma(Y_i), \sigma(Y_j)$ shows

$$\mathrm{E}\left(X_i X_j Y_i Y_j\right) = \mathrm{E}(X_i X_j)\mathrm{E}(Y_i)\mathrm{E}(Y_j) = 0\,.$$

Thus,

$$\mathrm{E}\left[(Z_n - Z_m)^2\right] = \sum_{i=m+1}^{n} \mathrm{E}(X_i^2)\,.$$

Given $\varepsilon > 0$, choosing N such that

$$\sum_{i=N+1}^{\infty} \mathrm{E}(X_i^2) \le \varepsilon\,,$$

which is possible from the given hypothesis, it holds that for $N \le m < n$,

$$\mathrm{E}\left[(Z_n - Z_m)^2\right] = \sum_{i=m+1}^{n} \mathrm{E}(X_i^2) \le \sum_{i=N+1}^{\infty} \mathrm{E}(X_i^2) \le \varepsilon\,,$$

showing $\{Z_n : n \ge 1\}$ is a Cauchy sequence in $L^2(\Omega)$.

7 Modes of convergence

Convergence in L^p and a.s. convergence have been defined in Subsections 3.4 and 6.3, respectively. In this section, a few more modes of convergence will be studied. The first one of them is convergence in probability, which is weaker than both convergence in L^p and a.s. convergence. As usual, $(\Omega, \mathcal{A}, P)$ is the probability space underlying all random variables which are real-valued, unless mentioned otherwise.

7.1 Convergence in probability

Definition. *For random variables $X, X_1, X_2, \ldots$, we say X_n* converges in probability *to X and write $X_n \xrightarrow{P} X$ if*

$$\lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0 \textit{ for all } \varepsilon > 0\,.$$

Theorem 7.1.1. *For random variables $X, X_1, X_2, \ldots$, $X_n \xrightarrow{P} X$ if either $X_n \to X$ in L^p for some $1 \le p \le \infty$ or $X_n \to X$ a.s.*

Proof. Markov's inequality implies that convergence in L^p implies convergence in probability, while DCT shows that a.s. convergence implies convergence in probability. □

Exercise 7.1.1. *If $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} X'$, then show that $X = X'$, a.s.*

Example 7.1.1. *Let $\Omega = (0,1]$, $\mathcal{A} = \mathcal{B}((0,1])$ and P be the restriction of Lebesgue measure to $(0,1]$. Define for all $\omega \in \Omega$,*

$$\begin{aligned} X_1(\omega) &= \mathbf{1}\left(0 < \omega \le \frac{1}{2}\right), \\ X_2(\omega) &= \mathbf{1}\left(\frac{1}{2} < \omega \le 1\right), \\ X_3(\omega) &= \mathbf{1}\left(0 < \omega \le \frac{1}{4}\right), \\ X_4(\omega) &= \mathbf{1}\left(\frac{1}{4} < \omega \le \frac{1}{2}\right), \\ X_5(\omega) &= \mathbf{1}\left(\frac{1}{2} < \omega \le \frac{3}{4}\right), \\ X_6(\omega) &= \mathbf{1}\left(\frac{3}{4} < \omega \le 1\right), \\ &\vdots \end{aligned}$$

and so on. Then, $X_n \xrightarrow{P} 0$ but

$$P\left(\lim_{n\to\infty} X_n = 0\right) = 0\,.$$

The above example shows that convergence in probability is a strictly weaker notion of convergence than almost sure convergence.

Exercise 7.1.2. *If $X_n \to Y$ in L^p for some $1 \le p < \infty$ and $X_n \to Z$ a.s., show that*

$$Y = Z \text{ a.s.}$$

Exercise 7.1.3.
1. *In Example 7.1.1, show that for all $p \in [1,\infty)$, $X_n \to 0$ in L^p.*
2. *Show that for $p \in [1,\infty)$, convergence in L^p neither implies nor is implied by a.s. convergence.*

Theorem 7.1.2. *If $X_n \xrightarrow{P} X$, then X_n has a subsequence X_{n_k} such that*

$$X_{n_k} \to X \text{ a.s.},$$

as $k \to \infty$.

Proof. Since $X_n \xrightarrow{P} X$, there exists n_1 such that

$$P(|X_{n_1} - X| > 1) \le \frac{1}{2}\,.$$

There exists N_2 such that

$$P\left(|X_n - X| > \frac{1}{2}\right) \le 2^{-2} \text{ for all } n \ge N_2 \, .$$

Define $n_2 = N_2 \vee (n_1 + 1)$. Proceeding similarly, we get positive integers $n_1 < n_2 < n_3 < \ldots$ such that

$$P\left(|X_{n_k} - X| > \frac{1}{k}\right) \le 2^{-k} \text{ for all } k \, .$$

Hence,

$$\sum_{k=1}^{\infty} P\left(|X_{n_k} - X| > \frac{1}{k}\right) < \infty \, .$$

The Borel-Cantelli Lemma implies that

$$P\left(|X_{n_k} - X| > \frac{1}{k} \text{ for infinitely many } k\right) = 0 \, .$$

Thus,

$$X_{n_k} \to X \text{ a.s.} \, ,$$

as $k \to \infty$. This completes the proof. $\square$

Theorem 7.1.3. *If $X_n \xrightarrow{P} X$ and $|X_n| \le Y$ for some Y with $\mathrm{E}(Y) < \infty$, then prove that*

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X) \, .$$

Proof. Exercise. $\square$

Exercise 7.1.4. *Prove or disprove the following claim. If X_n and X are random variables such that any subsequence $\{X_{n_k} : k \ge 1\}$ of X_n has a further subsequence $\{X_{n_{k_l}} : l \ge 1\}$ such that*

$$X_{n_{k_l}} \to X \text{ a.s.} \, ,$$

then $X_n \to X$ a.s.

Exercise 7.1.5. *Show that the following are equivalent for random variables X_n and X.*

1. *As $n \to \infty$, $X_n \xrightarrow{P} X$.*
2. *Every subsequence $\{X_{n_k} : k \ge 1\}$ of $\{X_n : n \ge 1\}$ has a further subsequence $\{X_{n_{k_l}} : l \ge 1\}$ such that as $l \to \infty$,*

$$X_{n_{k_l}} \to X \text{ a.s.}$$

3. *Every subsequence $\{X_{n_k} : k \ge 1\}$ of $\{X_n : n \ge 1\}$ has a further subsequence $\{X_{n_{k_l}} : l \ge 1\}$ such that as $l \to \infty$,*

$$X_{n_{k_l}} \xrightarrow{P} X \, .$$

7.2 Weak convergence

Definition. *For probability measures $\mu, \mu_1\mu_2, \ldots$ on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, μ_n* converges weakly *to μ, written as $\mu_n \Rightarrow \mu$, if*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f\, d\mu_n = \int_{\mathbb{R}^d} f\, d\mu \text{ for all bounded continuous function } f : \mathbb{R}^d \to \mathbb{R}\,.$$

For $\mathbb{R}^d$-valued random variables $X, X_1, X_2, \ldots$, X_n converges weakly, in law or in distribution *to X, written as $X_n \Rightarrow X$, if*

$$P \circ X_n^{-1} \Rightarrow P \circ X^{-1},\, n \to \infty\,.$$

Theorem 7.2.1. *If $\mu, \mu', \mu_1, \mu_2, \ldots$ are probability measures on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ such that $\mu_n \Rightarrow \mu$ and $\mu_n \Rightarrow \mu'$, then*

$$\mu = \mu'\,.$$

The proof uses the following exercise which is a simple application of the good set principle.

Exercise 7.2.1. *Show that for any probability measure P on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$,*

$$P(A) = \inf\{P(U) : U \text{ open}, U \supset A\} = \sup\{P(F) : F \text{ closed}, F \subset A\}\,.$$

Proof of Theorem 7.2.1. In view of Exc 7.2.1, it suffices to show that

$$\mu(F) = \mu'(F) \text{ for all closed } F \subset \mathbb{R}^d\,. \tag{7.2.1}$$

Fix a closed $F \subset \mathbb{R}^d$.

Let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^d$. Define

$$d(x, F) = \inf\left\{\|x - y\| : y \in F\right\},\, x \in \mathbb{R}^d\,. \tag{7.2.2}$$

We shall first show that $d(\cdot, F)$ is a continuous function with respect to the norm $\|\cdot\|$. For $x, x' \in \mathbb{R}^d$ and $y \in F$, the triangle inequality implies

$$\|x - y\| \le \|x - x'\| + \|x' - y\|\,.$$

Taking inf over $y \in F$ shows

$$d(x, F) \le \|x - x'\| + d(x', F)\,.$$

Interchanging the roles of x and x', in conjunction with the above, implies

$$\big|d(x, F) - d(x', F)\big| \le \|x - x'\|,\, x, x' \in \mathbb{R}^d\,. \tag{7.2.3}$$

Thus $d(\cdot, F)$ is a continuous function. Fix $\varepsilon > 0$ and define

$$f_\varepsilon(x) = 1 \wedge \left(\varepsilon^{-1} d(x, F)\right),\, x \in \mathbb{R}^d\,. \tag{7.2.4}$$

Since $d(\cdot, F)$ is continuous, so is f_ε. Since f_ε is bounded as well, the hypothesis implies

$$\int f_\varepsilon \, d\mu = \int f_\varepsilon \, d\mu' . \tag{7.2.5}$$

Since

$$\mathbf{1}_{F_\varepsilon^c} \le f_\varepsilon \le \mathbf{1}_{F^c}, \tag{7.2.6}$$

where $F_\varepsilon = \{x \in \mathbb{R}^d : d(x, F) \le \varepsilon\}$, (7.2.5) implies

$$\mu\left(F_\varepsilon^c\right) \le \mu'\left(F^c\right),$$

that is,

$$\mu(F_\varepsilon) \ge \mu'(F) .$$

Since F is closed,

$$F = \{x : d(x, F) = 0\} = \bigcap_{k=1}^{\infty} F_{1/k},$$

showing that

$$\mu(F_{1/k}) \downarrow \mu(F), \; k \to \infty . \tag{7.2.7}$$

Thus $\mu(F) \ge \mu'(F)$. Interchanging the roles of μ and μ' implies the reverse inequality, from which, (7.2.1) follows and completes the proof. $\square$

The following is a trivial exercise, the solution of which follows from the definition.

Exercise 7.2.2. *Show that for $\mathbb{R}^d$-valued random variables $X, X_1, X_2, \ldots$, $X_n \Rightarrow X$ if and only if*

$$\lim_{n\to\infty} \mathrm{E}\left(f(X_n)\right) = \mathrm{E}\left(f(X)\right)$$

for all bounded continuous $f : \mathbb{R}^d \to \mathbb{R}$.

Theorem 7.2.2 (Continuous mapping theorem)**.** *If $X, X_1, X_2, \ldots$ are $\mathbb{R}^{d_1}$-valued random variables and $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ is a continuous function, then*

$$f(X_n) \Rightarrow f(X) \text{ in } \mathbb{R}^{d_2} .$$

Proof. Follows from the definition of weak convergence. $\square$

Theorem 7.2.3 (Portmanteau theorem on $\mathbb{R}^d$)**.** *The following are equivalent for probability measures $\mu, \mu_1\mu_2, \ldots$ on $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$.*

1. *As $n \to \infty$, $\mu_n \Rightarrow \mu$.*

2. *For all bounded uniformly continuous functions $f : \mathbb{R}^d \to \mathbb{R}$,*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f \, d\mu_n = \int_{\mathbb{R}^d} f \, d\mu .$$

3. For all closed set $F \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} \mu_n(F) \le \mu(F)\,.$$

4. For all open set $U \subset \mathbb{R}^d$,

$$\liminf_{n\to\infty} \mu_n(U) \ge \mu(U)\,.$$

Proof. We shall show that 1⇒2⇒3⇒4⇒1.

Proof of 1 ⇒2. Follows tautologically from the definition of weak convergence. □

Proof of 2 ⇒3. Assume that 2 holds. Fix a closed set $F \subset \mathbb{R}^d$ and $\varepsilon > 0$. Let $d(\cdot, F)$ and $f_\varepsilon(\cdot)$ be as in (7.2.2) and (7.2.4), respectively. It follows from (7.2.3) that $d(\cdot, F)$ is uniformly continuous, and hence so if f_ε. A restatement of (7.2.6) is that

$$\mathbf{1}_F \le 1 - f_\varepsilon \le \mathbf{1}_{F_\varepsilon},$$

where F_ε is as defined there. Thus,

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(F) &\le \limsup_{n\to\infty} \int (1 - f_\varepsilon)\, d\mu_n \\ \text{(by hypothesis of 2)} &= \int (1 - f_\varepsilon)\, d\mu \\ &\le \mu(F_\varepsilon)\,.\end{aligned}$$

Since F is a closed set, (7.2.7) holds, which shows 3. □

Proof of 3 ⇒4. Follows trivially from the fact that the complement of an open set is closed. □

Proof of 4 ⇒1. Assume 4. Let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded and continuous. Thus there exist $a \in \mathbb{R}$ and $b > 0$ such that $0 < b(f - a) < 1$. Since $\mu, \mu_1, \mu_2, \ldots$ are probability measures, for the sake of showing

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu, \tag{7.2.8}$$

it can be assumed without loss of generality that $0 < f < 1$.

Fix $0 < \varepsilon < 1$. Since $\{x \in \mathbb{R} : \mu\left(f^{-1}\{x\}\right) > 0\}$ is a countable set, there exist $\varepsilon/2 < a_1 < \varepsilon$ such that $\mu\left(f^{-1}\{a_1\}\right) = 0$. Proceeding inductively, $0 = a_0 < a_1 < \ldots < a_k = 1$ can be chosen such that

$$\max_{1\le i\le k} (a_i - a_{i-1}) \le \varepsilon \text{ and } \mu\left(f^{-1}\{a_0, \ldots, a_k\}\right) = 0\,.$$

Thus, letting

$$g = \sum_{i=1}^{k} a_i f^{-1}(a_{i-1}, a_i],$$

ensures $g - \varepsilon \le f \le g$. Hence

$$\begin{aligned}
\liminf_{n\to\infty} \int f\, d\mu_n &\ge \liminf_{n\to\infty} \int (g-\varepsilon)\, d\mu_n \\
&= -\varepsilon + \liminf_{n\to\infty} \sum_{i=1}^{k} a_i \mu_n \left(f^{-1}(a_{i-1}, a_i]\right) \\
(a_i \ge 0) &\ge -\varepsilon + \liminf_{n\to\infty} \sum_{i=1}^{k} a_i \mu_n \left(f^{-1}(a_{i-1}, a_i)\right) \\
(a_i \ge 0) &\ge -\varepsilon + \sum_{i=1}^{k} a_i \liminf_{n\to\infty} \mu_n \left(f^{-1}(a_{i-1}, a_i)\right) \\
&\ge -\varepsilon + \sum_{i=1}^{k} a_i \mu \left(f^{-1}(a_{i-1}, a_i)\right),
\end{aligned}$$

the last line following from the hypothesis of 4 and that $f^{-1}(a_{i-1}, a_i)$ is an open set, which is a consequence of continuity of f. The choices of $a_0, \ldots, a_k$ imply that

$$\sum_{i=1}^{k} a_i \mu \left(f^{-1}(a_{i-1}, a_i)\right) = \sum_{i=1}^{k} a_i \mu \left(f^{-1}(a_{i-1}, a_i]\right) = \int g\, d\mu \ge \int f\, d\mu,$$

thus showing

$$\liminf_{n\to\infty} \int f\, d\mu_n \ge \int f\, d\mu - \varepsilon\,.$$

Since ε is arbitrary, the desired lower bound on $\liminf$ is obtained. Replacing f by $-f$ yields the reverse inequality, which establishes 1. ☐

This completes the proof. ☐

Exercise 7.2.3. *Suppose that $X_1, \ldots, X_\infty$ are $\mathbb{R}^d$-valued random variables and $X_n = (X_{n1}, \ldots, X_{nd})$ for $n = 1, \ldots, \infty$.*

1. *If*

$$X_{ni} \xrightarrow{P} X_{\infty i},\ n \to \infty,\ i = 1, \ldots, d,$$

show that $X_n \Rightarrow X_\infty$.

2. *Show that the converse holds when X_∞ is a degenerate random variable, that is, there exists $x \in \mathbb{R}^d$ such that $X_\infty = x$ a.s.*

8 Characteristic function and other transforms

The concept of Fourier transform in analysis is essentially called characteristic function in probability theory.

8.1 Fourier transform

For studying Fourier transforms, the theory of integration for functions taking values in $\mathbb{C}$, the complex plane, has to be developed. This follows the predictable path of defining it for the real and imaginary parts separately. The real and imaginary parts of a complex number will be denoted by $\Re$ and $\Im$, respectively. That is, for $z = a + ib \in \mathbb{C}$, where $i = \sqrt{-1}$ and $a, b \in \mathbb{R}$,

$$\Re(z) = a \quad \text{and} \quad \Im(z) = b\,.$$

Definition. *Given a measure space $(\Omega, \mathcal{A}, \mu)$, a function $f : \Omega \to \mathbb{C}$ is* integrable *if both $\Re(f)$ and $\Im(f)$ belong to $L^1(\Omega, \mathcal{A}, \mu)$, and in that case, the* integral *of f with respect to μ is defined by*

$$\int_\Omega f\,d\mu = \int_\Omega \Re(f)\,d\mu + i\int_\Omega \Im(f)\,d\mu\,.$$

Exercise 8.1.1. *Given a measure space $(\Omega, \mathcal{A}, \mu)$ and a function $f : \Omega \to \mathbb{C}$, show that f is integrable if and only if $\Re(f)$ and $\Im(f)$ are Borel functions from Ω to $\mathbb{R}$ and*

$$\int |f|\,d\mu < \infty\,.$$

The following result is essentially the version of Theorem 3.2.2 for $\mathbb{C}$-valued functions.

Theorem 8.1.1. *For a measure space $(\Omega, \mathcal{A}, \mu)$, the following hold for integrable functions $f, g : \Omega \to \mathbb{C}$:*

1. *if $f = g$ a.e., then*
$$\int f\,d\mu = \int g\,d\mu,$$
2. *for $\alpha \in \mathbb{C}$,*
$$\int \alpha f\,d\mu = \alpha \int f\,d\mu,$$
3. *$f + g$ is integrable and*
$$\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu,$$
4. *and*
$$\left|\int f\,d\mu\right| \le \int |f|\,d\mu\,.$$

Proof. The proof of 1, 2 and 3 is exactly the same as that of 2, 3 and 5 of Theorem 3.2.2, respectively. For the proof of 4, define

$$\alpha = \int f \, d\mu \, .$$

If $\alpha = 0$, then 4 is automatic as the left hand side of the claimed inequality is zero in that case. Assuming $\alpha \in \mathbb{C} \setminus \{0\}$ without loss of generality, 2 implies

$$\begin{aligned} \int \alpha^{-1} f \, d\mu &= \alpha^{-1} \int f \, d\mu \\ &= 1; \end{aligned}$$

the equality in the second line follows from the definition of α. Thus

$$1 = \int \Re(\alpha^{-1} f) \, d\mu \le \int |\alpha^{-1} f| \, d\mu = |\alpha|^{-1} \int |f| \, d\mu,$$

from which the proof of 4 follows by multiplying throughout by $|\alpha|$. □

The following versions of Fubini and DCT can easily be proven and hence their proofs are left as exercises.

Theorem 8.1.2 (Fubini for complex-valued functions)**.** *Suppose* $(\Omega_i, \mathcal{A}_i, \mu_i)$ *is a* σ*-finite measure space for* $i = 1, 2$. *Then for an integrable function* $f : \Omega_1 \times \Omega \to \mathbb{C}$, *it holds that*

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, \mu_2(d\omega_2) \, \mu_1(d\omega_1) = \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1) \, \mu_2(d\omega_2) \, .$$

Proof. Exercise. □

The above theorem can be generalized to any finite product, which is also left as an exercise.

Theorem 8.1.3 (DCT for complex-valued functions)**.** *If* $f_1, \ldots, f_\infty, g$ *are integrable functions from a measure space* $(\Omega, \mathcal{A}, \mu)$ *to* $\mathbb{C}$ *such that* $f_n \to f_\infty$ *as* $n \to \infty$ *and* $|f_n| \le |g|$ *for all* n, *then*

$$\lim_{n \to \infty} \int f_n \, d\mu = \int f_\infty \, d\mu \, .$$

Proof. Exercise. □

Fourier transforms can now be defined.

Definition. *Let $L^1(\mathbb{R}^d \to \mathbb{C})$ be the collection of functions from $\mathbb{R}^d$ to $\mathbb{C}$ which are integrable with respect to the Lebesgue measure on $\mathbb{R}^d$. For $f \in L^1(\mathbb{R}^d \to \mathbb{C})$, its Fourier transform $\widehat{f}$ is a function from $\mathbb{R}^d$ to $\mathbb{C}$ defined by*

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{i\langle \xi, x\rangle}\, dx,\ \xi \in \mathbb{R}^d,$$

where $i = \sqrt{-1}$ and $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^d$.

Example 8.1.1. *Suppose f is the Gaussian density on $\mathbb{R}$, that is,*

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2},\ x \in \mathbb{R}\,.$$

Though we wish to calculate the Fourier transform of f, we start with observing that for $\xi \in \mathbb{R}$,

$$\begin{aligned} \int_{\mathbb{R}} e^{\xi x} f(x)\, dx &= (2\pi)^{-1/2} \int_{\mathbb{R}} e^{\xi x - x^2/2}\, dx \\ \left(\xi x - \frac{x^2}{2} = \frac{\xi^2}{2} - \frac{(x-\xi)^2}{2}\right) &= (2\pi)^{-1/2} e^{\xi^2/2} \int_{\mathbb{R}} e^{-(x-\xi)^2/2}\, dx \\ &= e^{\xi^2/2}\,. \end{aligned}$$

For those with knowledge of complex analysis, $h : \mathbb{C} \to \mathbb{C}$ defined by

$$h(z) = \int_{\mathbb{R}} e^{zx} f(x)\, dx,\ z \in \mathbb{C},$$

which is defined because $\int_{\mathbb{R}} e^{\xi x} f(x)\, dx < \infty$ for all $\xi > 0$, is an entire function. Since

$$h(z) = e^{z^2/2} \text{ for all } z \in \mathbb{R},$$

and the right hand side is also an entire function of $z \in \mathbb{C}$, the result from complex analysis, that any two entire functions which agree on the real line agree on $\mathbb{C}$, implies

$$h(z) = e^{z^2/2} \text{ for all } z \in \mathbb{C}\,.$$

Putting $z = i\xi$ for $\xi \in \mathbb{R}$, it follows that

$$\widehat{f}(\xi) = e^{-\xi^2/2},\ \xi \in \mathbb{R}\,.$$

Those who are not familiar with complex analysis may derive the above via the following exercise.

Exercise 8.1.2. *If $f : \mathbb{R} \to [0, \infty)$ is a Borel function such that*

$$\int_{\mathbb{R}} e^{tx} f(x)\, dx < \infty \text{ and } \int_{\mathbb{R}} e^{-tx} f(x)\, dx < \infty \text{ for some } t > 0,$$

then show that

$$\int_{\mathbb{R}} |x|^n f(x)\,dx < \infty \text{ for all } n = 0, 1, 2, \ldots$$

and

$$\widehat{f}(\xi) = \sum_{n=0}^{\infty} \frac{1}{n!}(i\xi)^n \int_{\mathbb{R}} x^n f(x)\,dx,\ -t \le \xi \le t\,.$$

Hint. *Show that* $\int_{\mathbb{R}} e^{|tx|} f(x)\,dx < \infty$ *and use DCT.*

The following is an easy application of Fubini and the above example.

Exercise 8.1.3. *Suppose for a fixed* $\sigma > 0$*,* $f : \mathbb{R}^d \to \mathbb{R}$ *is defined by*

$$f(x) = (2\pi)^{-d/2}\sigma^{-d}e^{-\|x\|^2/2\sigma^2},\ x \in \mathbb{R}^d,$$

$\|\cdot\|$ *being the usual (Euclidean)* L^2 *norm on* $\mathbb{R}^d$*:*

$$\|x\| = \sqrt{\sum_{i=1}^{d} x_i^2},\ x = (x_1, \ldots, x_d) \in \mathbb{R}^d\,.$$

Show that

$$\widehat{f}(\xi) = e^{-\sigma^2\|\xi\|^2/2},\ \xi \in \mathbb{R}^d\,.$$

Theorem 8.1.4. *For* $f \in L^1(\mathbb{R}^d \to \mathbb{C})$*, its Fourier transform* $\widehat{f}$ *is an uniformly continuous function from* $\mathbb{R}^d$ *to* $\mathbb{C}$ *and satisfies*

$$\sup_{\xi\in\mathbb{R}^d} |\widehat{f}(\xi)| \le \int_{\mathbb{R}^d} |f(x)|\,dx\,.$$

Proof. For $\xi \in \mathbb{R}^d$,

$$\begin{aligned}\left|\widehat{f}(\xi)\right| &= \left|\int_{\mathbb{R}^d} f(x)e^{i\langle\xi,x\rangle}\,dx\right| \\ &\le \int_{\mathbb{R}^d} \left|f(x)e^{i\langle\xi,x\rangle}\right|\,dx \\ &= \int_{\mathbb{R}^d} |f(x)|\,dx,\end{aligned}$$

and for $\xi, \xi' \in \mathbb{R}^d$, a similar calculation shows

$$\left|\widehat{f}(\xi) - \widehat{f}(\xi')\right| \le \int_{\mathbb{R}^d} |f(x)|\left|e^{i\langle\xi-\xi',x\rangle} - 1\right|\,dx\,.$$

It follows from DCT that

$$\lim_{u\to 0}\int_{\mathbb{R}^d} |f(x)|\left|e^{i\langle u,x\rangle} - 1\right|\,dx = 0,$$

thus showing $\widehat{f}$ is uniformly continuous. □

Theorem 8.1.5 (Fourier inversion theorem). *If $f \in L^1(\mathbb{R}^d \to \mathbb{C})$ is such that $\widehat{f} \in L^1(\mathbb{R}^d \to \mathbb{C})$, then*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle}\, d\xi \text{ for a.e. } x \in \mathbb{R}^d . \tag{8.1.1}$$

Before proceeding to the proof, let us record an implication of (8.1.1).

Corollary 8.1.1. *If f satisfies the hypothesis of Theorem 8.1.5 and is continuous, then*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle}\, d\xi \text{ for all } x \in \mathbb{R}^d .$$

Remark 8.1.1. *In the harmonic analysis literature, the Fourier transform of $f \in L^1(\mathbb{R}^d \to \mathbb{C})$ is often defined as*

$$\widetilde{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{2\pi i \langle \xi, x\rangle}\, dx,\ \xi \in \mathbb{R}^d .$$

With the above definition, for an f satisfying the hypothesis of Theorem 8.1.5, (8.1.1) *takes the form*

$$f(-x) = \widetilde{\left(\widetilde{f}\right)}(x) \text{ for a.e. } x \in \mathbb{R}^d . \tag{8.1.2}$$

Indeed, (8.1.1) *can be rewritten as that for a.e. $x \in \mathbb{R}^d$,*

$$\begin{aligned} f(x) &= (2\pi)^{-d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y) e^{i\langle \xi, y-x\rangle}\, dy \right) d\xi \\ \left(\text{putting } \xi = 2\pi\theta,\ d\xi = (2\pi)^d \theta\right) &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y) e^{i\langle 2\pi\theta, y-x\rangle}\, dy \right) d\theta \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y) e^{2\pi i\langle \theta, y-x\rangle}\, dy \right) d\theta \\ &= \int_{\mathbb{R}^d} \widetilde{f}(\theta) e^{2\pi i \langle \theta, -x\rangle}\, d\theta \\ &= \widetilde{\left(\widetilde{f}\right)}(-x), \end{aligned}$$

which is equivalent to (8.1.2).

The proof of Theorem 8.1.5 uses the following lemma which is important in analysis for several other reasons.

Lemma 8.1.1. *If $L^1(\mathbb{R}^d \to \mathbb{C})$ is equipped with the usual L^1 metric, then*

$$C_c(\mathbb{R}^d, \mathbb{C}) = \left\{ f : \mathbb{R}^d \to \mathbb{C} \,:\, f \text{ is continuous and compactly supported} \right\}$$

is dense in $L^1(\mathbb{R}^d \to \mathbb{C})$.

Proof. Let $g : \mathbb{R}^d \to [0, \infty)$ be integrable and $\varepsilon > 0$. An argument similar to that leading to (3.5.4) shows that

$$g = \sum_{n=1}^{\infty} c_n \mathbf{1}_{A_n} \tag{8.1.3}$$

for some $c_1, c_2, \ldots \in (0, \infty)$ and $A_1, A_2, \ldots \in \mathcal{B}(\mathbb{R}^d)$. Integrability of g with respect to λ, the Lebesgue measure on $\mathbb{R}^d$, and that $c_n > 0$ imply

$$\lambda(A_n) < \infty \text{ for all } n \geq 1\,.$$

Exercise 3.5.4, which is valid on $\mathbb{R}^d$ as well, shows that for all n, there exists K_n compact and U_n open with $K_n \subset A_n \subset U_n$ and

$$\lambda(U_n \setminus K_n) < \frac{1}{c_n} \varepsilon 2^{-n-1}\,.$$

Since U_n is open, U_n^c is a closed set which is disjoint from K_n. Thus

$$\delta_n = \inf\{\|x - y\| : x \in K_n,\, y \in U_n^c\} > 0\,.$$

Let N be such that

$$\sum_{n=N+1}^{\infty} c_n \lambda(A_n) < \frac{\varepsilon}{2},$$

which exists because of (8.1.3) and the assumption that g is integrable. Clearly,

$$\mathbf{1}_{K_n}(x) \leq 1 - (1 \wedge \delta_n^{-1} d(x, K_n)) \leq \mathbf{1}_{U_n}(x),\, n = 1, \ldots, N\,.$$

Define

$$f(x) = \sum_{n=1}^{N} c_n \left(1 - (1 \wedge \delta_n^{-1} d(x, K_n))\right),\, x \in \mathbb{R}^d\,.$$

Then f is a compactly supported continuous function with

$$\begin{aligned}
\|f - g\|_1 &\leq \sum_{n=N+1}^{\infty} c_n \lambda(A_n) + \sum_{n=1}^{N} c_n \lambda(U_n \setminus K_n) \\
&< \frac{\varepsilon}{2} + \sum_{n=1}^{\infty} \varepsilon 2^{-n-1} \\
&= \varepsilon\,.
\end{aligned}$$

For $g \in L^1(\mathbb{R}^d \to \mathbb{C})$, applying the same argument to the positive and negative parts of each of $\Re(g)$ and $\Im(g)$, the proof follows. $\square$

Proof of Theorem 8.1.5. Let f be as given, that is, $f, \widehat{f} \in L^1(\mathbb{R}^d \to \mathbb{C})$. For fixed $\varepsilon > 0$ and $x \in \mathbb{R}^d$, write

$$\begin{aligned}\int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y) e^{i\langle \xi, y\rangle}\, dy \right) e^{-i\langle \xi, x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi \\ &= \int_{\mathbb{R}^d} f(y) \left(\int_{\mathbb{R}^d} e^{i\langle \xi, y-x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi \right) dy \end{aligned} \tag{8.1.4}$$

the second line following from Fubini and the observation

$$\left(\int_{\mathbb{R}^d} e^{-\varepsilon^2\|\xi\|^2/2}\, d\xi \right) \int_{\mathbb{R}^d} |f(y)|\, dy < \infty\,.$$

Putting $\sigma = \varepsilon^{-1}$, Exc 8.1.3 shows that

$$\int_{\mathbb{R}^d} e^{i\langle \xi, y-x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi = (2\pi)^{d/2} \varepsilon^{-d} e^{-\|y-x\|^2/2\varepsilon^2}\,.$$

Thus (8.1.4) shows

$$\begin{aligned}\int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi &= (2\pi)^{d/2}\varepsilon^{-d} \int_{\mathbb{R}^d} f(y) e^{-\|y-x\|^2/2\varepsilon^2}\, dy \\ \big(\text{substitute } z = \varepsilon^{-1}(y-x)\big) &= (2\pi)^{d/2} \int_{\mathbb{R}^d} f(x+\varepsilon z) e^{-\|z\|^2/2}\, dz\,.\end{aligned}$$

Rewrite the above as

$$\int_{\mathbb{R}^d} f(x+\varepsilon z)(2\pi)^{-d/2} e^{-\|z\|^2/2}\, dz = (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi\,. \tag{8.1.5}$$

DCT and integrability of $\widehat{f}$ imply

$$\lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle} e^{-\varepsilon^2 \|\xi\|^2/2}\, d\xi = \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\langle \xi, x\rangle}\, d\xi, \ \text{ for all } x \in \mathbb{R}^d\,.$$

Thus the claim would follow once it is shown that there exist integers $1 \le k_1 < k_2 < \dots$ such that

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f(x + z/k_n)(2\pi)^{-d/2} e^{-\|z\|^2/2}\, dz = f(x) \text{ for a.e. } x \in \mathbb{R}^d\,.$$

The above would follow from Theorem 3.4.5 and the observation

$$\int_{\mathbb{R}^d} e^{-\|z\|^2/2}\, dz = (2\pi)^{d/2}$$

if it can be shown that

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x+z/k) - f(x)|\, e^{-\|z\|^2/2}\, dz\, dx = 0\,. \tag{8.1.6}$$

Proceeding towards (8.1.6), using Tonelli write

$$
\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x+z/k)-f(x)|\, e^{-\|z\|^2/2}\, dz\, dx \\
&= \int_{\mathbb{R}^d} e^{-\|z\|^2/2} \int_{\mathbb{R}^d} |f(x+z/k)-f(x)|\; dx\, dz \\
&= \int_{\mathbb{R}^d} e^{-\|z\|^2/2} g(z/k)\, dz,
\end{aligned}
$$

where

$$
g(y) = \int_{\mathbb{R}^d} |f(x+y)-f(x)|\; dx,\; y \in \mathbb{R}^d\,.
$$

Since $0 \le g \le 2\|f\|_1$, DCT would imply (8.1.6) once it is shown that

$$
\lim_{y\to 0} g(y) = 0\,. \tag{8.1.7}
$$

Fix $\delta > 0$ and use Lemma 8.1.1 to get $h \in C_c(\mathbb{R}^d, \mathbb{C})$ satisfying

$$
\|f-h\|_1 \le \frac{\delta}{3}\,.
$$

Since h is a continuous function supported on a compact set, say $K \subset \mathbb{R}^d$, it is uniformly continuous. Thus there exists $\eta > 0$ such that

$$
|h(y)-h(z)| \le \frac{1}{6\lambda(K)}\delta \text{ whenever } \|y-z\| \le \eta,\; y, z \in \mathbb{R}^d\,.
$$

For $y \in \mathbb{R}^d$ with $\|y\| \le \eta$,

$$
\begin{aligned}
g(y) &\le 2\|f-h\|_1 + \int_{\mathbb{R}^d} |h(x+y)-h(x)|\; dx \\
&\le \frac{2}{3}\delta + \int_{\mathbb{R}^d} |h(x+y)-h(x)|\; dx \\
&= \frac{2}{3}\delta + \int_{K\cup(K-y)} |h(x+y)-h(x)|\; dx \\
\text{(Choice of } \eta) &\le \frac{2}{3}\delta + \frac{\delta}{6\lambda(K)}\lambda(K\cup(K-y)) \\
&\le \frac{2}{3}\delta + \frac{\delta}{6\lambda(K)} 2\lambda(K) \\
&= \delta\,.
\end{aligned}
$$

This establishes (8.1.7), from which the proof follows. □

Remark 8.1.2. *Equation* (8.1.5) *is often called an "approximate identity". Indeed,* (8.1.1) *essentially follows by letting* $\varepsilon \downarrow 0$ *in* (8.1.5).

8.2 Characteristic function

Definition. *For a probability measure μ on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, its* characteristic function *ϕ_μ is defined by*

$$\phi_\mu(\xi) = \int_{\mathbb{R}^d} e^{i\langle \xi, x\rangle}\, dx,\ \xi \in \mathbb{R}^d\,.$$

For an $\mathbb{R}^d$-valued random variable X defined on a probability space $(\Omega, \mathcal{A}, P)$, its characteristic function *is defined as the characteristic function of the measure $P \circ X^{-1}$ on $\mathbb{R}^d$.*

Henceforth, $\mathcal{B}(\mathbb{R}^d)$ is the σ-field associated with $\mathbb{R}^d$. The following is a trivial exercise.

Exercise 8.2.1. *If μ is a probability measure on $\mathbb{R}^d$ which has a density f with respect to Lebesgue measure, then show that*

$$\phi_\mu = \widehat{f},$$

where ϕ_μ is the characteristic function of μ and $\widehat{f}$ is the Fourier transform of f.

The proof of the following theorem is very similar to that of Theorem 8.1.4, and is hence left as an exercise.

Theorem 8.2.1. *If ϕ is the characteristic function of a probability measure on $\mathbb{R}^d$, then ϕ is uniformly continuous and satisfies*

$$\sup_{t\in\mathbb{R}} |\phi(t)| \le 1 = \phi(0)\,.$$

Proof. Exercise. $\square$

Definition. *For probability measures μ_1, μ_2 on $\mathbb{R}^d$, the* convolution *of μ_1 and μ_2 is the probability measure $\mu_1 * \mu_2$ defined by*

$$\mu_1 * \mu_2(A) = \int_{\mathbb{R}^d} \mu_1(A - x)\, \mu_2(dx),\ A \in \mathcal{B}(\mathbb{R}^d),$$

where $A - x = \{a - x : a \in A\}$.

The following result provides an useful interpretation of convolution, namely it is the push-forward of the product measure under sum.

Lemma 8.2.1. *Suppose μ_1 and μ_2 are probability measures on $\mathbb{R}^d$. Define $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ by $T(x, y) = x + y$. Then*

$$\mu_1 * \mu_2 = (\mu_1 \otimes \mu_2) \circ T^{-1}\,,$$

that is, $\mu_1 * \mu_2$ is the push-forward measure of the product measure $\mu_1 \otimes \mu_2$ on $\big(\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d})\big)$ under T.

Proof. For $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
(\mu_1 \otimes \mu_2)\,(T^{-1}B) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}(x+y \in B)\mu_1(dx)\mu_2(dy) \\
&= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} \mathbf{1}(x \in B-y)\mu_1(dx)\right)\mu_2(dy) \\
&= \int_{\mathbb{R}^d} \mu_1(B-y)\,\mu_2(dy) \\
&= \mu_1 * \mu_2(B)\,,
\end{aligned}$$

and hence the proof. □

An immediate corollary of Lemma 8.2.1 is the following.

Corollary 8.2.1. *The convolution operator is commutative, that is, $\mu_1 * \mu_2 = \mu_2 * \mu_1$ for all probability measures μ_1, μ_2 on $\mathbb{R}^d$.*

Exercise 8.2.2. *If X_1 and X_2 are independent $\mathbb{R}^d$-valued random variables defined on $(\Omega, \mathcal{A}, P)$ such that $P \circ X_i^{-1} = \mu_i$ for $i = 1, 2$, show that*

$$P \circ (X_1 + X_2)^{-1} = \mu_1 * \mu_2\,.$$

The following result connects characteristic functions with convolutions.

Lemma 8.2.2. *For probability measures μ_1 and μ_2 on $\mathbb{R}^d$,*

$$\phi_{\mu_1 * \mu_2}(\xi) = \phi_{\mu_1}(\xi)\phi_{\mu_2}(\xi) \text{ for all } \xi \in \mathbb{R}^d,$$

where $\phi_{\mu_1}, \phi_{\mu_2}, \phi_{\mu_1 * \mu_2}$ are the characteristic functions of $\mu_1, \mu_2, \mu_1 * \mu_2$, respectively.

Proof. For $\xi \in \mathbb{R}^d$,

$$\begin{aligned}
\phi_{\mu_1 * \mu_2}(\xi) &= \int_{\mathbb{R}^d} e^{i\langle \xi, x\rangle}\,\mu_1 * \mu_2(dx) \\
\text{(Lemma 8.2.1)} &= \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{i\langle \xi, y+z\rangle}\mu_1(dy) \otimes \mu_2(dz) \\
\text{(Fubini)} &= \phi_{\mu_1}(\xi)\phi_{\mu_2}(\xi)\,.
\end{aligned}$$

as claimed. □

Exercise 8.2.3. *If μ_1 and μ_2 are probability measures on $\mathbb{R}^d$, show that $\mu_1 * \mu_2$ has a density with respect to Lebesgue if so does either μ_1 or μ_2.*

The next result is essentially a consequence of Theorem 8.1.5.

Theorem 8.2.2. *If the characteristic function ϕ_μ of a probability measure μ on $\mathbb{R}^d$ is integrable, that is,*

$$\int_{\mathbb{R}^d} |\phi_\mu(\xi)|\, d\xi < \infty\,, \tag{8.2.1}$$

then

$$\mu(B) = \int_B f(x)\, dx \text{ for all } B \in \mathcal{B}(\mathbb{R}^d),$$

where

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \xi, x\rangle} \phi_\mu(\xi)\, d\xi,\ x \in \mathbb{R}^d\,. \tag{8.2.2}$$

Proof. Let X be an $\mathbb{R}^D$-valued random variable, defined on some probability space $(\Omega, \mathcal{A}, P)$, having distribution μ, that is, $P \circ X^{-1} = \mu$. What has to be shown is that f, defined by (8.2.2), is the density of X.

Suppose $Z = (Z_1, \ldots, Z_d)$ is a random variable defined on the same probability space (for which Ω may have to be expanded), independent of X, such that $Z_1, \ldots, Z_d$ are independent from standard normal, that is, they have density

$$\phi(z) = (2\pi)^{-1/2} e^{-z^2/2},\ z \in \mathbb{R}\,.$$

Define

$$Y_\varepsilon = X + \varepsilon Z,\ \varepsilon > 0\,.$$

Exc 8.2.3 shows that for al $\varepsilon > 0$, Y_ε has a density, say f_ε. Lemma 8.2.2 shows that for fixed $\varepsilon > 0$,

$$\widehat{f_\varepsilon}(\xi) = \widehat{\phi_\mu}(\xi) e^{-\varepsilon^2 \|\xi\|^2/2},\ \xi \in \mathbb{R}^d\,.$$

Theorem 8.1.5 shows that for a.e. x,

$$\begin{aligned} f_\varepsilon(x) &= (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{\phi_\mu}(\xi) e^{-i\langle \xi, x\rangle - \varepsilon^2\|\xi\|^2/2}\, d\xi \\ &\to (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{\phi_\mu}(\xi) e^{-i\langle \xi, x\rangle}\, d\xi \end{aligned} \tag{8.2.3}$$

as $\varepsilon \downarrow 0$ by DCT. That is,

$$\lim_{n\to\infty} f_{1/n} = f \text{ a.e.} \tag{8.2.4}$$

Hence $f \geq 0$ a.e. Faotou's lemma implies

$$\int_{\mathbb{R}^d} f(x)\, dx \leq \liminf_{n\to\infty} f_{1/n}(x)\, dx = 1\,.$$

Thus f is non-negative and integrable.

Rewrite (8.2.2) as

$$\widehat{\phi_\mu}(x) = (2\pi)^{-d} f(-x),\ x \in \mathbb{R}^d,$$

and use Corollary 8.1.1 and Theorem 8.2.1 along with integrability of f to argue that for all $\xi \in \mathbb{R}^d$,

$$\phi_\mu(\xi) = (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{\phi_\mu}(x) e^{-i\langle \xi, x\rangle}\, dx = \int_{\mathbb{R}^d} f(-x) e^{-i\langle \xi, x\rangle}\, dx\,.$$

Putting $\xi = 0$ and using Theorem 8.2.1 again, it follows that

$$1 = \int_{\mathbb{R}^d} f(-x)\, dx = \int_{\mathbb{R}^d} f(x)\, dx\,.$$

This, in conjunction with (8.2.4) and Theorem 3.4.2, which is Scheffé's lemma, implies

$$f_{1/n} \to f \text{ in } L^1,\ n \to \infty\,. \tag{8.2.5}$$

Recall from the definition of Y_ε that as $\varepsilon \downarrow 0$, $Y_\varepsilon \to X$. Exc 7.2.3 implies $Y_\varepsilon \Rightarrow X$, as $\varepsilon \downarrow 0$. However, (8.2.5) means

$$Y_{1/n} \Rightarrow V, \tag{8.2.6}$$

where V is a random variable with density f. Theorem 7.2.1 shows $V \overset{d}{=} X$, that is, f is the density of X, which was required to be shown. Hence the proof follows. $\square$

The next result connects characteristic functions with weak convergence.

Theorem 8.2.3 (Lévy's continuity theorem)**.** *Suppose $\mu_1, \mu_2, \ldots, \mu_\infty$ are probability measures on $\mathbb{R}^d$ with respective characteristic functions $\phi_1, \phi_2 \ldots \ldots, \phi_\infty$. Then $\mu_n \Rightarrow \mu_\infty$ if and only if*

$$\lim_{n\to\infty} \phi_n(\xi) = \phi_\infty(\xi) \text{ for all } \xi \in \mathbb{R}^d\,. \tag{8.2.7}$$

Proof. The "only if" part follows trivially from the definition of weak convergence. For the "if" part, assume (8.2.7). Suppose $X_1, \ldots, X_\infty$ are random variables with respective distributions $\mu_1, \ldots, \mu_\infty$. By an appeal to Theorem 7.2.3, it suffices to show that for all bounded uniformly continuous function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathrm{E}\left(f(X_n)\right) = \mathrm{E}\left(f(X_\infty)\right)\,. \tag{8.2.8}$$

Let $Y = (Y^{(1)}, \ldots, Y^{(d)})$ be independent of $(X_1, \ldots, X_\infty)$ where $Y^{(1)}, \ldots \ldots, Y^{(d)}$ are i.i.d. real-valued random variables with density

$$h(y) = (1 - |y|)\,\mathbf{1}(|y| \le 1)\,.$$

It can be checked that

$$\widehat{h}(t) = \begin{cases} 2t^{-2}(1 - \cos t), & t \ne 0, \\ 1, & t = 0\,. \end{cases}$$

Define

$$Y_{n\varepsilon} = X_n + \varepsilon Y,\, \varepsilon > 0,\, n = 1, 2, \ldots, \infty\,.$$

Thus the characteristic function $\phi_{Y_{n\varepsilon}}$ of $Y_{n\varepsilon}$ is

$$\phi_{Y_{n\varepsilon}}(\xi) = \phi_n(\xi) \prod_{j=1}^{d} \widehat{h}(\varepsilon \xi_j),\, \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d\,.$$

Integrability of $\widehat{h}$ on $\mathbb{R}$ and (8.2.7) imply by DCT that for all fixed $\varepsilon > 0$,

$$\phi_{Y_{n\varepsilon}} \to \phi_{Y_{\infty\varepsilon}} \text{ in } L^1,\, n \to \infty\,.$$

Proceeding with the help of Theorem 8.1.5 and Scheffé's lemma, as in (8.2.3)-(8.2.6), it can be argued from the above that

$$Y_{n\varepsilon} \Rightarrow Y_{\infty\varepsilon},\, n \to \infty,\, \varepsilon > 0\,. \tag{8.2.9}$$

To prove (8.2.8), fix a bounded uniformly continuous $f : \mathbb{R}^d \to \mathbb{R}$ and $\delta > 0$. Choose $\varepsilon > 0$ such that

$$|f(x) - f(y)| \le \delta \text{ whenever } \|x - y\|_\infty \le \varepsilon,\, x, y \in \mathbb{R}^d,$$

where $\|\cdot\|_\infty$ is the L^∞ norm on $\mathbb{R}^d$ as in (4.3.2). Since $\|Y\|_\infty \le 1$, it is immediate that

$$\|Y_{n\varepsilon} - X_n\|_\infty \le \varepsilon,\, n = 1, \ldots, \infty\,.$$

Thus $|f(X_n) - f(Y_{n\varepsilon})| \le \delta$ for all $n = 1, \ldots, \infty$. This in conjunction with (8.2.9) establishes

$$\limsup_{n\to\infty} |\mathrm{E}(f(X_n)) - \mathrm{E}(f(X_\infty))| \le 2\delta\,.$$

Since δ is arbitrary, (8.2.8) follows, which completes the proof. $\square$

The following is a consequence of the above result and Theorem 7.2.1.

Corollary 8.2.2. *The characteristic function of a probability measure is unique, that is, if μ_1 and μ_2 are probability measures on $\mathbb{R}^d$ with respective characteristic functions ϕ_1 and ϕ_2, then*

$$\mu_1 = \mu_2 \iff \phi_1(\xi) = \phi_2(\xi) \textit{ for all } \xi \in \mathbb{R}^d\,.$$

The following result is usually the most used one for proving weak convergence in $\mathbb{R}^d$.

Theorem 8.2.4 (Cramér-Wold device)**.** *For $\mathbb{R}^d$ valued random variables $X, X_1, X_2, \ldots,$*

$$X_n \Rightarrow X$$

if and only if

$$\langle \xi, X_n \rangle \Rightarrow \langle \xi, X \rangle \textit{ for all } \xi \in \mathbb{R}^d\,.$$

Proof. The "only if" part follows from Theorem 7.2.2 and the fact that $x \mapsto \langle \xi, x \rangle$ is a continuous map from $\mathbb{R}^d$ to $\mathbb{R}$ for all $\xi \in \mathbb{R}^d$. The "if" part follows from Theorem 8.2.3. $\square$

Corollary 8.2.3. *For $\mathbb{R}^d$-valued random variables X and Y,*

$$X \stackrel{d}{=} Y \text{ if and only if } \langle \xi, X \rangle \stackrel{d}{=} \langle \xi, Y \rangle \text{ for all } \xi \in \mathbb{R}^d .$$

8.3 Moments generating function

Definition. *For a probability measure μ on $\mathbb{R}^d$, its* moments generating function *(MGF) is a function $\psi : \mathbb{R}^d \to (0, \infty]$ defined by*

$$\psi(\xi) = \int_{\mathbb{R}^d} e^{\langle \xi, x \rangle} \mu(dx),\ \xi \in \mathbb{R}^d .$$

For an $\mathbb{R}^d$-valued random variable X, its moments generating function *is the MGF of $P \circ X^{-1}$.*

Theorem 8.3.1. *Suppose X and Y are $\mathbb{R}^d$-valued random variables with respective MGFs ψ_X and ψ_Y. If there exists $\varepsilon > 0$ such that*

$$\psi_X(\lambda) = \psi_Y(\lambda) < \infty \text{ whenever } \lambda \in \mathbb{R}^d, \|\lambda\| \le \varepsilon, \tag{8.3.1}$$

where $\| \cdot \|$ is any norm on $\mathbb{R}^d$, then $X \stackrel{d}{=} Y$.

Proof. Let us first prove this for the case $d = 1$. That is, assume X and Y are real-valued random variables with

$$\mathrm{E}\left(e^{\theta X}\right) = \mathrm{E}\left(e^{\theta Y}\right) < \infty \text{ for all } -\varepsilon \le \theta \le \varepsilon . \tag{8.3.2}$$

It has to be shown that $X \stackrel{d}{=} Y$.

For those familiar with complex analysis, the functions

$$f_X(z) = \mathrm{E}\left(e^{zX}\right) \text{ and } f_Y(z) = \mathrm{E}\left(e^{zY}\right) \text{ for all } z \in \mathbb{C},\ |\Re(z)| < \varepsilon$$

are holomorphic functions on the strip $\{z \in \mathbb{C} : -\varepsilon < \Re(z) < \varepsilon\}$; (8.3.2) simply means f_X and f_Y agree on $(-\varepsilon, \varepsilon)$. Recall the complex analysis result that if two holomorphic functions g and h on a connected open subset U of $\mathbb{C}$ agree on some set which has a limit point in U, then $g = h$ on U. The said result implies

$$f_X(z) = f_Y(z) \text{ for all } z \in \mathbb{C},\ |\Re(z)| < \varepsilon .$$

Putting $z = i\xi$, where $i = \sqrt{-1}$, for $\xi \in \mathbb{R}$, shows that the characteristic functions of X and Y are the same. Corollary 8.2.2 with $d = 1$ implies $X \stackrel{d}{=} Y$.

For a proof without using complex analysis, let ϕ_X and ϕ_Y be the characteristic functions of X and Y, respectively. An implication of (8.3.2) is

$$\mathrm{E}\left(e^{\varepsilon|X|}\right) \le \mathrm{E}\left(e^{\varepsilon X} + e^{-\varepsilon X}\right) < \infty\,.$$

Hence

$$\infty > \mathrm{E}\left(e^{\varepsilon|X|}\right) = \mathrm{E}\left(\sum_{n=0}^{\infty} \frac{1}{n!}\varepsilon^n |X|^n\right) = \sum_{n=0}^{\infty} \frac{1}{n!}\varepsilon^n \mathrm{E}\left(|X|^n\right)\,. \tag{8.3.3}$$

Therefore, all moments of X are finite and

$$\sum_{n=0}^{\infty} \frac{1}{n!}\left|\mathrm{E}(X^n)\right| \varepsilon^n < \infty\,.$$

Using Fubini, it can thus be shown that

$$\phi_X(t) = \sum_{n=0}^{\infty} \frac{1}{n!}\mathrm{E}(X^n)(it)^n,\ 0 \le t \le \varepsilon\,. \tag{8.3.4}$$

A similar argument works for Y and (8.3.2) shows

$$\mathrm{E}(X^n) = \mathrm{E}(Y^n),\ n \ge 1\,.$$

The above combined with (8.3.4) and its equivalent for Y implies

$$\phi_Y(t) = \phi_X(t),\ 0 \le t \le \varepsilon\,. \tag{8.3.5}$$

For $-\varepsilon \le t \le \varepsilon$, (8.3.3) implies

$$\sum_{n=0}^{\infty} \frac{1}{n!}|it|^n \left|\mathrm{E}\left(X^n e^{i\varepsilon X}\right)\right| \le \sum_{n=0}^{\infty} \frac{1}{n!}\varepsilon^n \mathrm{E}\left(|X|^n\right) < \infty\,.$$

Apply Fubini to get

$$\phi_X(\varepsilon+t) = \mathrm{E}\left(e^{i\varepsilon X}\sum_{n=0}^{\infty} \frac{1}{n!}(itX)^n\right) = \sum_{n=0}^{\infty} \frac{1}{n!}(it)^n\, \mathrm{E}\left(e^{i\varepsilon X}X^n\right),\ -\varepsilon \le t \le \varepsilon\,. \tag{8.3.6}$$

A similar argument works for Y which agrees with the above power series for $-\varepsilon \le t \le 0$ by (8.3.5). Thus

$$\mathrm{E}\left(e^{i\varepsilon X}X^n\right) = \mathrm{E}\left(e^{i\varepsilon Y}Y^n\right),\ n \ge 1\,.$$

For $0 < t \le \varepsilon$, (8.3.6), its equivalent for Y and the above show that

$$\phi_X(\varepsilon+t) = \phi_Y(\varepsilon+t)\,.$$

In other words,
$$\phi_X(s) = \phi_Y(s),\, \varepsilon < s \le 2\varepsilon\,.$$
Like (8.3.5) implies the above, proceeding inductively it can be shown that for all $n \ge 1$,
$$\phi_X(t) = \phi_Y(t),\, (n-1)\varepsilon < t \le n\varepsilon\,.$$
Since this holds for all $n \ge 1$, ϕ_X and ϕ_Y agree on $[0,\infty)$. A similar argument works for the negative half-line, and shows ϕ_X is identical to ϕ_Y, a consequence of which is $X \stackrel{d}{=} Y$. This proves the claim for $d = 1$.

For a general d, assume (8.3.1). Corollary 8.2.3 shows that if suffices to prove
$$\langle \xi, X\rangle \stackrel{d}{=} \langle \xi, Y\rangle,\, \xi \in \mathbb{R}^d\,. \tag{8.3.7}$$
Fix $\xi \in \mathbb{R}^d$. As the above trivially holds for $\xi = 0$, assume without loss of generality that $\|\xi\| > 0$. Letting $\delta = \varepsilon/\|\xi\|$, (8.3.1) with $\lambda = \theta\xi$, for $-\delta \le \theta \le \delta$ shows
$$\psi_X(\theta\xi) = \psi_Y(\theta\xi) < \infty,$$
which is the same as
$$\mathrm{E}\left(e^{\theta\langle\xi,X\rangle}\right) = \mathrm{E}\left(e^{\theta\langle\xi,Y\rangle}\right) < \infty,\, -\delta \le \theta \le \delta\,.$$
By the claim for $d = 1$, which has already been proven, (8.3.7) follows and so does the proof. □

9 The central limit theorems

In order to state the central limit theorems in higher dimensions, the multivariate normal distribution has to be first studied. This is done in the first subsection.

9.1 Multivariate normal distribution

Exercise 9.1.1. *Let $X_1, \ldots, X_n$ be i.i.d. from standard normal, and*
$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}.$$
Fix $\mu \in \mathbb{R}^n$ and let Σ be a $n \times n$ real symmetric positive definite (p.d.) matrix, that is, $\Sigma^T = \Sigma$ and $x^T\Sigma x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$. Let $\Sigma^{1/2}$ be the p.d. square root of Σ, that is, $\Sigma^{1/2}$ is the unique p.d. matrix whose square is Σ. Define
$$Y = \mu + \Sigma^{1/2}X\,. \tag{9.1.1}$$

Show that the density of $Y = (Y_1, \ldots, Y_n)$ *is*

$$g(y) = \frac{1}{(2\pi)^{n/2}\det(\Sigma^{1/2})}\exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right), y \in \mathbb{R}^n .$$

Soln.: Follows from Theorem 4.3.2.

The density obtained in the above exercise is the density of the so-called multivariate normal distribution which is formally defined below. The next several results are devoted to understanding the properties of this distribution. Observing that $\det(\Sigma^{1/2}) = \sqrt{\det(\Sigma)}$, the following definition makes perfect sense.

Definition. *If* $X = (X_1, \ldots, X_n)$ *has density*

$$f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), x \in \mathbb{R}^n , \quad (9.1.2)$$

for some $\mu \in \mathbb{R}^n$ *and* $n \times n$ *p.d. matrix* Σ*, then* X *follows* n*-dimensional* multivariate normal distribution *with parameters* μ *and* Σ*, which is written as*

$$X \sim N_n(\mu, \Sigma) .$$

The interpretation of μ and Σ in the distribution $N_n(\mu, \Sigma)$ will be clear after a couple of results. The following theorem is essentially the converse of Exc 9.1.1.

Theorem 9.1.1. *If* $X \sim N_n(\mu, \Sigma)$ *and* $(Y_1, \ldots, Y_n) = Y = \Sigma^{-1/2}(X - \mu)$*, then* $Y_1, \ldots, Y_n$ *are i.i.d. from standard normal.*

Proof. The density of X is

$$f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), x \in \mathbb{R}^n$$

Writing

$$Y = \Sigma^{-1/2}X - \Sigma^{-1/2}\mu ,$$

Theorem 4.3.2 implies that the density of Y is

$$\begin{aligned}
g(y) &= \frac{1}{\det(\Sigma^{-1/2})} f\left(\Sigma^{1/2}\left(y + \Sigma^{-1/2}\mu\right)\right) \\
&= \frac{1}{\det(\Sigma^{-1/2})} f\left(\Sigma^{1/2}y + \mu\right) \\
&= (2\pi)^{-n/2}\exp\left(-\frac{1}{2}\left(\Sigma^{1/2}y\right)^T\Sigma^{-1}\left(\Sigma^{1/2}y\right)\right) \\
&= (2\pi)^{-n/2}\exp\left(-\frac{1}{2}y^Ty\right) ,
\end{aligned}$$

the last line follows from the fact that $\Sigma^{1/2}$ is symmetric and

$$\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2} = I\,.$$

Exercise 6.1.4 shows that $Y_1, \ldots, Y_n$ are i.i.d. from standard normal, which completes the proof. □

The next theorem shows that if $X \sim N_n(\mu, \Sigma)$, then μ and Σ are the "mean vector" and the "covariance matrix" of X, respectively.

Theorem 9.1.2. *If* $X \sim N_n(\mu, \Sigma)$ *where*

$$\mu = (\mu_1, \ldots, \mu_n) \text{ and } \Sigma = ((\sigma_{ij}))_{1 \le i,j \le n}\,,$$

then

$$\begin{aligned} \mathrm{E}(X_i) &= \mu_i\,, i = 1, \ldots, n\,, \\ \mathrm{Cov}(X_i, X_j) &= \sigma_{ij}\,, 1 \le i, j \le n\,. \end{aligned}$$

In particular, $\mathrm{Var}(X_i) = \sigma_{ii}$ *for* $i = 1, \ldots, n$.

Proof. Let $(Y_1, \ldots, Y_n) = Y = \Sigma^{-1/2}(X - \mu)$; $Y_1, \ldots, Y_n$ are i.i.d. from standard normal by Theorem 9.1.1. Rewrite the above as

$$X = \mu + \Sigma^{1/2} Y\,,$$

or

$$X_i = \mu_i + \sum_{j=1}^n \theta_{ij} Y_j\,, i = 1, \ldots, n\,,$$

where $\Sigma^{1/2} = ((\theta_{ij}))_{1 \le i,j \le n}$. Since $Y_1, \ldots, Y_n$ are zero mean random variables, it immediately follows $\mathrm{E}(X_i) = \mu_i$ for $i = 1, \ldots, n$. Exercise 3.6.5 shows that for fixed $1 \le i, j \le n$,

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \mathrm{Cov}\left(\sum_{k=1}^n \theta_{ik} Y_k, \sum_{l=1}^n \theta_{jl} Y_l \right) \\ &= \sum_{k=1}^n \sum_{l=1}^n \theta_{ik}\theta_{jl} \mathrm{Cov}\,(Y_k, Y_l) \\ &= \sum_{k=1}^n \theta_{ik}\theta_{jk}\,, \end{aligned}$$

the last line following from the fact that $Y_1, \ldots, Y_n$ are independent and each has variance one. Recalling that θ_{ik} is the (i,k)-th entry of $\Sigma^{1/2}$ which is a symmetric matrix, write

$$\begin{aligned}\sum_{k=1}^{n} \theta_{ik}\theta_{jk} &= \sum_{k=1}^{n} \theta_{ik}\theta_{kj} \\ &= (i,j)\text{-th entry of } \Sigma^{1/2}\Sigma^{1/2} \\ &= (i,j)\text{-th entry of } \Sigma \\ &= \sigma_{ij}\,.\end{aligned}$$

It thus follows that

$$\mathrm{Cov}(X_i, X_j) = \sigma_{ij}\,, 1 \le i,j \le n\,.$$

Taking $i = j$ implies $\mathrm{Var}(X_i) = \sigma_{ii}$ and completes the proof. □

Example 9.1.1. *Let $Z_1, \ldots, Z_d$ be i.i.d. from standard normal and $Z = (Z_1, \ldots, Z_d)$. Fix a $d \times d$ symmetric non-negative definite (n.n.d.) matrix Σ and $\mu \in \mathbb{R}^d$ and define*

$$X = \mu + \Sigma^{1/2} Z\,, \tag{9.1.3}$$

where elements of $\mathbb{R}^d$ are to be interpreted as column vectors by convention. Let us calculate the characteristic function of X. Fix $\lambda \in \mathbb{R}^d$ and write

$$\lambda^T X = \lambda^T \mu + \theta^T Z\,,$$

where

$$\theta = \Sigma^{1/2}\lambda\,.$$

Recall that $\theta^T Z$ follows $N(0, \|\theta\|^2)$, where $\|\cdot\|$ is the L^2-norm, if $\|\theta\| > 0$; $\theta^T Z$ is degenerate at zero otherwise. Assuming for a moment that $t = \|\theta\| > 0$,

$$\begin{aligned}\mathrm{E}\left(e^{i\theta^T Z}\right) &= \mathrm{E}\left(\exp\left(it\frac{\theta^T Z}{\|\theta\|}\right)\right) \\ \left(\textit{because } \frac{\theta^T Z}{\|\theta\|} \sim N(0,1)\right) &= e^{-t^2/2} \\ &= e^{-\|\theta\|^2/2} \\ &= \exp\left(-\frac{1}{2}\theta^T\theta\right) \\ &= \exp\left(-\frac{1}{2}\lambda^T\Sigma\lambda\right)\,.\end{aligned}$$

If $\|\theta\| = 0$, that is, θ is the zero vector, then also

$$\mathrm{E}\left(e^{i\theta^T Z}\right) = \exp\left(-\frac{1}{2}\lambda^T\Sigma\lambda\right)\,,$$

because both sides equal 1 *in this case. Thus,*

$$\mathrm{E}\left(e^{i\lambda^T X}\right) = e^{i\lambda^T \mu}\mathrm{E}\left(e^{i\theta^T Z}\right) = \exp\left(i\lambda^T \mu - \frac{1}{2}\lambda^T \Sigma \lambda\right).$$

In other words, the characteristic function ϕ_X *of* X *is*

$$\phi_X(\lambda) = \exp\left(i\lambda^T \mu - \frac{1}{2}\lambda^T \Sigma \lambda\right), \lambda \in \mathbb{R}^d.$$

Definition. *An* $\mathbb{R}^d$*-valued random variable* X *follows* $N_d(\mu, \Sigma)$ *for* $\mu \in \mathbb{R}^d$ *and a* $d \times d$ *symmetric n.n.d. matrix* Σ*, if the characteristic function of* X *is*

$$\phi_X(\lambda) = \exp\left(i\lambda^T \mu - \frac{1}{2}\lambda^T \Sigma \lambda\right), \lambda \in \mathbb{R}^d.$$

The above definition is consistent with (9.1.2) in the following sense. If Σ is p.d. and $X \sim N_d(\mu, \Sigma)$ according to the above definition, then the density of X is f as in (9.1.2). Indeed, (9.1.3) should be compared with (9.1.1) to see this immediately.

Remark 9.1.1. *The distribution* $N_d(\mu, \Sigma)$ *is called a "singular normal distribution" if* Σ *is n.n.d. but not p.d. It should be noted that a singular normal distribution in one dimension is a degenerate distribution.*

Exercise 9.1.2. *Show that a* $N_d(\mu, \Sigma)$ *distribution has a density if and only if* Σ *is p.d.*

Theorem 9.1.3. *For an* $\mathbb{R}^d$*-valued random variable* X*,* $\mu \in \mathbb{R}^d$ *and a* $d \times d$ *n.n.d. matrix* Σ*,*

$$X \sim N_d(\mu, \Sigma) \iff \langle \lambda, X \rangle \sim N\left(\lambda^T \mu, \lambda^T \Sigma \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d.$$

Proof. For the "$\Rightarrow$ part", assume $X \sim N_d(\mu, \Sigma)$ and fix $\lambda \in \mathbb{R}^d$. Then for $t \in \mathbb{R}$,

$$\begin{aligned}\mathrm{E}\left(e^{it\langle \lambda, X \rangle}\right) &= \mathrm{E}\left(e^{i\langle t\lambda, X \rangle}\right) \\ &= \exp\left(i(t\lambda)^T \mu - \frac{1}{2}(t\lambda)^T \Sigma (t\lambda)\right) \\ &= e^{it\theta - \sigma^2 t^2/2},\end{aligned}$$

where $\theta = \lambda^T \mu$ and $\sigma^2 = \lambda^T \Sigma \lambda$. As the above is true for all $t \in \mathbb{R}$, $\langle \lambda, X \rangle \sim N(\theta, \sigma^2)$. This proves the "$\Rightarrow$ part".

For the reverse implication, assume that

$$\langle \lambda, X \rangle \sim N\left(\lambda^T \mu, \lambda^T \Sigma \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d.$$

Let $Y \sim N_d(\mu, \Sigma)$. The already proven "$\Rightarrow$ part" shows that

$$\langle \lambda, Y \rangle \sim N\left(\lambda^T \mu, \lambda^T \Sigma \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d .$$

Thus $\langle \lambda, X \rangle \stackrel{d}{=} \langle \lambda, Y \rangle$ for all $\lambda \in \mathbb{R}^d$. The Cramér-Wold device shows

$$X \stackrel{d}{=} Y ,$$

from which the "$\Leftarrow$ part" follows. This completes the proof. $\square$

Exercise 9.1.3. *If X is a $\mathbb{R}^d$-valued random vector such that for all $\lambda \in \mathbb{R}^d$, $\lambda^T X$ follows one-dimensional normal, show that $X \sim N_d(\mu, \Sigma)$, where μ and Σ are the mean vector and the variance-covariance matrix of X, respectively.*

Exercise 9.1.4. *If $X_i \sim N_d(\mu_i, \Sigma_i)$ for $i = 1, \ldots, n$ and $X_1, \ldots, X_n$ are independent, show that*

$$\sum_{i=1}^N X_i \sim N_d\left(\sum_{i=1}^n \mu_i , \sum_{i=1}^n \Sigma_i\right) .$$

The above property of the normal distribution is of utmost importance and will be crucially used in the proof of the Lindeberg central limit theorem in the next subsection.

9.2 The central limit theorems in $\mathbb{R}$

The following result, due to Lindeberg, is arguably the most general central limit theorem (CLT) on $\mathbb{R}$ for independent summands.

Theorem 9.2.1 (Lindeberg's CLT on $\mathbb{R}$)**.** *Suppose that for $n \in \mathbb{N}$, $X_{n1}, \ldots$ $\ldots, X_{nn}$ are independent $\mathbb{R}$-valued random variables satisfying the following:*

$$\mathrm{E}(X_{ni}) = 0 , i = 1, \ldots, n,\ n = 1, 2, \ldots ,$$

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2\right) = \sigma^2 < \infty ,$$

and

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)\right) = 0 ,\ \textit{for every } \varepsilon > 0 . \tag{9.2.1}$$

Then, as $n \to \infty$,

$$\sum_{i=1}^n X_{ni} \Rightarrow Z ,$$

where $Z \sim N(0, \sigma^2)$.

The proof uses the following exercises.

Exercise 9.2.1. *If $X_n \sim N(0, \sigma_n^2)$ and $0 \le \sigma_n \to \sigma < \infty$, show that $X_n \Rightarrow X$ where $X \sim N(0, \sigma^2)$.*
Hint. *If $Z \sim N(0,1)$, then $X_n \stackrel{d}{=} \sigma_n Z \to \sigma Z$.*

Exercise 9.2.2. *Suppose $X, X_1, X_2, \ldots$ are random variables such that for all thrice differentiable bounded $f : \mathbb{R} \to \mathbb{R}$ whose first three derivatives are bounded, it holds that*

$$\lim_{n\to\infty} \mathrm{E}\left(f(X_n)\right) = \mathrm{E}\left(f(X)\right) .$$

Show that $X_n \Rightarrow X$.
Hint. *Let*

$$f(x) = \begin{cases} 1, & x \le 0\,, \\ (1-x^4)^4, & 0 < x < 1\,, \\ 0, & x \ge 1\,. \end{cases}$$

Observe that for $w < y$,

$$\mathbf{1}_{(-\infty,w]}(x) \le f\left(\frac{x-w}{y-w}\right) \le \mathbf{1}_{(-\infty,y]}(x) \text{ for all } x \in \mathbb{R}\,.$$

Proof of Theorem 9.2.1. Using Exc 9.2.2, it suffices to show that

$$\lim_{n\to\infty} \mathrm{E}\left(f(S_n)\right) = \mathrm{E}\left(f(Z)\right) , \tag{9.2.2}$$

for all thrice differentiable $f : \mathbb{R} \to \mathbb{R}$ such that f and its first three derivatives are bounded, where

$$S_n = \sum_{i=1}^n X_{ni}\,, n \ge 1.$$

Fix such f.

Let $(Z_1, Z_2, \ldots)$ be a collection of i.i.d. standard normal random variables which is independent of the triangular array $\{X_{ni} : 1 \le i \le n, n \ge 1\}$. Set

$$\sigma_{ni} = \sqrt{\mathrm{E}(X_{ni}^2)}\,, 1 \le i \le n\,, n = 1, 2, \ldots\,,$$

and

$$\sigma_n = \sqrt{\sum_{i=1}^n \sigma_{ni}^2}\,, n \ge 1\,.$$

Since

$$\sum_{i=1}^n \sigma_{ni} Z_i \sim N(0, \sigma_n^2)\,, n = 1, 2, \ldots\,, \tag{9.2.3}$$

and $\sigma_n^2 \to \sigma^2$, Exc 9.2.1 shows

$$\lim_{n\to\infty} \mathrm{E}\left(f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right) = \mathrm{E}\left(f(Z)\right).$$

Thus, (9.2.2) would follow once it is shown that

$$\lim_{n\to\infty} \mathrm{E}\left(f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right) = 0. \tag{9.2.4}$$

Fix $n \in \{1, 2, \ldots\}$ and write

$$f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right) = \sum_{i=1}^n (f(Y_{i-1}) - f(Y_i)),$$

where

$$Y_i = \sum_{j=i+1}^n X_{nj} + \sum_{j=1}^i \sigma_{nj} Z_j, i = 0, 1, \ldots, n,$$

with the usual interpretation of the sum as zero if the lower limit exceeds the upper limit. Thus,

$$\left|\mathrm{E}\left(f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right)\right| \leq \sum_{i=1}^n \left|\mathrm{E}(f(Y_{i-1}) - f(Y_i))\right|. \tag{9.2.5}$$

Fix $i \in \{1, \ldots, n\}$ and write

$$Y_i = W + \sigma_{ni} Z_i,$$

and

$$Y_{i-1} = W + X_{ni},$$

where

$$W = \sum_{j=i+1}^n X_{nj} + \sum_{j=1}^{i-1} \sigma_{nj} Z_j.$$

It is immediate that W, X_{ni}, Z_i are independent. Taylor's theorem implies

$$\begin{aligned} f(Y_{i-1}) &= f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(\xi_1) &\quad (9.2.6)\\ &= f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) + \frac{1}{6} X_{ni}^3 f'''(\xi_2), &\quad (9.2.7) \end{aligned}$$

for some ξ_1 and ξ_2 between W and Y_{i-1}, where f', f'', f''' are the first three derivatives of f, respectively. Let

$$K = \sup_{x\in\mathbb{R}} \left(|f(x)| \vee |f'(x)| \vee |f''(x)| \vee |f'''(x)|\right),$$

which is finite by assumption. A consequence of (9.2.6) is that

$$\begin{aligned}
&\left|f(Y_{i-1}) - \left(f(W) + X_{ni}f'(W) + \frac{1}{2}X_{ni}^2 f''(W)\right)\right| \\
&= \frac{1}{2}X_{ni}^2 \left|f''(\xi_1) - f''(W)\right| \\
&\le \frac{1}{2}X_{ni}^2 \left(|f''(\xi_1)| + |f''(W)|\right) \\
&\le K X_{ni}^2 \,.
\end{aligned}$$

Similarly, (9.2.7) shows

$$\left|f(Y_{i-1}) - \left(f(W) + X_{ni}f'(W) + \frac{1}{2}X_{ni}^2 f''(W)\right)\right| \le \frac{1}{6}K|X_{ni}|^3 \le K|X_{ni}|^3 \,.$$

Thus,

$$\left|f(Y_{i-1}) - \left(f(W) + X_{ni}f'(W) + \frac{1}{2}X_{ni}^2 f''(W)\right)\right| \le K(X_{ni}^2 \wedge |X_{ni}|^3) \,.$$

Therefore,

$$\begin{aligned}
K\mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) &\ge \mathrm{E}\left|f(Y_{i-1}) - \left(f(W) + X_{ni}f'(W) + \frac{1}{2}X_{ni}^2 f''(W)\right)\right| \\
&\ge \left|\mathrm{E}(f(Y_{i-1})) - \mathrm{E}\left(f(W) + X_{ni}f'(W) + \frac{1}{2}X_{ni}^2 f''(W)\right)\right| \\
&= \left|\mathrm{E}(f(Y_{i-1})) - \mathrm{E}(f(W)) - \frac{1}{2}\sigma_{ni}^2 \mathrm{E}(f''(W))\right| \,,
\end{aligned}$$

the last line following from the independence of W and X_{ni} and that the mean and variance of X_{ni} are zero and σ_{ni}^2, respectively. A similar calculation shows

$$\left|\mathrm{E}(f(Y_i)) - \mathrm{E}(f(W)) - \frac{1}{2}\sigma_{ni}^2 \mathrm{E}(f''(W))\right| \le K\mathrm{E}\left(|\sigma_{ni}Z_i|^3\right) = C\sigma_{ni}^3 \,,$$

where $C = K\mathrm{E}(|Z_1|^3)$. Combine the two inequalities obtained to get

$$\left|\mathrm{E}\left(f(Y_{i-1}) - f(Y_i)\right)\right| \le K\mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) + C\sigma_{ni}^3 \,.$$

Summing the above inequality over $i = 1, \ldots, n$ and using (9.2.5), we get

$$\left|\mathrm{E}\left(f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni}Z_i\right)\right)\right| \le C\sum_{i=1}^n \sigma_{ni}^3 + K\sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) \,.$$

Thus, (9.2.4) would follow, which would complete the proof, once the following are shown:

$$\lim_{n\to\infty}\sum_{i=1}^n \sigma_{ni}^3 = 0\,, \tag{9.2.8}$$

$$\text{and } \lim_{n\to\infty}\sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) = 0\,. \tag{9.2.9}$$

For (9.2.8), write

$$\begin{aligned}\sum_{i=1}^n \sigma_{ni}^3 &\le \sigma_n^2 \max_{1\le i\le n} \sigma_{ni} \\ &= \sigma_n^2 \sqrt{\max_{1\le i\le n} \sigma_{ni}^2}\,.\end{aligned}$$

Since $\sigma_n^2 \to \sigma^2 < \infty$, (9.2.8) would follow if it can be shown that

$$\lim_{n\to\infty}\max_{1\le i\le n} \sigma_{ni}^2 = 0\,.$$

Fix $\varepsilon > 0$ and write

$$\sigma_{ni}^2 = \mathrm{E}(X_{ni}^2\mathbf{1}(|X_{ni}| \le \varepsilon)) + \mathrm{E}(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon)) \le \varepsilon^2 + \mathrm{E}(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon))\,.$$

Hence

$$\begin{aligned}\max_{1\le i\le n} \sigma_{ni}^2 &\le \varepsilon^2 + \max_{1\le i\le n} \mathrm{E}(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon)) \\ &\le \varepsilon^2 + \sum_{i=1}^n \mathrm{E}(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon))\,.\end{aligned}$$

Invoke (9.2.1) to argue

$$\limsup_{n\to\infty}\max_{1\le i\le n} \sigma_{ni}^2 \le \varepsilon^2\,.$$

Since ε is arbitrary,

$$\lim_{n\to\infty}\max_{1\le i\le n} \sigma_{ni}^2 = 0\,,$$

which shows (9.2.8).

Finally, for (9.2.9), fix $\varepsilon > 0$ and write

$$\begin{aligned}\sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) &\le \sum_{i=1}^n \mathrm{E}\left(|X_{ni}|^3\mathbf{1}(|X_{ni}| \le \varepsilon\right) + \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon)\right) \\ &\le \varepsilon \sum_{i=1}^n \mathrm{E}(X_{ni}^2) + \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon)\right) \\ &= \varepsilon\sigma_n^2 + \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2\mathbf{1}(|X_{ni}| > \varepsilon)\right)\,.\end{aligned}$$

Let $n \to \infty$ and use (9.2.1) to get

$$\limsup_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) \le \varepsilon\sigma^2 .$$

Since ε is arbitrary, (9.2.9) follows. This in conjunction with (9.2.8) shows (9.2.4), which completes the proof. □

Remark 9.2.1. *The above proof is transparent in that it displays the property of normal that has been used. Indeed,* (9.2.3) *does use the fact that the sum of independent normal random variables also follows normal.*

The assumption (9.2.1) is called Lindeberg's condition. The family $\{X_{ni} : 1 \le i \le n,\, n = 1, 2, \ldots\}$ is called a triangular array, which is why, Theorem 9.2.1 is also known as CLT for triangular arrays. The following special case of Theorem 9.2.1 is known as the central limit theorem for i.i.d. random variables with finite variance.

Theorem 9.2.2. *Suppose $X_1, X_2, \ldots$ are i.i.d. zero-mean random variables with finite variance σ^2. Then*

$$n^{-1/2} \sum_{i=1}^{n} X_i \Rightarrow Z,\ n \to \infty,$$

where Z follows $N(0, \sigma^2)$.

Proof. Define

$$X_{ni} = \frac{1}{\sqrt{n}} X_i\,, 1 \le i \le n,\ n \ge 1 .$$

Then

$$\sum_{i=1}^{n} \mathrm{E}(X_{ni}^2) = \sigma^2 \text{ for all } n = 1, 2, \ldots$$

and DCT implies that for $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathrm{E}\left(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)\right) = \mathrm{E}\left(X_1^2 \mathbf{1}(|X_1| > \sqrt{n}\varepsilon)\right) \to 0, n \to \infty .$$

Theorem 9.2.1 completes the proof. □

The following corollary of Theorem 9.2.1 gives an easy way of checking the Lindeberg condition, which is based on the inequality

$$\mathrm{E}\left(|X|^{2+\delta}\right) \ge \varepsilon^\delta \mathrm{E}\left(X^2 \mathbf{1}(|X| > \varepsilon)\right),\ \delta, \varepsilon > 0 .$$

Corollary 9.2.1 (Lyapunov CLT). *Suppose that for $n \in \mathbb{N}$, $X_{n1}, \ldots, X_{nn}$ are independent $\mathbb{R}$-valued random variables satisfying the following:*

$$\mathrm{E}(X_{ni}) = 0\,, i = 1, \ldots, n,\ n = 1, 2, \ldots\,,$$

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(X_{ni}^2\right) = \sigma^2 < \infty\,,$$

and for some $\delta > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(|X_{ni}|^{2+\delta}\right) = 0\,.$$

Then, as $n \to \infty$,

$$\sum_{i=1}^{n} X_{ni} \Rightarrow Z\,,$$

where $Z \sim N(0, \sigma^2)$.

9.3 The central limit theorems in $\mathbb{R}^d$

The following is the generalization of Theorem 9.2.1 to $\mathbb{R}^d$.

Theorem 9.3.1 (Lindeberg CLT on $\mathbb{R}^d$). *Suppose for all fixed $n = 1, 2, \ldots$, $X_{n1}, \ldots, X_{nn}$ are independent $\mathbb{R}^d$-valued random variables with zero mean and covariance matrices $\Sigma_{n1}, \ldots, \Sigma_{nn}$, respectively. Assume that*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \Sigma_{ni} = \Sigma$$

for some $d \times d$ matrix Σ, where the above limit is taken entry by entry. Assume furthermore that for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(\|X_{ni}\|^2 \mathbf{1}(\|X_{ni}\| > \varepsilon)\right) = 0,$$

where $\|\cdot\|$ is the Euclidean L^2-norm. Then Σ is n.n.d. and

$$\sum_{i=1}^{k_n} X_{ni} \Rightarrow Z,\ n \to \infty,$$

where $Z \sim N_d(0, \Sigma)$.

Proof. Since Σ is the entrywise limit of n.n.d. matrices, it is n.n.d. Thus there exists an $\mathbb{R}^d$-valued Z which follows $N_d(0, \Sigma)$. Theorem 8.2.4, which is the Cramér-Wold device, would complete the proof once it is shown that for all $\xi \in \mathbb{R}^d$,

$$\left\langle \xi, \sum_{i=1}^n X_{ni} \right\rangle \Rightarrow \langle \xi, Z \rangle,\ n \to \infty\,. \tag{9.3.1}$$

Fix $\xi \in \mathbb{R}^d \setminus \{0\}$, which is interpreted as a column vector, and write

$$\left\langle \xi, \sum_{i=1}^n X_{ni} \right\rangle = \sum_{i=1}^n Y_{ni} \text{ where } Y_{ni} = \langle \xi, X_{ni} \rangle,\ 1 \le i \le n,\ n \in \mathbb{N}\,.$$

It is immediate that $Y_{n1}, \ldots, Y_{nn}$ are independent,

$$\mathrm{E}(Y_{ni}) = 0 \text{ and } \mathrm{Var}(Y_{ni}) = \xi' \Sigma_{ni} \xi,\ i = 1, \ldots, n,\ n \ge 1\,.$$

Thus

$$\sum_{i=1}^n \mathrm{Var}(Y_{ni}) \to \xi' \Sigma \xi\,.$$

Finally, for $\varepsilon > 0$, the Cauchy-Schwarz inequality in $\mathbb{R}^d$ implies for $\varepsilon > 0$,

$$\mathrm{E}\left(Y_{ni}^2 \mathbf{1}(|Y_{ni}| > \varepsilon)\right) \le \|\xi\|^2 \mathrm{E}\left(\|X_{ni}\|^2 \mathbf{1}\left(\|X_{ni}\| > \|\xi\|^{-1}\varepsilon\right)\right),$$

showing that the triangular array $\{Y_{ni} : i = 1, \ldots, n,\ n = 1, 2, \ldots\}$ satisfies Lindeberg's condition. Theorem 9.2.1 shows

$$\sum_{i=1}^n Y_{ni} \Rightarrow Y,$$

where $Y \sim N(0, \xi' \Sigma \xi)$. Theorem 9.1.3 shows $Y \stackrel{d}{=} \langle \xi, Z \rangle$ and hence the above is the same as (9.3.1), from which the proof follows. $\square$

The following is the generalization of Theorem 9.2.2 and follows from Theorem 9.3.1 in the same way as the former does from Theorem 9.2.1.

Corollary 9.3.1. *If $X_1, X_2, \ldots$ are i.i.d. $\mathbb{R}^d$-valued random variables with mean zero and covariance matrix Σ, where entries of Σ are all finite,*

$$n^{-1/2} \sum_{i=1}^n X_i \Rightarrow Z,$$

where $Z \sim N_d(0, \Sigma)$.

10 Conditional expectation

10.1 Conditional expectation as an L^2 projection

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space and $X \in L^2(\Omega)$ and $\mathcal{F} \subset \mathcal{A}$ is a σ-field. The goal of this subsection is to show that X has a "projection" on $L^2(\Omega, \mathcal{F}, P)$ and the same will be defined as the "conditional expectation of X given $\mathcal{F}$".

Theorem 10.1.1. *There exists $Z \in L^2(\Omega, \mathcal{F}, P)$ such that*

$$\|X - Z\|_2 = \inf\{\|X - Y\|_2 : Y \in L^2(\Omega, \mathcal{F}, P)\} < \infty\,.$$

Proof. Let

$$v = \inf\{\|X - Y\|_2 : Y \in L^2(\Omega, \mathcal{F}, P)\}\,.$$

Taking $Y = 0$ in the right hand side, it is immediate that $v \le \|X\|_2 < \infty$. There exist $Z_1, Z_2, \ldots \in L^2(\Omega, \mathcal{F}, P)$ such that

$$\|Z_n - X\|_2 \to v\,.$$

We shall show that $\{Z_n\}$ is a Cauchy sequence in $L^2(\Omega)$.

Fix $\varepsilon > 0$. Fix N such that

$$\|Z_n - X\|_2 \le \sqrt{v^2 + \frac{\varepsilon^2}{4}},\ n \ge N\,.$$

For fixed $m, n \ge N$, denote $U = Z_m - X$ and $V = Z_n - X$ and write

$$\begin{aligned}
\|Z_m - Z_n\|_2^2 &= \mathrm{E}\left((U - V)^2\right) \\
&= 2\mathrm{E}(U^2 + V^2) - \mathrm{E}\left[(U + V)^2\right] \\
&= 2\mathrm{E}(U^2 + V^2) - 4\mathrm{E}\left[\left(\frac{Z_m + Z_n}{2} - X\right)^2\right] \\
\left(\text{as } \frac{Z_m + Z_n}{2} \in L^2(\Omega, \mathcal{F}, P)\right) &\le 2\mathrm{E}(U^2 + V^2) - 4v^2 \\
&= 2\left(\|Z_m - X\|_2^2 + \|Z_n - X\|_2^2 - 2v^2\right) \\
(\text{choice of } N) &\le 2\left(v^2 + \frac{\varepsilon^2}{4} + v^2 + \frac{\varepsilon^2}{4} - 2v^2\right) \\
&= \varepsilon^2\,.
\end{aligned}$$

In other words, $\|Z_m - Z_n\|_2 \le \varepsilon$ for all $m, n \ge N$, that is, $\{Z_n\}$ is Cauchy in $L^2(\Omega, \mathcal{F}, P)$.

Theorem 3.4.6 implies $\{Z_n\}$ is a convergent sequence, that is, there exists $Z \in L^2(\Omega, \mathcal{F}, P)$ such that $\|Z_n - Z\|_2 \to 0$. Therefore

$$\|Z - X\|_2 = \lim_{n\to\infty} \|Z_n - X\|_2 = v,$$

thus completing the proof. □

Definition. *The* conditional expectation *of X given $\mathcal{F}$, denoted by* $\mathrm{E}(X|\mathcal{F})$, *is $Z \in L^2(\Omega, \mathcal{F}, P)$ satisfying*

$$\|X - Z\|_2 = \inf\{\|X - Y\|_2 : Y \in L^2(\Omega, \mathcal{F}, P)\}. \tag{10.1.1}$$

The following result can be thought of as the Pythagoras theorem in $L^2(\Omega, \mathcal{A}, P)$.

Theorem 10.1.2. *For all $Y \in L^2(\Omega, \mathcal{F}, P)$,*

$$\mathrm{E}\left[(X - \mathrm{E}(X|\mathcal{F}))\, Y\right] = 0\,.$$

The above holds, in particular, with Y replaced by $Z = \mathrm{E}(X|\mathcal{F})$ and hence

$$\|X\|_2^2 = \|X - Z\|_2^2 + \|Z\|_2^2\,.$$

Proof. Let $Z = \mathrm{E}(X|\mathcal{F})$ and $Y \in L^2(\Omega, \mathcal{F}, P)$. Define

$$f(\alpha) = \|X - (Z + \alpha Y)\|_2^2,\ \alpha \in \mathbb{R}\,.$$

The definition of Z implies

$$f(\alpha) \geq f(0)\,.$$

Rewriting

$$f(\alpha) = \mathrm{E}\left[((X - Z) - \alpha Y)^2\right] = \mathrm{E}\left[(X - Z)^2\right] - 2\alpha \mathrm{E}\left[(X - Z)Y\right] + \alpha^2 \mathrm{E}(Y^2),$$

elementary calculus shows that

$$0 = \left.\frac{df(\alpha)}{d\alpha}\right|_{\alpha=0} = -2\mathrm{E}\left[(X - Z)Y\right]\,.$$

This proves the first claim. The second claim follows by putting $Y = Z$ and the last one follows trivially from that. $\square$

Theorem 10.1.3. *For all $Z \in L^2(\Omega, \mathcal{F}, P)$, $Z = \mathrm{E}(X|\mathcal{F})$ if and only if*

$$\int_A X\, dP = \int_A Z\, dP \text{ for all } A \in \mathcal{F}\,.$$

Proof. The "only if" part follows from Theorem 10.1.2 by putting $Y = \mathbf{1}_A$ for any $A \in \mathcal{F}$. Conversely, suppose that

$$\int_A X\, dP = \int_A Z\, dP \text{ for all } A \in \mathcal{F}\,.$$

Let $Z' = \mathrm{E}(X|\mathcal{F})$. The already proven "only if" part shows that for $A \in \mathcal{F}$,

$$\int_A Z'\, dP = \int_A X\, dP = \int_A Z\, dP\,.$$

In other words, $Z - Z'$ is an $\mathcal{F}$-measurable integrable ($L^2(\Omega) \subset L^1(\Omega)$) random variable satisfying

$$\int_A (Z - Z')\, dP = 0 \text{ for all } A \in \mathcal{F}\,.$$

Thus $Z = Z'$ a.s. Hence $Z = \mathrm{E}(X|\mathcal{F})$, proving the "if" part. $\square$

10.2 Conditional expectation for L^1 functions

Theorem 10.1.3 will help us define conditional expectation for L^1 functions, that is, integrable random variables. As before $(\Omega, \mathcal{A}, P)$ is a probability space and $\mathcal{F} \subset \mathcal{A}$ is a σ-field. Now $X \in L^1(\Omega, \mathcal{A}, P)$.

Theorem 10.2.1. *There exists $Z \in L^1(\Omega, \mathcal{F}, P)$ such that*

$$\int_A Z\, dP = \int_A X\, dP \text{ for all } A \in \mathcal{F}\,.$$

Proof. First assume $X \geq 0$. Define $X_n = X \wedge n$ and

$$Z_n = \mathrm{E}(X_n|\mathcal{F}),\ n \geq 1\,.$$

Theorem 10.1.3 shows for all $A \in \mathcal{F}$,

$$\int_A Z_n\, dP = \int_A X_n\, dP \leq \int_A X_{n+1}\, dP = \int_A Z_{n+1}\, dP\,.$$

Thus $Z_{n+1} - Z_n$ is $\mathcal{F}$-measurable and

$$\int_A (Z_{n+1} - Z_n)\, dP \geq 0 \text{ for all } A \in \mathcal{F}\,.$$

Hence $Z_n \leq Z_{n+1}$ a.s. A similar argument shows $Z_n \geq 0$ a.s.

Letting $Z = \sup_n Z_n$, it thus follows that $0 \leq Z_n \uparrow Z$ a.s. MCT shows that for all $A \in \mathcal{F}$,

$$\int_A Z\, dP = \lim_{n\to\infty} \int_A Z_n\, dP = \lim_{n\to\infty} \int_A X_n\, dP = \int_A X\, dP,$$

showing the desired equality. For an integrable X which is not necessarily non-negative, splitting $X = X^+ - X^-$, the proof follows. □

Definition. *For $X \in L^1(\Omega, \mathcal{F}, P)$, its* conditional expectation *is defined as* $\mathrm{E}(X|\mathcal{F}) = Z$ *if $Z \in L^1(\Omega, \mathcal{F}, P)$ and satisfies*

$$\int_A Z\, dP = \int_A X\, dP \text{ for all } A \in \mathcal{F},$$

which exists by Theorem 10.2.1. This definition is consistent with (10.1.1) *for square-integrable X by Theorem 10.1.3.*

Exercise 10.2.1. *Show that for integrable X_1, X_2,*

$$\mathrm{E}(X_1 + X_2|\mathcal{F}) = \mathrm{E}(X_1|\mathcal{F}) + \mathrm{E}(X_2|\mathcal{F})\,.$$

Theorem 10.2.2. *If X, XY are integrable and Y is $\mathcal{F}$-measurable, then*

$$\mathrm{E}\,(XY|\mathcal{F}) = Y\,\mathrm{E}(X|\mathcal{F})\,.$$

Proof. Assume $X \geq 0$ and let $Z = \mathrm{E}(X\mathcal{F})$. If $Y = \mathbf{1}_F$ for some $F \in \mathcal{F}$, then for any $A \in \mathcal{F}$,

$$\begin{aligned} \int_A XY\,dP &= \int_{A\cap F} X\,dP \\ \text{(definition of } Z,\ A\cap F \in \mathcal{F}) &= \int_{A\cap F} Z\,dP \\ &= \int_A YZ\,dP\,. \end{aligned}$$

Routine arguments via simple functions and MCT show that

$$\int_A XY\,dP = \int_A YZ\,dP$$

for all $\mathcal{F}$-measurable $Y \geq 0$. Taking $A = \Omega$ and using the fact that XY is integrable, it follows that so is YZ. In other words,

$$\mathrm{E}(XY|\mathcal{F}) = YZ = Y\mathrm{E}(X|\mathcal{F})\,.$$

For integrable X and any $\mathcal{F}$-measurable Y such that XY is integrable, write $X = X^+ - X^-$ and likewise for Y. Since

$$0 \leq X^{\pm}Y^{\pm} \leq |XY|,$$

from what we showed it follows that for $i, j \in \{+,-\}$,

$$\mathrm{E}\left(X^iY^j|\mathcal{F}\right) = Y^j\mathrm{E}\left(X^i|\mathcal{F}\right)\,.$$

This completes the proof. □

Exercise 10.2.2. *If X is integrable and $\mathcal{F}$ is independent of $\sigma(X)$, show that*

$$\mathrm{E}(X|\mathcal{F}) = \mathrm{E}(X)\,.$$

Theorem 10.2.3 (Tower property)**.** *If X is integrable and $\mathcal{F} \subset \mathcal{G} \subset \mathcal{A}$ are σ-fields, then*

$$\mathrm{E}\left(\mathrm{E}(X|\mathcal{G})\middle|\mathcal{F}\right) = \mathrm{E}(X|\mathcal{F})\,.$$

Proof. Let $Y = \mathrm{E}(X|\mathcal{F})$ and $Z = \mathrm{E}(X|\mathcal{G})$. For $A \in \mathcal{F}$, the definition of Y implies

$$\begin{aligned} \int_A Y\,dP &= \int_A X\,dP \\ (A \in \mathcal{F} \subset \mathcal{G} \text{ and definition of } Z) &= \int_A Z\,dP\,. \end{aligned}$$

Since this holds for all $A \in \mathcal{F}$ and Y is $\mathcal{F}$-measurable, we get

$$\mathrm{E}(Z|\mathcal{F}) = Y,$$

which is the claimed equality. □

Taking $\mathcal{F} = \{\emptyset, \Omega\}$ in the above theorem, we get the following result.

Corollary 10.2.1. *If X is integrable and $\mathcal{G} \subset \mathcal{A}$ is a σ-field, then*

$$\mathrm{E}\big(\mathrm{E}(X|\mathcal{G})\big) = \mathrm{E}(X)\,.$$

10.3 Martingales

As before, $(\Omega, \mathcal{A}, P)$ is a probability space.

Definition. *If $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{A}$ are σ-fields and $X_1, X_2, \ldots \in L^1(\Omega)$ are such that*

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n \textit{ for all } n = 1, 2, \ldots,$$

then $(X_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale*. The collection $\{\mathcal{F}_n\}_{n\geq 1}$ is a* filtration*.*

Exercise 10.3.1. *If $(X_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale, show that*

$$\mathrm{E}\left(X_n|\mathcal{F}_m\right) = X_m \textit{ for all } 1 \leq m < n, \tag{10.3.1}$$

and

$$\mathrm{E}(X_n) = \mathrm{E}(X_1) \textit{ for all } n \geq 1\,. \tag{10.3.2}$$

Example 10.3.1. *Suppose $(X_n : n = 0, 1, 2, \ldots)$ is a simple symmetric random walk (SSRW) on $\mathbb{Z}$ starting from the origin and $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Show that $(X_n, \mathcal{F}_n)_{n\geq 0}$ and $(X_n^2 - n, \mathcal{F}_n)$ are martingales.*

Definition. *An $\mathbb{N}$-valued random variable τ is a stopping time with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 1}$ if $[\tau = n] \in \mathcal{F}_n$ for all $n \geq 1$.*

The following result essentially shows that (10.3.2) holds if n is replaced by a bounded stopping time.

Theorem 10.3.1 (Optional stopping theorem)**.** *If $(X_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale and τ is a bounded stopping time with respect to the filtration $\{\mathcal{F}_n\}_{n\geq 1}$, then*

$$\mathrm{E}(X_\tau) = \mathrm{E}(X_1)\,.$$

Proof. Since τ is bounded, suppose $\tau \leq n$ and write

$$\begin{aligned}
\mathrm{E}(X_\tau) &= \mathrm{E}\left(X_\tau \sum_{i=1}^n \mathbf{1}(\tau = i)\right) \\
&= \sum_{i=1}^n \mathrm{E}\left(X_i \mathbf{1}_{[\tau=i]}\right) \\
\left([\tau = i] \in \mathcal{F}_i,\ (10.3.1) \text{ implies } \mathrm{E}(X_n|\mathcal{F}_i) = X_i\right) &= \sum_{i=1}^n \mathrm{E}\left(X_n \mathbf{1}_{[\tau=i]}\right) \\
&= \mathrm{E}(X_n) \\
&= \mathrm{E}(X_1),
\end{aligned}$$

(10.3.2) implying the last line. This completes the proof. ∎

The above result has interesting applications in gambler's ruin probability.

Example 10.3.2. *Suppose $(X_n : n = 0, 1, 2, \ldots)$ is an SSRW on $\mathbb{Z}$ and $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Let $\alpha, \beta \in \mathbb{N}$ and*

$$\tau = \inf\{n \geq 1 : X_n \text{ equals either of } -\alpha, \beta\}\,. \tag{10.3.3}$$

In other words, $[X_\tau = -\alpha]$ is the event that a gambler, who enters a casino (where on each bet, the gambler either wins a rupee or loses a rupee, each with probability $1/2$, independently of all previous bets) with α rupees and wishes to continue until they win β rupees, is ruined before achieving the target. We wish to calculate the ruin probability, which is $P(X_\tau = -\alpha)$.

For a fixed $n \geq 1$, $\tau \wedge n$ is a stopping time with respect to $\{\mathcal{F}_n\}_{n\geq 1}$ because

$$[\tau \wedge n = i] = [\tau = i] \in \mathcal{F}_i \text{ for } i = 1, \ldots, n-1,$$

and

$$[\tau \wedge n = n] = [\tau \geq n] = [\tau \leq n-1]^c \in \mathcal{F}_{n-1} \subset \mathcal{F}_n\,.$$

The optional stopping theorem implies

$$\mathrm{E}(X_{\tau\wedge n}) = 0\,.$$

It is easy to see that $\tau < \infty$ a.s. Indeed,

$$\begin{aligned}
P(\tau < \infty) &\geq P\left(\bigcup_{n=0}^{\infty} \bigcap_{i=n(\alpha+\beta)+1}^{(n+1)(\alpha+\beta)} [X_i - X_{i-1} = 1]\right) \\
&= 1 - P\left(\bigcap_{n=0}^{\infty} \left(\bigcap_{i=n(\alpha+\beta)+1}^{(n+1)(\alpha+\beta)} [X_i - X_{i-1} = 1]\right)^c\right) \\
&= 1 - \prod_{n=0}^{\infty} \left(1 - 2^{-\alpha-\beta}\right) \\
&= 1\,.
\end{aligned}$$

Thus $\tau \wedge n = \tau$ for large n a.s. Hence

$$X_{\tau\wedge n} \to X_\tau \text{ a.s.}$$

Since $|X_{\tau\wedge n}| \leq \alpha \vee \beta$, DCT implies

$$\mathrm{E}(X_\tau) = \lim_{n\to\infty} \mathrm{E}(X_{\tau\wedge n}) = 0\,.$$

Recalling that X_τ equals either $-\alpha$ or β, it thus follows that

$$P(X_\tau = -\alpha) = \frac{\beta}{\alpha+\beta}\,. \tag{10.3.4}$$

Exercise 10.3.2. *For τ as in* (10.3.3)*, show that*

$$\mathrm{E}(\tau) = \alpha\beta\,.$$

Hint. *Use* (10.3.4) *and the fact that $(X_n^2 - n, \mathcal{F}_n)_{n\geq 1}$ is a martingale.*

The following is another interesting consequence of the optional stopping theorem.

Theorem 10.3.2 (Wald's identity)**.** *If $Z_1, Z_2, \ldots$ are i.i.d. random variables with a finite mean μ, $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$ and τ is an $\{\mathcal{F}_n\}_{n\geq 1}$-stopping time with a finite mean, then*

$$\mathrm{E}\left(\sum_{i=1}^{\tau} Z_i\right) = \mu \mathrm{E}(\tau)\,.$$

Proof. Let us first show the identity when $Z_1 \geq 0$. Setting

$$X_n = \sum_{i=1}^{n} Z_i - n\mu,\ n = 1, 2, \ldots,$$

it is obvious that $(X_n, \mathcal{F}_n)$ is a martingale. The optional stopping theorem shows that

$$\mathrm{E}(X_{\tau\wedge n}) = \mathrm{E}(X_1) = 0,\ n \geq 1\,.$$

That is,

$$\mathrm{E}\left(\sum_{i=1}^{\tau\wedge n} Z_i\right) = \mu \mathrm{E}(\tau \wedge n)\,.$$

Letting $n \to \infty$ and using MCT, the desired identity follows when $Z_1 \geq 0$.

For i.i.d. $Z_1, Z_2, \ldots$ which are not necessarily non-negative, $Z_1^+, Z_2^+, \ldots$ are i.i.d. The proved identity for non-negative random variables shows

$$\mathrm{E}\left(\sum_{i=1}^{\tau} Z_i^+\right) = \mathrm{E}\left(Z_1^+\right)\mathrm{E}(\tau),$$

and likewise for the negative parts. Combining the two completes the proof. □

Exercise 10.3.3. *If $Z_1, Z_2, \ldots$ are independent non-negative random variables each having a finite mean μ, $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$ and τ is an $\{\mathcal{F}_n\}_{n\geq 1}$-stopping time with a finite mean, show that*

$$\mathrm{E}\left(\sum_{i=1}^{\tau} Z_i\right) = \mu \mathrm{E}(\tau)\,.$$