Measure Theory for Probability

Leonardo T. Rolla

12th July 2022

Reusing this material
This licence allows users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The licence allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified version under identical terms.

Attribution
This is an updated and expanded version of the lecture notes "Mathematics of Random Events" written in collaboration with Nikolaos Constantinou in 2020. Those notes were largely based on material typeset by Nikolaos Constantinou as a student, when the ST342 module was taught by Wei Wu in the Autumn 2019 term at Warwick. The module taught by Wei Wu was largely based on handwritten lecture notes received from Larbi Alili, which in turn were based on material previously developed by Sigurd Assing, Anastasia Papavasileiou, Jon Warren and Wilfrid Kendall.

Contents

Notation and terminology

Notation

$\mathbb{N}$	$\{1, 2, 3, 4, \dots\}$
$\mathbb{N}_0$	$\{0, 1, 2, 3, \dots\}$
A^c	$\Omega \setminus A$ (note that Ω is implicit)
$A_n \uparrow A$	$A_n \subseteq A_{n+1}$ for all n and $\cup_n A_n = A$
$A_n \downarrow A$	$A_n \supseteq A_{n+1}$ for all n and $\cap_n A_n = A$
$x_n \uparrow x$	x_n is non-decreasing and $x_n \to x$
$x_n \downarrow x$	x_n is non-increasing and $x_n \to x$
$\{f < g\}$	$\{\omega \in \Omega : f(\omega) < g(\omega)\}$
$\sup_n f_n$	the function h defined by $h(\omega) = \sup_n f_n(\omega)$ for each ω
$\max\{f_1, f_2\}$	$\sup_n f_n$ over $n \in \{1, 2\}$
$[x]^+$	$\max\{0, x\}$, the positive part of x
$[x]^-$	$\max\{0, -x\}$, the negative part of x, so $x = [x]^+ - [x]^-$
$\lfloor x \rfloor$	the largest $n \in \mathbb{Z}$ such that $n \leqslant x \in \mathbb{R}$, or x if $x = \pm\infty$
μ-a.e.	except on a set A with $\mu(A) = 0$
a.e.	μ-a.e. when μ is obvious from the context
μ-a.e. ω	every $\omega \in \Omega \setminus A$, for some set $A \in \mathcal{F}$ with $\mu(A) = 0$
$\mathbb{P}$-a.s.	with probability 1
a.s. on ω	for $\mathbb{P}$-a.e. ω

Terminology

σ-additivity	Countable additivity
Set	a sample space or a subset of a sample space
Event	a measurable subset of a sample space
Collection	a set of subsets of a sample space
Class	a collection with some structure
Family	a set of classes or functions

1 Infinity

Can we toss a coin infinitely many times?

Why is countable additivity important?

We know that $\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y$, but why? It doesn't seem to follow from the usual definition that treats discrete and continuous random variables as hermetically separated entities. So, what is expectation, really?

Do continuous random variables even exist?

How small can a class $\mathcal{A}$ of events be so that their probabilities determine $\mathbb{P}$?

Why can we differentiate moment generating functions to compute moments?

Our aim is to study concepts of Measure Theory useful to Probability Theory, providing a solid ground for the latter. A *measure* generalises the notion of area for arbitrary sets in Euclidean spaces $\mathbb{R}^d$, $d \geqslant 1$. We introduce the theory of measurable spaces, measurable functions, integral with respect to a measure, density of measures, product measures, and convergence of functions and random variables.

Below we give examples that motivate the need for such a theory, discuss in which sense modern Measure Theory is the best we can hope for, and introduce the concept of infinite numbers and infinite sums used throughout the remaining chapters.

1.1 Events at infinity

We know that $A_n \uparrow A$ implies $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ which, assuming that $\mathbb{P}$ is non-negative and finitely additive, is equivalent to $\mathbb{P}$ being σ-additive (a shorthand for countably additive). Below we see examples where interesting models and events require some sort of limiting process in their study.

Example 1.1. Coin flips $\{0, 1\}$.

(i) Coin flip N times, so the sample space is $\Omega = \{(x_1, \dots, x_N) : x_i \in \{0, 1\}\}$. For $N = 7$, define the event $A = \{(x_1, \dots, x_N) : x_1 = 1, x_5 = 0\}$. Rigorously, the probability of this event is $\mathbb{P}(A) = 1/4$.

(ii) Coin flip infinitely many times (if we can!), so the sample space in this case is $\Omega = \{(x_1, x_2, \dots) : x_i \in \{0, 1\}\}$. Let B be the event of 10^6 consecutive zeros, i.e. $B = \bigcup_{n=1}^{\infty} \{\omega \in \Omega : x_n = x_{n+1} = \cdots = x_{n+10^6-1} = 0\}$. The probability $\mathbb{P}(B)$ is computed by first an estimate, and then taking limit.

△

Example 1.2 (Ruin probability). Gambling with initial wealth $X_0 \in \mathbb{N}_0$. For any $t \geqslant 1$, we bet an integer amount and reach a wealth denoted by X_t. If at any point in time, wealth amounts to 0, it remains 0 forever. The sample space that indicates the wealth process is $\Omega = \{(x_0, x_1, \dots) : x_i \in \mathbb{N}_0\}$. We define the function of wealth after the n-th gamble by, $X_n : \Omega \to \mathbb{N} \cup \{0\}$, where $X_n(\omega) = x_n$. In fact, $(X_n)_{n \geqslant 0}$ is a Markov process. Then

$$\{\text{stay in state 0 eventually}\} = \bigcup_{n=0}^{\infty} \bigcap_{m=n}^{\infty} \{\omega \in \Omega : X_m(\omega) = 0\}. \tag{1.3}$$

By monotonicity and continuity, its probability is

$$\mathbb{P}(\{\text{stay in state 0 eventually}\}) = \lim_{n \to \infty} \mathbb{P}(\{\omega \in \Omega : X_n(\omega) = 0\}). \qquad \triangle$$

Example 1.4 (Brownian motion). It is possible to construct a sequence of continuous piece-wise linear functions, which in the limit give a continuous nowhere differentiable random path that is at the core of Stochastic Analysis. $\triangle$

Example 1.5 (Uniform variable). If $X \sim U[0, 1]$, then $\mathbb{P}(X \neq x) = 1$ for every $x \in \mathbb{R}$. Nevertheless, $\mathbb{P}(X \neq x \text{ for every } x \in \mathbb{R}) = 0$, so there will be some unlucky x that will happen to be hit by X. Now for a countable set A, by σ-additivity we have $\mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x) = 0$. Since $\mathbb{Q}$ is countable, we see that a uniform random variable is always irrational. Well, this is unless there is an uncountable number of such variables in the same probability space, in which case some unlucky variables may happen to take rational values. $\triangle$

Example 1.6 (Strong law of large numbers). Let $X_1, X_2, X_3, \dots$ be independent and take value ± 1 with probability $\frac{1}{2}$ each, and take $S_n = X_1 + \cdots + X_n$. The strong law of large numbers says that, almost surely (a shorthand for "$\mathbb{P}(\cdots) = 1$"), $\frac{S_n}{n} \to 0$ as $n \to \infty$. In order to show this, we need to express this event as a limit and compute its probability. The first part is simply:

$$\{\lim_{n \to \infty} \tfrac{S_n}{n} = 0\} = \bigcap_{k=1}^{\infty} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{|\tfrac{S_n}{n}| < \tfrac{1}{k}\}. \qquad \triangle$$

Example 1.7 (Recurrence of a random walk). For the sequence $(S_n)_n$ of the previous example, we know that $\frac{S_n}{n} \to 0$ with probability one. We want to consider whether $\frac{S_n}{n}$ converges to zero from above, from below, or oscillating, which is the same as asking whether $S_n = 0$ infinitely often. $\triangle$

1.2 The measure problem

It is clear (it should be!) from previous examples that we want to work with measures that have nice *continuity* properties, so we can take limits. However, when the mass is spread over uncountably many sample points $\omega \in \Omega$, it is not possible to assign a measure to all subsets of Ω in a reasonable way.

We would like to define a random variable uniformly distributed on $[0, 1]$, by means of a function that assigns a weight to subsets of this interval. For instance, what is the probability that this number is irrational? What is the probability that its decimal expansion does not contain a 3?

This is the same problem as assigning a 'length' to subsets of $\mathbb{R}$. We are also interested in defining a measure of 'area' on $\mathbb{R}^2$, 'volume' on $\mathbb{R}^3$, and so on.

Of course, a good measure of length/area/volume/etc. on $\mathbb{R}^d$ should:

(i) give the correct value on obvious sets, such as intervals and balls;

(ii) give the same value if we rotate or translate a set;

(iii) be σ-additive.

We stress again that σ-additivity is equivalent to a measure being continuous, and we are not willing to resign that. On the other hand, we do not want more than that: each of the uncountably many points in $[0,1]$ alone has length zero, but all together they have length one; likewise, each sequence of coin tosses in $\{0,1\}^{\mathbb{N}}$ has probability zero, but all together they have probability one.

The measure problem is the following.

> There is no measure m defined on all subsets of $\mathbb{R}^d$ which satisfy all the reasonable properties listed above. What modern Measure Theory does is to work with measures that are defined on a class of sets which is large enough to be useful and small enough for these properties to hold.

The next example shows that there is no measure m defined on all subsets of $\mathbb{R}^3$ which satisfies these three properties.

Example 1.8 (Banach-Tarski paradox). Consider the ball

$$B = \{x \in \mathbb{R}^3 : \|x\| \leqslant 1\}.$$

There exist[1] $k \in \mathbb{N}$, *disjoint* sets $A_1, \dots, A_{2k}$, and *isometries* (maps that preserve distances and angles) $S_1, \dots, S_{2k} : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$B = (A_1 \cup \cdots \cup A_k) \cup (A_{k+1} \cup \cdots \cup A_{2k}),$$

$$B = S_1 A_1 \cup \cdots \cup S_k A_k, \quad B = S_{k+1} A_{k+1} \cup \cdots \cup S_{2k} A_{2k}.$$

So B was decomposed into *finitely many* pieces, which were later on moved around *rigidly* and *recombined* to produce two copies of B! Why is it a paradox? Finitely many pieces is not the issue in itself, since $\mathbb{N}$ can be decomposed into even and odd numbers, and they can be *compressed* (or stretched, in some sense) to produce two copies of $\mathbb{N}$. Rigidity alone is not the issue either, since we can move each of the *uncountably many* points of the segment $[0,1]$ to form the segment $[0,2]$. The paradox is that this magic was done with *rigid movements on finitely many pieces.* And here we can see the *measure problem*: if all these disjoint sets $A_1, \dots, A_{2k}$ were to have a volume $V_1, \dots, V_{2k} \geqslant 0$, what would be the volume of the ball B? △

[1] See https://youtu.be/s86-Z-CbaHA for a nice overview of the proof.

The next example is not nearly as effective in impressing friends at a party, and would certainly not make a youtube video with 31 million views, but it has two advantages. First, it shows directly that the measure problem already occurs on $d = 1$. Second, we can actually explain its proof in a third of a page rather than a dozen.

Example 1.9 (Vitali set). Consider the unit circle $\mathbb{S}^1$ with points indexed by turns instead of degrees or radians. This is the same as the interval $\mathbb{S}^1 = [0, 1)$ with the angle addition operation $x \oplus y = x + y \mod 1$. There exists a set $E \subseteq [0, 1)$ such that $\mathbb{S}^1$ is decomposed into disjoint $\{E_n\}_{n\in\mathbb{N}}$ which are translations of E. And here we see again the *measure problem*: by σ-additivity, if the length of E is zero, then the length of the circle is zero; and if the length of E is non-zero, then the length of the circle is infinite. So E is not measurable. △

Sketch of proof. Write $\mathbb{Q}\cap[0, 1) = \{r_n\}_{n=1,2,3,\dots}$. For $E \subseteq \mathbb{S}^1$, let $E_n = \{x\oplus r_n : x \in E\}$ be the *translation* of E by r_n. We want to find a set E such that

(i) The sets $E_1, E_2, E_3, \dots$ are disjoint,

(ii) The union satisfies $\cup_n E_n = \mathbb{S}^1$.

Start with a small set that satisfies the first property, such as $E = \{0\}$. Enlarge the set E by adding a point $x \in \mathbb{S}^1 \setminus (\cup_n E_n)$. Adding such point does not break the first property (proof omitted), and may help the second one. Keep adding points this way, until it is no longer possible. When it is no longer possible, it can only be so because the second property is also satisfied.[2] □

Remark 1.10. It is often emphasised that the Banach-Tarski paradox and Vitali set depend crucially on the Axiom of Choice (for the above sketch of proof, it is concealed in the expression "keep adding until"). We may wonder what happens if we do not accept this axiom. In this case, we cannot prove the Banach-Tarski paradox, nor the existence of a Vitali set. But neither can we prove that they do not exist, so the measure problem persists. △

1.3 Infinite numbers and infinite sums

We now define the set of extended real numbers and briefly discuss some its useful properties, then discuss the meaning of infinite sums, and move on to other perhaps philosophical questions about this theory.

1.3.1 Extended real numbers

We are about to start working with measures, and because measures can be infinite, and integrals can be negative infinite, we work with the set of *extended real numbers* $\overline{\mathbb{R}} := [-\infty, +\infty]$ that extends $\mathbb{R}$ by adding two symbols $\pm\infty$. The

[2] See [Coh13, 1.4.9] for a complete proof.

novelty is of course to conveniently allow operations and comparisons involving these symbols.

Basically, we can safely operate as one would reasonably guess:

$$-\infty < -1 < 0 < 5 < +\infty,\ -7 + \infty = +\infty,\ (-2) \times (-\infty) = +\infty,$$
$$|-\infty| = +\infty,\ (+\infty) \times (-\infty) = -\infty,\ a \leqslant b \implies a + x \leqslant b + x$$
$$\lim_{n\to\infty} (2 + n^2) = \lim_{n\to\infty} 2 + \lim_{n\to\infty} n^2 = 2 + \infty = +\infty,$$

etc. Since we will never need to divide by infinity, let us leave $\frac{x}{\infty}$ undefined (otherwise we would need to check that x is finite).

The non-obvious definition is $0 \cdot \infty = 0$. In Calculus, it would have been considered an indeterminate form, but in Measure Theory it is convenient to define it this way because the integral of a function that takes value 0 on an interval of infinite length and $+\infty$ at a few points should still be 0. That is, the area of a rectangle having zero width and infinite length is zero.

Now some caveats. First, $\lim_n(a_n b_n) = (\lim_n a_n)(\lim_n b_n)$ may fail in case it gives $0 \cdot \infty$. Also, note that now $a + b = a + c$ does not imply $b = c$. This can be false when $a = \pm\infty$. Likewise, $a < b$ no longer implies that $a + x < b + x$. So we should be careful with cancellations.

> The one thing that is definitely not allowed, and that Measure Theory does not handle well, is
> $$\text{“ } +\infty - \infty \text{ ” !}$$
> This is simply forbidden, and if we will ever write this, it will be in quotation marks and just in order to say that this case is excluded.

The reader should consult [Coh13, §§B.4–B.6] and [Tao11, p. xi] for a more complete description of operations on $\overline{\mathbb{R}}$.

1.3.2 Infinite sums

Infinite sums of numbers on $[0, +\infty]$ are always well-defined through a rather simple formula. If Λ is an index set and $x_\alpha \in [0, +\infty]$ for all $\alpha \in \Lambda$, we define:

$$\sum_{\alpha\in\Lambda} x_\alpha = \sup_{\substack{A\subseteq\Lambda \\ A \text{ finite}}} \sum_{\alpha\in A} x_\alpha.$$

The set Λ can be uncountable, but the sum can be finite only if $\Lambda_+ = \{\alpha : x_\alpha > 0\}$ is countable (proof omitted). If $x_\alpha \in [-\infty, +\infty]$, we define $\Lambda_- = \{\alpha : x_\alpha < 0\}$ and

$$\sum_{\alpha\in\Lambda} x_\alpha = \sup_{\substack{A\subseteq\Lambda_+ \\ A \text{ finite}}} \sum_{\alpha\in A} x_\alpha \ - \sup_{\substack{A\subseteq\Lambda_- \\ A \text{ finite}}} \sum_{\alpha\in A} -x_\alpha, \tag{1.11}$$

as long as this difference does not give "$+\infty - \infty$" !

The theory of conditionally convergent sums as

$$\sum_{j\in\mathbb{N}} x_j = \lim_n \sum_{j=1}^{n} x_j \tag{1.12}$$

is hardly meaningful to us. In case the expression in (1.11) is well-defined, we can write $(\Lambda_- \cup \Lambda_+) = \{\alpha_j\}_{j\in\mathbb{N}}$ by ordering these indices in any way we want (assuming for simplicity that these sets are countable), and formula (1.12) will give the same result as (1.11). Pretty robust.

However, in case (1.12) converges but (1.11) is not well-defined (so it gives "$+\infty - \infty$"), we can re-order the index set $\mathbb{N}$ so that (1.12) will give any number we want. This is definitely not the type of delicacy we want to handle here.

For this reason, we will only use (1.12) when either

$$x_j \in [0, +\infty]$$

for all j, or when

$$\sum_j |x_j| < \infty.$$

So there are two overlapping cases where we can work comfortably: non-negative extended numbers, or series which are absolutely summable.

1.3.3 Two different and overlapping theories

The above tradeoff is already a good prelude to something rather deep that will appear constantly in upcoming chapters. Borrowing from [Tao11]:

> Because of this tradeoff, we will see two overlapping types of measure and integration theory: the *non-negative* theory, which involves quantities taking values in $[0, +\infty]$, and the *absolutely integrable* theory, which involves quantities taking values in $\mathbb{R}$ or $\mathbb{C}$.

However, at the risk of leaving it for the reader to figure out some corner cases, we can (and will) extend these theories to a theory on $[-\infty, +\infty]$ by doing what we just did above. Namely, whereas the absolutely integrable theory requires that both terms in (1.11) be finite, and the non-negative theory requires that one of them be zero, we only require that one of them be finite.

We end this chapter with the simplest example of these overlapping theories.

Theorem 1.13 (Tonelli Theorem for series)**.** *Let $x_{m,n} \in [0, +\infty]$ be a doubly-indexed sequence. Then*

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{m,n} = \sum_{(m,n)\in\mathbb{N}^2} x_{m,n} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{m,n}.$$

Proof. A proof is given in §7.2 as an application of Tonelli Theorem, but here is a bare hands proof. For $A \subseteq \mathbb{N}^2$ finite, there exist $m, n \in \mathbb{N}$ such that $\sum_A x_{j,k} \leqslant \sum_{j=1}^n \sum_{k=1}^n x_{j,k}$. Conversely, given $m, n \in \mathbb{N}$, there is $A \subseteq \mathbb{N}^2$ such that $\sum_A x_{j,k} = \sum_{j=1}^n \sum_{k=1}^n x_{j,k}$. Hence,

$$\sum_{(j,k)\in\mathbb{N}^2} x_{j,k} = \sup_A \sum_{(j,k)\in A} x_{j,k} = \sup_{m,n\in\mathbb{N}} \sum_{j=1}^m \sum_{k=1}^n x_{j,k}$$

and therefore

$$\begin{aligned}\sum_{(j,k)\in\mathbb{N}^2} x_{j,k} &= \sup_{m\in\mathbb{N}} \sup_{n\in\mathbb{N}} \sum_{j=1}^m \sum_{k=1}^n x_{j,k} = \sup_{m\in\mathbb{N}} \lim_n \sum_{j=1}^m \sum_{k=1}^n x_{j,k} = \\ &= \sup_{m\in\mathbb{N}} \sum_{j=1}^m \left(\lim_n \sum_{k=1}^n x_{j,k} \right) = \sup_{m\in\mathbb{N}} \sum_{j=1}^m \left(\sup_{n\in\mathbb{N}} \sum_{k=1}^n x_{j,k} \right) = \sum_{j=1}^\infty \left(\sum_{k=1}^\infty x_{j,k} \right).\end{aligned}$$

The other equality is proved in identical way. □

Theorem 1.14 (Fubini Theorem for series)**.** *Let $x_{m,n} \in [-\infty, +\infty]$ be a doubly-indexed sequence. If*

$$\sum_{m=1}^\infty \sum_{n=1}^\infty |x_{m,n}| < \infty,$$

then

$$\sum_{m=1}^\infty \sum_{n=1}^\infty x_{m,n} = \sum_{(m,n)\in\mathbb{N}^2} x_{m,n} = \sum_{n=1}^\infty \sum_{m=1}^\infty x_{m,n}.$$

Proof. Given in §7.2 as a application of Fubini Theorem. □

Example 1.15 (Failure)*.* Consider the doubly-indexed sequence

$$x_{m,n} = \begin{array}{|cccccc} 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 0 & 1 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{array}$$

which is not absolutely summable. Note that summing columns and then rows we get 1, whereas summing rows and then columns we get 0. △

2 Measure Spaces

The goal in this chapter is to define the notion of a measure space $(\Omega, \mathcal{F}, \mu)$ and work on examples to get familiar with the relevant concepts.

Later on, we will focus our attention on two types of measures: the Lebesgue measure m and probability measures $\mathbb{P}$.

2.1 Classes of sets

This may be the least pleasant topic, but we *need* to go through it.

2.1.1 σ-algebras

We start with the notion of σ-algebra on a sample space Ω.

Definition 2.1 (σ-algebra)**.** Let Ω be a sample space. A collection $\mathcal{F}$ of subsets of Ω is called a *σ-algebra* (or σ-field) on Ω, if

(i) $\Omega \in \mathcal{F}$,

(ii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,

(iii) $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*. The sets $A \in \mathcal{F}$ are called *$\mathcal{F}$-measurable sets*, or *measurable sets* when $\mathcal{F}$ is clear from the context.

Remark 2.2. In this case, it also satisfies

(iv) $\emptyset \in \mathcal{F}$,

(v) $A_1, A_2, \dots \in \mathcal{F} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$. △

Let us review the following examples to build our intuition.

Example 2.3. For a given sample space Ω, let $\mathcal{P}(\Omega) := \{A : A \subseteq \Omega\}$ be the power set of Ω. Then $\mathcal{P}(\Omega)$ is a σ-algebra on Ω. △

Example 2.4. For a given sample space Ω, let $\mathcal{A} = \{\emptyset, \Omega\}$. Then $\mathcal{A}$ defines a σ-algebra on Ω, and it is called the *trivial σ-algebra*. △

Exercise 2.5. Let Ω be a set and $\mathcal{A}$ the class of all subsets which are either countable or whose complement is countable. Prove that $\mathcal{A}$ is a σ-algebra. △

Exercise 2.6. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be σ-algebras on a given sample space Ω, and let $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ be the class made of the sets that belong to both $\mathcal{A}_1$ and $\mathcal{A}_2$. Prove that $\mathcal{A} := \mathcal{A}_1 \cap \mathcal{A}_2$ is also a σ-algebra on Ω. △

Exercise 2.7 (Pull-back). If $f : \Omega_1 \to \Omega_2$ and $\mathcal{F}_2$ is a σ-algebra on Ω_2, prove that $f^{-1}(\mathcal{F}_2) := \{f^{-1}(A) : A \in \mathcal{F}_2\}$ is a σ-algebra on Ω_1. △

So when a function goes from a sample space Ω_1 to another space Ω_2, it pulls back any σ-algebra $\mathcal{F}$ on Ω_2 to a σ-algebra $\mathcal{G}$ on Ω_1, as in the following diagram:

$$\begin{tikzcd} \Omega_1 \arrow[r, dotted, no head] \arrow[d, "f"] & \mathcal{G} \\ \Omega_2 \arrow[r, dotted, no head] & \mathcal{F} \arrow[u, "f^{-1}"'] \end{tikzcd} \tag{2.8}$$

2.1.2 Algebras

On a few occasions, we will use a class with less structure than σ-algebras.

Definition 2.9 (Algebra)**.** Let Ω be a sample space. A class $\mathcal{A}$ of subsets of Ω is called an *algebra on* Ω, if

(i) $\Omega \in \mathcal{A}$,

(ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$,

(iii) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$.

Remark 2.10. In this case, it also satisfies:

(iv) $\emptyset \in \mathcal{A}$,

(v) $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$.

In other words, an algebra on Ω is a class of subsets of Ω, which contains Ω and is *stable under finitely many set operations*. △

Example 2.11. Every σ-algebra is an algebra. △

Remark 2.12. If Ω is a sample space and $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \cdots$ are σ-algebras, then $\mathcal{A} := \cup_n \mathcal{F}_n$ may not be a σ-algebra, but it is still an algebra. △

Exercise 2.13. Prove that $\mathcal{A}$ in the previous remark is indeed an algebra. △

Example 2.14. Let $\Omega = \mathbb{R}$ and

$$\mathcal{E} = \{(a, b] : -\infty < a < b < +\infty\} \cup \{(a, +\infty) : a \in \mathbb{R}\} \cup \{(-\infty, b] : b \in \mathbb{R}\}.$$

Consider the class

$$\mathcal{A} = \{I_1 \cup \cdots \cup I_n : n \in \mathbb{N}, I_1, \dots, I_n \in \mathcal{E}, I_k \cap I_j = \emptyset \text{ for } j \neq k\} \cup \{\mathbb{R}, \emptyset\}.$$

Then $\mathcal{A}$ is an algebra. △

Exercise 2.15. Prove that $\mathcal{A}$ in the previous example is an algebra. Suggestion: show that $\mathcal{E} \cup \{\emptyset\}$ is closed under intersections, use this to show that $\mathcal{A}$ is closed under intersections, also show that $\mathcal{A}$ is closed under complement, and conclude that $\mathcal{A}$ is closed under unions. △

2.1.3 The σ-algebra generated by a class

It will often be very useful to consider the σ-algebra "spanned" by a class.

Definition 2.16 (σ-algebra generated by a class of subsets)**.** Let $\mathcal{E}$ be a class of subsets of Ω. We define *the σ-algebra generated by $\mathcal{E}$*, denoted $\sigma(\mathcal{E})$, as the *unique* class of subsets of Ω with the following properties:

(i) $\sigma(\mathcal{E})$ is a σ-algebra,

(ii) $\sigma(\mathcal{E}) \supseteq \mathcal{E}$,

(iii) if $\mathcal{F} \supseteq \mathcal{E}$ and $\mathcal{F}$ is a σ-algebra, then $\mathcal{F} \supseteq \sigma(\mathcal{E})$.

So $\sigma(\mathcal{E})$ is the smallest σ-algebra that contains $\mathcal{E}$.

We now prove that $\sigma(\mathcal{E})$ is indeed well-defined.

Proof by exercise. Let $\{\mathcal{A}_\alpha\}_{\alpha\in\Lambda}$ be a family of σ-algebras on Ω, where Λ is an arbitrary *non-empty* index set. Prove that $\bigcap_{\alpha\in\Lambda} \mathcal{A}_\alpha$ is also a σ-algebra. Let $\mathcal{E}$ be a class of subsets of Ω. Prove that there exists *at least one* σ-algebra on Ω that contains $\mathcal{E}$ as a subclass. Prove that $\sigma(\mathcal{E})$ exists by considering the family of σ-algebras that contain $\mathcal{E}$. Finally, prove that $\sigma(\mathcal{E})$ is unique. $\square$

One might describe $\sigma(\mathcal{E})$ as the collection of sets that can be obtained by countably many operations of taking union, intersection and complement of sets in $\mathcal{E}$. This description may indeed give some comfort to those facing such bitter abstraction for the first time, but we should not rely on it because it is hard to make it precise. It is better to work directly with the definition.

The examples below illustrate this.

Example 2.17. Let Ω be a sample space, $A \subseteq \Omega$ and $\mathcal{E} = \{A\}$. We claim that the σ-algebra generated by $\mathcal{E}$ is $\mathcal{A} := \{\emptyset, \Omega, A, A^c\}$. Indeed, $\mathcal{A}$ is a σ-algebra, it contains $\mathcal{E}$, and any σ-algebra that contains $\mathcal{E}$ must contain $\mathcal{A}$. $\triangle$

Example 2.18. Let $\mathcal{E} = \{A_1, \dots, A_n\}$, be a partition of a sample space Ω. Then the σ-algebra generated by $\mathcal{E}$ is $\mathcal{A} = \{A_{i_1} \cup \dots \cup A_{i_m} : i_1, \dots, i_m \in \{1, \dots, n\}\} \cup \{\emptyset\}$. Indeed, $\mathcal{A}$ is a σ-algebra, it contains $\mathcal{E}$, and any σ-algebra containing $\mathcal{E}$ must contain $\mathcal{A}$. $\triangle$

Exercise 2.19. Let $\Omega = \{1, 2, 3\}$ and $\mathcal{E} = \{\{1\}, \{2\}, \{3\}\}$. What is $\sigma(\mathcal{E})$? $\triangle$

Exercise 2.20. Let $\mathcal{E} = \{\{x\} : x \in \Omega\}$ be the class of singletons. What is $\sigma(\mathcal{E})$? $\triangle$

Proposition 2.21. *Suppose $\mathcal{E}_1$ and $\mathcal{E}_2$ are classes of subsets of Ω and $\mathcal{E}_1 \subseteq \sigma(\mathcal{E}_2)$. Prove that $\sigma(\mathcal{E}_1) \subseteq \sigma(\mathcal{E}_2)$.*

Exercise 2.22. Prove the above proposition. $\triangle$

The Borel sets on $\mathbb{R}$ and $\overline{\mathbb{R}}$

See page 92 for a very brief recall of basic facts from Metric Spaces. When working on $\mathbb{R}$ or $\overline{\mathbb{R}}$, the most important σ-algebra is the following.

Definition 2.23 (Borel sets on $\mathbb{R}$)**.** For the space $\Omega = \mathbb{R}$, the *Borel σ-algebra* $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by open sets. It is also the smallest σ-algebra that contains all the open intervals, or all the closed intervals.

For the extended line, we define $\mathcal{B}(\overline{\mathbb{R}}) := \{A \subseteq \overline{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\}$, that is, we allow adding the points $\pm\infty$ to sets in $\mathcal{B}(\mathbb{R})$.

When it is clear that $\Omega = \mathbb{R}$ or $\overline{\mathbb{R}}$, we may write $\mathcal{B}$ instead of $\mathcal{B}(\mathbb{R})$ or $\mathcal{B}(\overline{\mathbb{R}})$.

Below we list more classes of sets, all of which generate $\mathcal{B}(\mathbb{R})$. Any "reasonable" set is a Borel set, and it is actually hard to construct one which is not, such as Vitali sets from Example 1.9. Nevertheless, we should keep in mind that all the theory works on σ-algebras, and $\mathcal{B}$ is the one we are most interested in.

Denote by $\mathcal{E}_1, \dots, \mathcal{E}_6$ the following classes of subsets of $\mathbb{R}$, respectively:

1. Closed subsets of $\mathbb{R}$.
2. Open subsets of $\mathbb{R}$.
3. Intervals of the form (a, b), $a, b \in [-\infty, +\infty]$.
4. Finite intervals of the form $(a, b]$, $a, b \in \mathbb{R}$.
5. Semi-infinite intervals of the form $(-\infty, b]$, $b \in \mathbb{R}$.
6. Semi-infinite intervals of the form $(-\infty, b)$, $b \in \mathbb{R}$.

One can show that $\sigma(\mathcal{E}_1) = \cdots = \sigma(\mathcal{E}_6)$ as follows:

$$\sigma(\mathcal{E}_2) \subseteq \sigma(\mathcal{E}_1) \subseteq \sigma(\mathcal{E}_2) \subseteq \sigma(\mathcal{E}_3) \subseteq \sigma(\mathcal{E}_4) \subseteq \sigma(\mathcal{E}_5) \subseteq \sigma(\mathcal{E}_6) \subseteq \sigma(\mathcal{E}_2).$$

We prove two inclusions and leave five as exercise.

Let $A \in \mathcal{E}_2$. Then A is the countable union of disjoint open intervals. So there exist $\{I_n\}_{n\in\mathbb{N}} \subseteq \mathcal{E}_3 \subseteq \sigma(\mathcal{E}_3)$ such that $A = \cup_n I_n$. Since $\sigma(\mathcal{E}_3)$ is a σ-algebra, it is closed under countable unions, and $A \in \sigma(\mathcal{E}_3)$. Thus, $\mathcal{E}_2 \subseteq \sigma(\mathcal{E}_3)$. By Proposition 2.21, $\sigma(\mathcal{E}_2) \subseteq \sigma(\mathcal{E}_3)$.

Let $A \in \mathcal{E}_5$. Then $A = (-\infty, b]$ for some $b \in \mathbb{R}$. Consider the sequence of sets $A_n = (-\infty, b + \frac{1}{n})$ in $\mathcal{E}_6$. Since $\sigma(\mathcal{E}_6) \supseteq \mathcal{E}_6$ is a σ-algebra, it is closed under countable intersections, and $A = \cap_n A_n \in \sigma(\mathcal{E}_6)$. Thus, $\mathcal{E}_5 \subseteq \sigma(\mathcal{E}_6)$. By Proposition 2.21, $\sigma(\mathcal{E}_5) \subseteq \sigma(\mathcal{E}_6)$.

Exercise 2.24. Prove the missing inclusions in the above chain. △

Exercise 2.25. Show that each of the classes below generate $\mathcal{B}(\overline{\mathbb{R}})$:

(a) $\{[-\infty, a)\}_{a\in\mathbb{R}}$,

(b) $\{[-\infty, a]\}_{a\in\mathbb{R}}$,

(c) $\{[a, +\infty]\}_{a\in\mathbb{R}}$,

(d) $\{(a, +\infty]\}_{a\in\mathbb{R}}$.

Hint: first show that, if a σ-algebra contains $\{-\infty, +\infty\}$ and $\mathcal{B}(R)$, then it contains $\mathcal{B}(\overline{\mathbb{R}})$, then use the previous exercise. △

2.2 Measures

Definition 2.26 (Measure and σ-finite measure)**.** Let $(\Omega, \mathcal{F})$ be a given measurable space. A function $\mu : \mathcal{F} \to [0, +\infty]$ is called a *measure* if $\mu(\emptyset) = 0$ and

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

for every sequence $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ of disjoint sets.

In this case, the triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*.

We say that μ is *finite* if $\mu(\Omega) < \infty$ and *σ-finite* if there exist measurable sets $(A_n)_{n\in\mathbb{N}}$ such that $\mu(A_n) < \infty$ for all n and $\cup_{n\in\mathbb{N}} A_n = \Omega$.

Let us start with some simple but important examples.

Example 2.27 (Dirac mass measure). Let Ω be a set, $\mathcal{F} = \mathcal{P}(\Omega)$ and $x \in \Omega$. The measure δ_x is defined as

$$\delta_x(A) = \mathbb{1}_A(x)$$

for all $A \subseteq \Omega$. △

Remark 2.28. Even though an indicator function and a Dirac mass function perform a similar test, they are different objects. The indicator function of a given $A \subseteq \Omega$ is a function

$$\mathbb{1}_A : \Omega \to \{0, 1\},$$

whereas the Dirac measure on a given $x \in \Omega$ is a map

$$\delta_x : \mathcal{P}(\Omega) \to \{0, 1\}.$$

In other words, $\mathbb{1}_A$ checks whether a point lies in A, whereas δ_x checks whether a set contains x. △

Example 2.29 (Counting measure). Let Ω be a set, $\mathcal{F} = \mathcal{P}(\Omega)$, and

$$\mu = \sum_{x\in\Omega} \delta_x.$$

So $\mu(A) = \sum_{x\in\Omega} \delta_x(A)$ counts how many elements A has. △

Example 2.30. For $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\Omega)$, the function $\mu(A) = \sum_{k\in A} k^{-1}$. △

Exercise 2.31. Let $\mu_1, \mu_2, \dots$ be a sequence of measures on a measurable space $(\Omega, \mathcal{F})$, and let $x_1, x_2, \dots \in [0, +\infty]$. Then

$$\mu = \sum_n x_n \mu_n$$

is a measure on $(\Omega, \mathcal{F})$. △

Example 2.32 (Coarsening). If $(\Omega, \mathcal{F}, \mu)$ is a measure space, and $\mathcal{G} \subseteq \mathcal{F}$ is also a σ-algebra, then $(\Omega, \mathcal{G}, \mu)$ is also a measure space. △

Example 2.33 (Restriction). If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $A \in \mathcal{F}$, then the restriction of μ to A, denoted $\mu_{|_A}$ and given by

$$\mu_{|_A}(B) = \mu(A \cap B)$$

is also a measure. △

Exercise 2.34. Prove that the functions described in previous examples are indeed measures. △

Proposition 2.35 (Properties of a measure)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then, we have the following properties:*

(i) *Countable sub-additivity: If $\{A_i\}_{i=1}^\infty \subseteq \mathcal{F}$, then*

$$\mu\big(\cup_{i=1}^\infty A_i\big) \leqslant \sum_{i=1}^\infty \mu(A_i).$$

(ii) *Continuity from below: If $\{A_i\}_{i=1}^\infty \subseteq \mathcal{F}, A_k \subseteq A_{k+1}, \forall k \in \mathbb{N}$, then $\mu(\bigcup_{i=1}^\infty A_i) := \mu(\lim\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} \mu(A_n)$.*

(iii) *Continuity from above: If $\{A_i\}_{i=1}^\infty \subseteq \mathcal{F}, A_k \supseteq A_{k+1}, \forall k \in \mathbb{N}$, and $\mu(A_j) < \infty$ for some $j \in \mathbb{N}$, then $\mu(\bigcap_{i=1}^\infty A_i) := \mu(\lim\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} \mu(A_n)$.*

Proof. For sub-additivity, set $B_k := A_k \setminus (\bigcup_{i=1}^{k-1} A_i), \forall k \in \mathbb{N}$, so that $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty B_i$, B_k are disjoint and $B_k \subseteq A_k, \forall k \in \mathbb{N}$. It follows that,

$$\mu\left(\bigcup_{i=1}^\infty A_i\right) = \mu\left(\bigcup_{i=1}^\infty B_i\right) = \sum_{i=1}^\infty \mu(B_i) \leqslant \sum_{i=1}^\infty \mu(A_i).$$

For continuity from below, let $A_0 = \emptyset$ and $B_k := A_k \setminus A_{k-1}, \forall k \in \mathbb{N}$, so that $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty B_i$ and B_k are disjoint. It follows that,

$$\mu\left(\bigcup_{i=1}^\infty A_i\right) = \mu\left(\bigcup_{i=1}^\infty B_i\right) = \sum_{i=1}^\infty \mu(B_i)$$

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

For continuity from above, let $B_k := A_j \setminus A_{j+k}, \forall k \in \mathbb{N}$, use continuity from below, and subtract both sides from $\mu(A_j)$. This concludes the proof. □

Remark 2.36. Continuity from above may be false without the assumption that $\mu(A_j)$ is finite for some j. For instance, if μ is the counting measure on $\mathbb{N}$ and $A_n = \{n, n+1, n+2, \dots\}$. Then $A_n \downarrow \emptyset =: A$. Since $\mu(A_n) = \infty$ and $\mu(A) = 0$, we have $\mu(A_n) \not\to \mu(A)$ even though $A_n \downarrow A$. △

We now state the most important measure in the theory.

Theorem 2.37 (Lebesgue Measure)**.** *There exists a unique measure m on $(\mathbb{R}, \mathcal{B})$ such that*

$$m\big((a,b]\big) = b - a$$

for all $a < b \in \mathbb{R}$. This measure m is called the Lebesgue Measure *on $\mathbb{R}$.*

Proof. Uniqueness will be proved in §3.1 and existence in §3.2. □

Example 2.38 (Countable sets). Any countable set has Lebesgue measure 0. Although $\mathbb{Q}$ is present everywhere in the sense that it intersects every tiny open interval, its length is actually zero. △

Example 2.39 (Restriction). Given real numbers $a < b$, the function μ given by

$$\mu(A) = m(A \cap [a,b])$$

is a measure on $(\mathbb{R}, \mathcal{B})$. △

Definition 2.40 (Probability)**.** A measure μ on a measurable space $(\Omega, \mathcal{F})$ is called a *probability measure* if $\mu(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, \mu)$ is called a *probability space.*

Probability measures are often denoted by P, $\mathbb{P}$ or $\mathbf{P}$ instead of μ.

Example 2.41 (Equiprobable outcomes). Let $n \in \mathbb{N}$ and let Ω be a set with n elements. Take $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P} = \frac{1}{n} \sum_{\omega\in\Omega} \delta_\omega$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is the model of equiprobable outcomes. △

Example 2.42. Let $n \in \mathbb{N}$, $\Omega = \{\omega_1, \dots, \omega_n\}$, and $\mathcal{E} = \{A_1, \dots, A_n\}$, where $A_i = \{\omega_i\}, i = 1, \dots, n$. Set $\mathbb{P}(A_i) = \mathbb{P}(\{\omega_i\}) := p_i \geqslant 0, i = 1, \dots, n$, such that $\sum_{i=1}^n p_i = 1$. We can extend the definition of $\mathbb{P}$ to $\mathcal{P}(\Omega)$ by finite additivity. Then, we can check that $\mathbb{P}$ defines a probability measure on $(\Omega, \mathcal{P}(\Omega))$. △

Exercise 2.43. Let $\mathbb{P}_1, \mathbb{P}_2, \dots$ be a sequence of probability measures on a measurable space $(\Omega, \mathcal{F})$, and let $p_1, p_2, \dots \in [0,1]$ be such that $\sum_n p_n = 1$ Show that

$$\mathbb{P} = \sum_n p_n \mathbb{P}_n$$

is a probability measure on $(\Omega, \mathcal{F})$. Give an interpretation to $\mathbb{P}$ and $(p_n)_n$. △

Definition 2.44 (Finite and σ-finite measure)**.** We say that a measure μ is *finite* if $\mu(\Omega) < \infty$ and *σ-finite* if there exist measurable sets $(A_n)_{n \in \mathbb{N}}$ such that $\mu(A_n) < \infty$ for all n and $\cup_{n \in \mathbb{N}} A_n = \Omega$.

If $\mu(A) < \infty$, then the restriction of μ to A is a finite measure. The Dirac mass is a probability measure, hence a finite measure. The counting measure on Ω is a finite measure if Ω is a finite set, is a σ-finite measure if Ω is countable, and is not a σ-finite measure if Ω is uncountable.

Every finite measure is σ-finite (indeed, one can check the definition by taking $A_n = \Omega$ for all n). The Lebesgue measure is σ-finite (take $A_n = [-n, n]$ for all n). The simplest example of a measure which is not σ-finite is the counting measure on $\mathbb{R}$.

Whereas most of the theory developed in later chapters apply to measures in general, several properties hold for finite and σ-finite measures only. For instance, uniqueness in Carathéodory Extension Theorem, Theorem 3.1, the Radon-Nikodým Theorem, uniqueness of a product measure, Tonelli Theorem, and Fubini Theorem, all require that measures be σ-finite. When a measure is finite, we have continuity from above, Proposition 3.5 holds, and bounded functions are integrable (see Chapter 5).

3 Existence and uniqueness

In this chapter we introduce two foundational theorems in measure and probability theories. The π-λ Theorem allows us to establish that certain properties hold on a large class of sets by checking them on a much smaller class. This is useful especially when we try to show that two measures are equal by checking that they assign the same value to sets in a specific collection.

The Carathéodory Extension Theorem provides existence of measure spaces starting from the specification on an algebra. We discuss how this tool is used to construct the Lebesgue measure on $\mathbb{R}$ as well as Lebesgue-Stieltjes measures. We conclude by discussing the Lebesgue measure on $\mathbb{R}^d$.

3.1 Dynkin's π-λ Theorem and uniqueness

In this section we will study Dynkin's π-λ Theorem and use it to prove the following.

Theorem 3.1 (Uniqueness of measures). *Let μ and ν be measures on a measurable space $(\Omega, \mathcal{F})$, and let $\mathcal{C} \subseteq \mathcal{F}$ be a class of subsets of Ω closed under intersection. Suppose that $\mu(A_n) = \nu(A_n) < \infty$ for some sequence $A_n \uparrow \Omega$ of sets in $\mathcal{C}$. If $\mu = \nu$ on $\mathcal{C}$, then $\mu = \nu$ on $\sigma(\mathcal{C})$.*

Of course, the uniqueness part of Theorem 2.37 is a particular case of the above theorem. For that, we take $\mathcal{C} = \{(a, b] : a, b \in \mathbb{R}\}$, since $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ and note that $(-n, n] \uparrow \mathbb{R}$ are in $\mathcal{C}$ and have finite measure.[3]

Before introducing the concepts used in its proof, let us discuss where they arise from. Suppose we want to check that two *finite* measures μ and ν on the measurable space $(\Omega, \mathcal{F})$ coincide on a large class of sets, such as $\mathcal{F}$. Suppose also that $\mu(\Omega) = \nu(\Omega)$. On how many sets A do we need to test that $\mu(A) = \nu(A)$? Testing all $A \in \mathcal{F}$ is a bit overkilling. Having tested for a disjoint sequence $(A_n)_n$, it will hold for A_1^c as well as $\cup_n A_n$ by elementary properties of finite measures. However, having tested for given sets $A, B \in \mathcal{F}$, there is no way to infer from the basic properties of measures alone that these two measures coincide on $A \cap B$. So the idea is to consider the class

$$\mathcal{D} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\} \tag{3.2}$$

and show that it is large. This class already has some structure inherited from the way finite measures treat disjoint unions, but that structure is still missing a small bit, namely being closed under intersections.

Definition 3.3 (π-system). A class $\mathcal{C}$ of sets is called a π-*system* if it is closed under intersections, that is, $A \cap B \in \mathcal{C}$ for all $A, B \in \mathcal{C}$.

[3] Note that $(a, b] = \emptyset$ when $b \leqslant a$, so $\emptyset \in \mathcal{C}$ and $\mathcal{C}$ is closed under intersections.

Good examples of π-systems are the classes listed in §2.1.3, if we add $\emptyset$ to them.

Definition 3.4 (λ-system)**.** Let Ω be a sample space. A class $\mathcal{D}$ of subsets of Ω is a *λ-system on Ω*, if $\Omega \in \mathcal{D}$ and it is closed under complement and countable disjoint unions, that is,

(i) $\Omega \in \mathcal{D}$,

(ii) $A \in \mathcal{D} \implies A^c \in \mathcal{D}$,

(iii) $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{D},\ A_i \cap A_j = \emptyset\ \forall i \neq j \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$.

These concepts are useful when the (large and complicated) class of sets satisfying a given property will naturally form a λ-system, and at the same time we can find a (smaller and simpler) *subclass* that forms a π-system and generates a large σ-algebra. This is the case for Theorem 3.1.

Proposition 3.5. *If μ and ν are measures on a measurable space $(\Omega, \mathcal{F})$ such that $\mu(\Omega) = \nu(\Omega) < \infty$, then the class $\mathcal{D}$ defined in (3.2) is a λ-system.*

Proof. Suppose $A \in \mathcal{D}$. Then $\mu(A^c) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^c)$, so $A^c \in \mathcal{D}$. Now consider a countable collection $\{A_n\}_n \subseteq \mathcal{D}$ of disjoint sets. Then $\mu(\cup_n A_n) = \sum_n \mu(A_n) = \sum_n \nu(A_n) = \nu(\cup_n A_n)$, so $\cup_n A_n \in \mathcal{D}$. □

We now have enough motivation for the following celebrated tool.

Theorem 3.6 (Dynkin's π-λ Theorem)**.** *Let Ω be a sample space. Suppose $\mathcal{C}$ is a π-system of subsets of Ω and $\mathcal{D}$ is a λ-system on Ω. If $\mathcal{D}$ contains $\mathcal{C}$, then $\mathcal{D}$ contains $\sigma(\mathcal{C})$.*

Proof. Given on page 94. □

Proof of Theorem 3.1. Assume for a moment that $\mu(\Omega) = \nu(\Omega) < \infty$. Define the class $\mathcal{D}$ as in (3.2). By Proposition 3.5, $\mathcal{D}$ is a λ-system. By the π-λ Theorem, $\sigma(\mathcal{C}) \subseteq \mathcal{D}$, which means exactly that $\mu = \nu$ on $\sigma(\mathcal{C})$ as claimed.

Now drop the previous assumption, and instead assume that $\mu(A_n) = \nu(A_n) < \infty$ for some sequence $A_n \uparrow \Omega$ of sets in $\mathcal{C}$. For each $n \in \mathbb{N}$, define the restriction measures μ_n and ν_n by $\mu_n(A) = \mu(A \cap A_n)$ and $\nu_n(A) = \nu(A \cap A_n)$. With this definition, $\mu_n(\Omega) = \mu(A_n) = \nu(A_n) = \nu_n(\Omega) < \infty$. Note also that, for each $C \in \mathcal{C}$, we have $C \cap A_n \in \mathcal{C}$ because $\mathcal{C}$ is a π-system. Hence, $\mu_n(C) = \mu(C \cap A_n) = \nu(C \cap A_n) = \nu_n(C)$ for every $C \in \mathcal{C}$.

So, for each n, the measures μ_n and ν_n are in the previous case, and, for every $A \in \sigma(\mathcal{C})$, we have $\mu_n(A) = \nu_n(A)$. Using continuity from below,

$$\mu(A) = \lim_n \mu(A \cap A_n) = \lim_n \mu_n(A) = \lim_n \nu_n(A) = \lim_n \mu(A \cap A_n) = \nu_n(A),$$

which concludes the proof.[4] $\square$

Another useful application of the π-λ Theorem is to study independent random variables. Suppose we know that the vector (X, Y) satisfies

$$\mathbb{P}\big((X,Y) \in (-\infty, r] \times (-\infty, t]\big) = \mathbb{P}\big(X \in (-\infty, r]\big)\mathbb{P}\big(Y \in (-\infty, t]\big)$$

for all $r, t \in \mathbb{R}$. Does it imply that

$$\mathbb{P}\big((X,Y) \in A \times B\big) = \mathbb{P}\big(X \in A\big)\mathbb{P}\big(Y \in B\big)$$

for every $A, B \in \mathcal{B}(\mathbb{R})$? The answer is yes, and the proof is a little more ingenious than the one we just showed. It uses twice the fact that $\{(-\infty, b] : b \in \mathbb{R}\}$ is a π-system that generates $\mathcal{B}(\mathbb{R})$. We leave it as a teaser for now.

3.2 Carathéodory Extension Theorem and existence

In this section, we introduce the key theorem that allows us to extend a "measure" μ from an algebra to a larger σ-algebra.

3.2.1 Pre-measures and the extension theorem

Given a sample space Ω and an algebra $\mathcal{A}$, we say that a function $\mu : \mathcal{A} \to [0, \infty]$ is a *finitely additive measure* if it is non-negative and finitely additive, that is,

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

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

Note that *finitely additive* measures are not *true* measures. The theory of finitely additive measures is rather limited, as one cannot take limits there.

Example 3.7 (Uniform "measure" on $\mathbb{N}$). Suppose we pick a large natural number at random. The chances that this number is odd should be $\frac{1}{2}$. The chances that a year is a leap year should be $\frac{97}{400}$. This notion of chance is finitely additive. For instance, the chances that a year is either odd or a leap year would be $\frac{297}{400}$, and the chance that it is not a leap year would be $\frac{303}{400}$. We could even talk about independent events. For instance, this random number being a multiple of 5 and a multiple of 6 would be independent, whereas a year being multiple of 5 and leap would not. This "frequentist" notion would be

$$\mu(A) = \lim_{n\to\infty} \frac{\#(A \cap [1, n])}{n},$$

and that this μ can be extended[5] to a finitely additive measure defined on all $\mathcal{P}(\mathbb{N})$, including sets A for which the limit does not exist. However, taking $A_n = \{n, n+1, n+2, \dots\}$, we have $\mu(A_n) = 1$ for every n in spite of $A_n \downarrow \emptyset$ and $\mu(\emptyset) = 0$. Not really the type of measure that we want to study here. $\triangle$

[4] Expanded from [Coh13, 1.6.4].

[5] This requires the Hahn-Banach Theorem which is a topic of Functional Analysis.

Fortunately, asking μ to be "σ-additive whenever possible" is enough to fix this.

Definition 3.8 (Pre-measure)**.** Given a sample space Ω and an algebra $\mathcal{A}$, we say that a finitely additive measure $\mu : \mathcal{A} \to [0, \infty]$ is a *pre-measure* if it is σ-additive, that is, $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ for every sequence of disjoint sets $A_n \in \mathcal{A}$ whose union A happens to be in $\mathcal{A}$.

We now state a fundamental theorem of Measure Theory and Probability.

> **Theorem 3.9** (Carathéodory Extension Theorem)**.** *Let $\mathcal{A}$ be an algebra on Ω. If μ is a pre-measure on $(\Omega, \mathcal{A})$, then μ extends to a measure on $(\Omega, \sigma(\mathcal{A}))$. Moreover, if the extension μ is σ-finite, then it is unique.*

We now outline the proof, and then consider its main applications.

3.2.2 Outer measures and measurable sets

Here we outline the proof of Theorem 3.9. Given a pre-measure μ on $(\Omega, \mathcal{A})$, we define the *outer measure* $\mu^* : \mathcal{P}(\Omega) \to [0, \infty]$ as follows. For each subset E of Ω, consider the collection $\mathcal{D}_E$ of all sequences $(A_n)_n$ of sets in $\mathcal{A}$ that cover E:

$$\mathcal{D}_E = \Big\{ (A_n)_n : A_n \in \mathcal{A} \text{ for all } n \text{ and } E \subseteq \bigcup\nolimits_n A_n \Big\}.$$

The outer measure μ^* is defined by

$$\mu^*(E) = \inf_{(A_n) \in \mathcal{D}_E} \sum_n \mu(A_n). \tag{3.10}$$

One might think of the outer measure as a conservative upper bound for what should be the measure of a set seen from the outside.

Lemma 3.11 (Properties of the outer measure)**.** *The outer measure μ^* satisfies:*

(i) *For every $E \subseteq F \subseteq \Omega$, $\mu^*(E) \leqslant \mu^*(F)$,*

(ii) *For $(E_n)_{n \in \mathbb{N}}$ subsets of Ω, then $\mu^*(\cup_n E_n) \leqslant \sum_n \mu^*(E_n)$,*

(iii) *For $A \in \mathcal{A}$, $\mu^*(A) = \mu(A)$.*

Proof. Given on page 94. □

Loosely speaking, problematic sets such as those of Examples 1.9 and 1.8 are those which look bigger from the outside than form the inside. It turns out, the way to exclude sets like this is to consider sets $A \subseteq \Omega$ such that,

$$\text{for every } E \subseteq \Omega,\ \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

The sets A satisfying this condition are called *μ^*-measurable*, and $\mathcal{F}^*$ denotes the collection of such sets.

The second property in the above lemma implies that

$$\mu^*(E) \leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c),$$

so problematic sets are those A for which inequality is strict. For a Vitali set B as described in Example 1.9, we have $\mu^*(B) + \mu^*([0,1) \cap B^c) > 1$, which one might interpret as B being larger seen from the outside than from the inside.

Lemma 3.12 (Carathéodory)**.** *The class $\mathcal{F}^*$ of μ^*-measurable sets is a σ-algebra on Ω which contains $\mathcal{A}$, and the restriction of μ^* to $\mathcal{F}^*$ is a measure.*

Proof. Given on page 95. □

From the two lemmas above, we have that $\mathcal{A} \subseteq \mathcal{F}^*$, hence $\sigma(\mathcal{A}) \subseteq \mathcal{F}^*$. Moreover, μ^*, restricted to $\sigma(\mathcal{A})$, is σ-additive, hence it is a measure. Furthermore, μ^*, restricted to $\mathcal{A}$, coincides with μ. This proves the existence part of Theorem 3.9.

For uniqueness, suppose μ^* is σ-finite, so there exist a countable collection $\{A_n\}_n$ in $\sigma(\mathcal{A})$ such that $\mu^*(A_n) < \infty$ for all n, and $\cup_n A_n = \Omega$. By (3.10), for each n there is a countable collection $\{A_{n,k}\}_k$ such that $\cup_k A_{n,k} \supseteq A_n$ and $\mu^*(A_{n,k}) < \infty$. Re-indexing the doubly-indexed countable collection $\{A_{n,k}\}_{n,k}$ as $\{B_j\}_{j\in\mathbb{N}}$ and defining $C_\ell = B_1 \cup \cdots \cup B_\ell$, we have $C_\ell \in \mathcal{A}$, $\mu(C_\ell) < \infty$ and $C_\ell \uparrow \Omega$. Noting that $\mathcal{A}$ is a π-system, it follows from the π-λ Theorem that μ^* is the unique measure on $\sigma(\mathcal{A})$ that agrees with μ on $\mathcal{A}$, concluding the proof of Theorem 3.9.

3.2.3 Lebesgue measure on $\mathbb{R}$

Here we prove the existence part in Theorem 2.37. Fix $\Omega = \mathbb{R}$ and

$$\mathcal{E} = \{(a,b] \cap \mathbb{R} : -\infty \leqslant a < b \leqslant +\infty\}$$

as in Example 2.14. We already know that the class $\mathcal{A}$ of sets A given by $\mathbb{R}$, $\emptyset$ and finite unions of disjoint left open intervals is an algebra. Set $m(I) = +\infty$ for infinite intervals I, and

$$m((a,b]) := b - a$$

for finite intervals $(a,b]$. We can extend the definition of m to $\mathcal{A}$ by

$$m(A) = \sum_{j=1}^{n} m(I_j), \tag{3.13}$$

where $A = I_1 \cup \cdots \cup I_n$ and $I_k \in \mathcal{E}$ are disjoint. Note that the above formula is well-defined even if A can be written in many different ways as such a finite disjoint union (for instance, if $I_1 = (0,1]$, $I_2 = (1,2]$, it gives $m((0,2]) = m((0,1]) + m((1,2]) = 2$). We omit the straightforward proof of this fact.

Lemma 3.14. *The set function $m : \mathcal{A} \to [0, \infty]$ is finitely additive.*

Proof. Immediate from the fact that (3.13) is well-defined. $\square$

What is not so easy to prove is that m is σ-additive. For instance, take $A_k = (\frac{1}{k+1}, \frac{1}{k}] \in \mathcal{E}$. Note that the A_k's are disjoint and $\bigcup_k A_k = (0, 1]$, which happens to be in $\mathcal{A}$. In this case,

$$\sum_{j=1}^{\infty} m(A_j) = \sum_{j=1}^{\infty} \left(\frac{1}{j} - \frac{1}{j+1} \right) = 1 = m((0, 1]).$$

To conclude the proof of Theorem 2.37 from Theorem 3.9, one needs to show that this property holds in general.

Lemma 3.15. *The set function $m : \mathcal{A} \to [0, \infty]$ is a pre-measure.*

Proof. Given on page 96. $\square$

Example 3.16. If we take $A \subseteq [0, 1)$ as a Vitali set, then $A \notin \mathcal{B}(\mathbb{R})$.

We sketch the proof. This is just expanding part of the sketch in Example 1.9. If A were in $\mathcal{B}(\mathbb{R})$, we could define $m(A) \geqslant 0$. Then by translation invariance, for each $r_n \in \mathbb{Q} \cap [0, 1)$, defining $A_n = \{x + r_n \mod 1 : x \in A\}$, we would have $m(A_n) = m(A)$. Moreover, these sets $(A_n)_n$ are mutually disjoint and $\cup_n A_n = [0, 1)$. So $m([0, 1)) = \sum_{n \in \mathbb{N}} m(A_n) = \sum_{n \in \mathbb{N}} m(A)$ which is either 0 or ∞ depending on whether $m(A) > 0$ or not. This contradicts $m([0, 1)) = 1$. △

3.2.4 Lebesgue-Stieltjes measures

Let $F : \mathbb{R} \to \mathbb{R}$ be a right-continuous non-decreasing function. Consider the sample space $\Omega = \mathbb{R}$ and the class $\mathcal{E} = \{(a, b] : -\infty < a < b < +\infty\}$. Define the set function $m_F : \mathcal{E} \to [0, \infty)$ by $m_F((a, b]) = F(b) - F(a)$.

We can define m_F on semi-infinite intervals by taking limits as $a \to -\infty$ or $b \to +\infty$. Consider the algebra $\mathcal{A}$ defined in Example 2.14 and revisited in §3.2.3. We can extend m_F to $\mathcal{A}$ in a unique (and obvious) way so that it is finitely additive. It can be shown that m_F is, in fact, a pre-measure on $\mathcal{A}$. Again, by Carathéodory Extension Theorem, there exists a measure m_F on $(\mathbb{R}, \mathcal{B})$ that gives the right values on $\mathcal{E}$. Moreover, since $\mathcal{E}$ is a π-system which contains a sequence $(-n, n] \uparrow \mathbb{R}$ of sets with finite measure, this measure is unique. Such measure is called the *Lebesgue-Stieltjes measure* induced by the function F on the space $\mathbb{R}$.

Remark 3.17. If $F(x) = x$, then m_F reduces to the Lebesgue measure on $\mathbb{R}$. △

Remark 3.18. We say that F is a *distribution function* if it is right-continuous, non-decreasing, $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. In this case, m_F is a probability measure on $\mathbb{R}$. Conversely, if $\mathbb{P}$ is a probability measure on $\mathbb{R}$, then

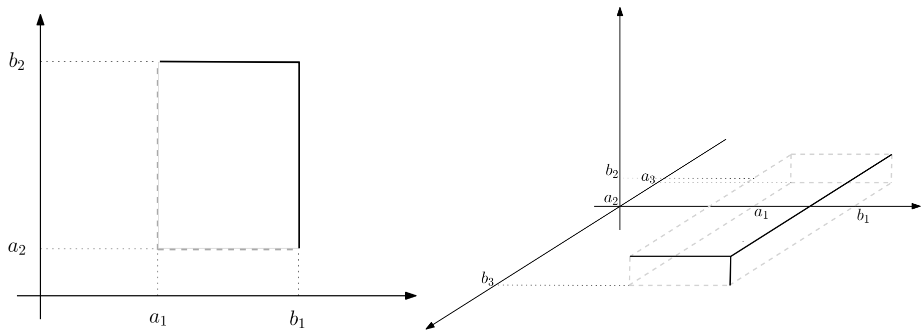


Figure 3.1: Elements of $\underline{\mathcal{E}}$ for $d = 2$ and $d = 3$.

$F(x) = \mathbb{P}(\{\omega : \omega \leqslant b\})$ is a distribution function. So these two types of objects are equivalent. $\triangle$

3.2.5 Lebesgue measure on $\mathbb{R}^d$

Consider the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ denotes *the Borel σ-algebra on* $\mathbb{R}^d$, $d \geqslant 1$. More precisely, we define $\mathcal{B}(\mathbb{R}^d) := \sigma(\underline{\mathcal{E}})$, where

$$\underline{\mathcal{E}} = \{(\underline{a}, \underline{b}] : -\infty < a_i < b_i < +\infty, i = 1, \ldots, d\}$$

and

$$(\underline{a}, \underline{b}] = (a_1, b_1] \times \ldots (a_d, b_d] \subseteq \mathbb{R}^d.$$

Figure 3.1 illustrates the structure of elements in $\underline{\mathcal{E}}$ for higher dimensions, which take the shape of rectangles and cuboids for $d = 2$ and $d = 3$.

It can be checked that the collection

$$\mathcal{A} := \{I_1 \cup \cdots \cup I_n : n \in \mathbb{N}_0, I_k \in \underline{\mathcal{E}} \text{ for } k = 1, \ldots, n\}$$

is an algebra of sets on $\mathbb{R}^d$.

Now, for any $(\underline{a}, \underline{b}] \in \underline{\mathcal{E}}$, we define the set function

$$m\big((\underline{a}, \underline{b}]\big) := (b_1 - a_1) \cdots (b_d - a_d)$$

as the d-dimensional volume of the cuboid $(\underline{a}, \underline{b}]$. As in §3.2.3, it is possible to prove that m defines a pre-measure on $\mathcal{A}$. Hence, by Carathéodory Extension Theorem, m extends to a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The measure m is called the *Lebesgue measure on* $\mathbb{R}^d$, which is the analogue to the 1-dimensional case. Also, $\underline{\mathcal{E}}$ is a π-system on $\mathbb{R}^d$ that generates $\mathcal{B}(\mathbb{R}^d)$, from which one can also show uniqueness of m.

Example 3.19. Take a point set $\{\underline{x}\} = \{(x_1, \ldots, x_d)\} \in \mathcal{B}(\mathbb{R}^d)$. Indeed, as one might expect, the Lebesgue measure m on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ assigns value of 0 for $\{\underline{x}\}$. To show this rigorously, consider $A_k = \{\underline{y} \in \mathbb{R}^d : x_i - \frac{1}{k} \leqslant y_i \leqslant$

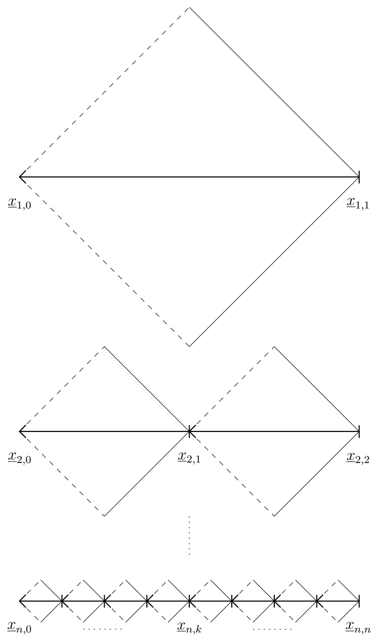


Figure 3.2: Covering L by smaller boxes as L_n.

$x_i, i = 1, \ldots, d\} \in \mathcal{B}(\mathbb{R}^d)$, where $A_k \supseteq A_{k+1}, \forall k \in \mathbb{N}$. Then, one see that $\{\underline{x}\} = \bigcap_{i=1}^{\infty} A_i$. Now, by continuity from above, it follows that,

$$m(\{\underline{x}\}) = m\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} m(A_n) = \lim_{n\to\infty}\left(\frac{1}{n}\right)^d = 0.$$

△

Remark 3.20. As in the above example, monotone-convergence properties of a measure are quite helpful when it comes to computing measures of sets. △

Remark 3.21. If a set in $\mathcal{B}(\mathbb{R}^d)$ is countable, then it has Lebesgue measure 0. Equivalently, if a set in $\mathcal{B}(\mathbb{R}^d)$ has strictly positive Lebesgue measure, then it is uncountable. △

Example 3.22. For $d \geqslant 2$, we look at a line segment $L = \{\underline{x} \in \mathbb{R}^d : \underline{x} = r\underline{\alpha} + \underline{\beta},\ r \in [0,1]\}$, where $\underline{\alpha}, \underline{\beta} \in \mathbb{R}^d$ are fixed. Set points on L by, $\underline{x}_{n,k} = \frac{k}{n}\underline{\alpha} + \underline{\beta}$, so as to consider $L_n = \bigcup_{k=1}^{n}(\underline{x}_{n,k-1}, \underline{x}_{n,k}] \in \mathcal{B}(\mathbb{R}^d)$, where $L_n \supseteq L_{n+1}, \forall n \in \mathbb{N}$. The idea is to iteratively split the line segment L into smaller segments and cover each such segment by via small squares L_n, which is captured in Figure 3.2. Now, we can see that, $L = \bigcap_{i=1}^{\infty} L_i$, so applying continuity from above yields,

$$m(L) = m\left(\bigcap_{i=1}^{\infty} L_i\right) = \lim_{n\to\infty} m(L_n) = 0.$$

△

Remark 3.23. Notice that, L lies in a subspace of $\mathbb{R}^d$. In general, if $A \in \mathcal{B}(\mathbb{R}^d)$ lies in a subspace of $\mathbb{R}^d$, then $m(A) = 0$. $\triangle$

4 Measurable functions and random variables

This chapter intends to introduce and construct the concept of measurable functions, through a number of important properties and theorems.

4.1 Definition and properties

As its name suggests, a measurable function between two measurable spaces preserves the measurability properties of sets.

Definition 4.1 (Measurable function)**.** Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. A function $f : \Omega_1 \to \Omega_2$ is called $\mathcal{F}_1/\mathcal{F}_2$-*measurable* if the pre-image of any $\mathcal{F}_2$-measurable set is $\mathcal{F}_1$-measurable. That is,

$$f^{-1}(A) \in \mathcal{F}_1, \ \forall A \in \mathcal{F}_2,$$

where $f^{-1}(A) := \{\omega_1 \in \Omega_1 : f(\omega_1) \in A\}$ is the pre-image of f. When the σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ are clear from context, we often drop '$\mathcal{F}_1/\mathcal{F}_2$-measurable' to just 'measurable' or '$\mathcal{F}_1$-measurable'.

Example 4.2. Let $(\Omega, \mathcal{F})$ be a measurable space and $A \in \mathcal{F}$. We show that the indicator function $\mathbb{1}_A : \Omega \to \{0, 1\}$ given by

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

is $\mathcal{F}/\mathcal{P}(\{0, 1\})$-measurable. Indeed, we have that

- $\mathbb{1}_A^{-1}(\emptyset) = \emptyset \in \mathcal{F}$,
- $\mathbb{1}_A^{-1}(\{0\}) = A^c \in \mathcal{F}$,
- $\mathbb{1}_A^{-1}(\{1\}) = A \in \mathcal{F}$,
- $\mathbb{1}_A^{-1}(\{0, 1\}) = \mathbb{1}_A^{-1}(\{0\}) \cup \mathbb{1}_A^{-1}(\{1\}) = A^c \cup A = \Omega \in \mathcal{F}$,

so $\mathbb{1}_A$ is indeed measurable. △

Naturally, one can expect from Definition 4.1 that a composition of measurable functions is measurable in any general measurable spaces.

Lemma 4.3 (Composition of measurable functions)**.** *Let* $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ *and* $(\Omega_3, \mathcal{F}_3)$ *be measurable spaces and* $f : \Omega_1 \to \Omega_2$, $g : \Omega_2 \to \Omega_3$ *be measurable functions. Then, the composition* $g \circ f$ *is* ($\mathcal{F}_1/\mathcal{F}_3$-)*measurable.*

Proof. Since g is measurable, then for any $A \in \mathcal{F}_3$, we have $g^{-1}(A) \in \mathcal{F}_2$. Now, since f is measurable, then $f^{-1}(g^{-1}(A)) \in \mathcal{F}_1$. But, this implies that

$$(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)) \in \mathcal{F}_1, \ \forall A \in \mathcal{F}_3,$$

so that the composition $g \circ f$ is indeed measurable. $\square$

Given a target measurable space $(\Omega_2, \mathcal{F}_2)$ and a function $f : \Omega_1 \to \Omega_2$, we can construct a σ-algebra on Ω_1 induced by f.

Definition 4.4 (σ-algebra generated by a function f)**.** Let $(\Omega_2, \mathcal{F}_2)$ be a measurable space and $f : \Omega_1 \to \Omega_2$ a function, where Ω_1 is a sample space. The *σ-algebra generated by f* on Ω_1 is defined as

$$\sigma(f) := \{f^{-1}(A) : A \in \mathcal{F}_2\},$$

From Exercise 2.7, $\sigma(f)$ is indeed a σ-algebra. See the diagram below:

$$\begin{tikzcd}
\Omega_1 \arrow[r, dotted, no head] \arrow[d, "f"] & \sigma(f) \\
\Omega_2 \arrow[r, dotted, no head] & \mathcal{F}_2 \arrow[u, "f^{-1}"']
\end{tikzcd}$$

Note that the notation $\sigma(f)$ does not make explicit its dependence on $\mathcal{F}_2$.

Remark 4.5. It follows from the definition of measurability that $\sigma(f)$ is the smallest σ-algebra on Ω_1 for which f is measurable. That is, for any σ-algebra $\mathcal{G}$ on Ω_1, the function f is $\mathcal{G}/\mathcal{F}_2$-measurable if and only if $\mathcal{G} \supseteq \sigma(f)$. The latter condition is interpreted as $\mathcal{G}$ being *finer* than $\sigma(f)$. $\triangle$

Definition 4.6. If f is a measurable function from a measure space $(\Omega_1, \mathcal{F}_1, \mu)$ to $(\Omega_2, \mathcal{F}_2)$, we can define the *push-forward* measure $f_*\mu$ on $(\Omega_2, \mathcal{F}_2)$ by

$$(f_*\mu)(A) = \mu(\{\omega \in \Omega_1 : f(\omega) \in A\}) = \mu(f^{-1}(A)).$$

This gives an enlarged diagram as follows:

$$\begin{tikzcd}
\Omega_1 \arrow[r, dotted, no head] \arrow[d, "f"] & \sigma(f) \arrow[r, dotted, no head] & \mu \arrow[d, "\circ f^{-1}"] \\
\Omega_2 \arrow[r, dotted, no head] & \mathcal{F}_2 \arrow[u, "f^{-1}"'] \arrow[r, dotted, no head] & f_*\mu
\end{tikzcd} \tag{4.7}$$

Quite often, it is unhandy to check measurability using the whole target space, in which case the following lemma comes in our aid.

Lemma 4.8 (Criterion for measurability)**.** *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. Suppose $\mathcal{F}_2 = \sigma(\mathcal{E})$ for some class $\mathcal{E}$ of subsets of Ω_2. Then, $f : \Omega_1 \to \Omega_2$ is a measurable function if and only if $f^{-1}(A) \in \mathcal{F}_1$ for all $A \in \mathcal{E}$.*

Proof. The direct implication follows from Definition 4.1: since any $A \in \mathcal{E}$ implies that $A \in \mathcal{F}_2$, then measurability of f gives $f^{-1}(A) \in \mathcal{F}_1$. So we only need to prove the converse implication. Suppose that $f^{-1}(A) \in \mathcal{F}_1$ for each $A \in \mathcal{E}$. We will first that the collection $\mathcal{G}$ given by

$$\mathcal{G} = \{A \subseteq \Omega_2 : f^{-1}(A) \in \mathcal{F}_1\}$$

is a σ-algebra on Ω_2, and then conclude that f is measurable. We check the definition of σ-algebra by using elementary properties of set operations:

(i) First, $\Omega_2 \in \mathcal{G}$, since $f^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$.

(ii) Let $A \in \mathcal{G}$, so $f^{-1}(A) \in \mathcal{F}_1$. Since f has domain Ω_1 and codomain Ω_2, we have $f^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus f^{-1}(A)$. Since $\mathcal{F}_1$ is closed under complements, we have $\Omega_1 \setminus f^{-1}(A) \in \mathcal{F}_1$, hence $A^c \in \mathcal{G}$.

(iii) Let $A_1, A_2, \dots \in \mathcal{G}$. Then, $f^{-1}(A_k) \in \mathcal{F}_1$, for each $k \in \mathbb{N}$. This in turn implies that $f^{-1}(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} f^{-1}(A_i) \in \mathcal{F}_1$. Thus, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}$.

Therefore, $\mathcal{G}$ defines a σ-algebra on Ω_2. Now, since $\mathcal{E} \subset \mathcal{G}$, by Definition 2.16 we have $\sigma(\mathcal{E}) \subseteq \mathcal{G}$. That is, every set $B \in \mathcal{F}_2$ is also in $\mathcal{G}$, which, by definition of $\mathcal{G}$ implies that f is measurable. □

Lemma 4.9 (Continuous functions)**.** *Let Ω_1 and Ω_2 be metric spaces. Take $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively, as the Borel σ-algebras, that is, those generated by the families of open sets. If $f : \Omega_1 \to \Omega_2$ is continuous, then it is measurable.*

Proof. Denote by $\mathcal{O}$ the class of open sets in Ω_2. Let $B \in \mathcal{O}$. Since f is continuous, $f^{-1}(B)$ is an open set in Ω_1, hence $f^{-1}(B) \in \mathcal{B}_1$. Since $\mathcal{B}_2 = \sigma(\mathcal{O})$, by Lemma 4.8, f is measurable. □

4.2 Extended Borel functions and random variables

We now discuss extended Borel functions and extended random variables.

4.2.1 Extended Borel functions

Definition 4.10 (Extended Borel function)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. If $f : \Omega \to \mathbb{R}$ is a measurable function, then it is called a *Borel function*. Likewise, a measurable function taking values on $(\overline{\mathbb{R}}, \mathcal{B})$ is called an *extended Borel function*.

Remark 4.11. One can think of a Borel function as simply an extended Borel function that never takes infinite values. So properties of extended Borel functions also hold for Borel functions. △

Recall from §2.1.3 the several classes of subsets of $\mathbb{R}$ that generate $\mathcal{B}(\mathbb{R})$ and hence $\mathcal{B}(\overline{\mathbb{R}})$. We can then apply Lemma 4.8 above to conclude that measurability suffices to hold for each such class on $\overline{\mathbb{R}}$.

Lemma 4.12 (Equivalent criteria). *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \overline{\mathbb{R}}$. The following are equivalent:*

(a) *f is an extended Borel function,*

(b) $\{f < a\} \in \mathcal{F}, \ \forall a \in \mathbb{R}$,

(c) $\{f \leqslant a\} \in \mathcal{F}, \ \forall a \in \mathbb{R}$,

(d) $\{f > a\} \in \mathcal{F}, \ \forall a \in \mathbb{R}$,

(e) $\{f \geqslant a\} \in \mathcal{F}, \ \forall a \in \mathbb{R}$.

Proof. By Exercise 2.25, each one of the classes $\{y < a\}_{a\in\mathbb{R}}$, $\{y \leqslant a\}_{a\in\mathbb{R}}$, $\{y > a\}_{a\in\mathbb{R}}$ and $\{y \geqslant a\}_{a\in\mathbb{R}}$ generates $\mathcal{B}(\overline{\mathbb{R}})$. Hence, the equivalence follows from Lemma 4.8. □

Lemma 4.13 (Pull-back of Borel sets). *Let Ω be a sample space and f a function from Ω to $(\overline{\mathbb{R}}, \mathcal{B})$. Then, $\sigma(f) = \sigma(\{\{f \leqslant a\} : a \in \mathbb{R}\})$.*

Proof. Since $[-\infty, a] \in \mathcal{B}$, we have $\{f \leqslant a\} \in \sigma(f)$, for each $a \in \mathbb{R}$. Thus, $\{f \leqslant a\}_{a\in\mathbb{R}} \subseteq \sigma(f)$, and so $\mathcal{G} := \sigma(\{f \leqslant a\}_{a\in\mathbb{R}}) \subseteq \sigma(\sigma(f)) = \sigma(f)$. On the other hand, by Lemma 4.12 f is $\mathcal{G}/\mathcal{B}(\overline{\mathbb{R}})$-measurable. Therefore, by Remark 4.5, $\mathcal{G} \supseteq \sigma(f)$. □

Example 4.14. Let $(\Omega, \mathcal{F})$ be a measurable space. Given a partition $A_1, \dots, A_n \in \mathcal{F}$ of Ω and distinct numbers $b_1, \dots, b_n \in \mathbb{R}$, define the function $f : \Omega \to \mathbb{R}$ by

$$f(\omega) := b_1 \mathbb{1}_{A_1}(\omega) + \cdots + b_n \mathbb{1}_{A_n}(\omega), \ \omega \in \Omega.$$

That is, $f(\omega) = b_j$ when $\omega \in A_j$. We will show that $\sigma(f) = \sigma(\{A_1, \dots, A_n\})$.

Let $a \in \mathbb{R}$. Denote by $\{b_{i_1}, \dots, b_{i_m}\}$ the collection of values $b \in \{b_1, \dots, b_n\}$ such that $b \leqslant a$. Now note that $\{f \leqslant a\} = \{\omega : \omega \in A_{i_j} \text{ for some } j = 1, \dots, m\}$, that is,

$$\{f \leqslant a\} = A_{i_1} \cup \cdots \cup A_{i_m},$$

so $\{f \leqslant a\} \in \sigma(\{A_1, \dots, A_n\})$. Since this holds for every $a \in \mathbb{R}$, we have

$$\{\{f \leqslant a\} : a \in \mathbb{R}\} \subseteq \sigma(\{A_1, \dots, A_n\}).$$

By Proposition 2.21, $\sigma(\{\{f \leqslant a\} : a \in \mathbb{R}\}) \subseteq \sigma(\{A_1, \dots, A_n\})$ and, by Lemma 4.13, $\sigma(f) \subseteq \sigma(\{A_1, \dots, A_n\})$.

Conversely, observe that

$$\{f = b_i\} = A_i \in \sigma(f), \quad \text{for } i = 1, \dots, n,$$

so $\{A_1, \dots, A_n\} \subset \sigma(f)$, and by Proposition 2.21 $\sigma(\{A_1, \dots, A_n\}) \subseteq \sigma(f)$. △

Exercise 4.15. Consider the measurable space $([0, 1], \mathcal{B}([0, 1]))$, define the functions $f, g : [0, 1] \to \{0, 1, 2, 3\}$ by

$$f(\omega)=\begin{cases}1, & \omega\in[0,\frac{1}{4}]\\ 2, & \omega\in(\frac{1}{4},\frac{1}{2}]\\ 3, & \omega\in(\frac{1}{2},1]\end{cases}\quad,\quad g(\omega)=\begin{cases}0, & \omega\in[0,\frac{1}{2}]\\ 1, & \omega\in(\frac{1}{2},1]\end{cases}\quad.$$

Find $\sigma(f)$, and check that $\sigma(g)\subset\sigma(f)$. △

4.2.2 Extended random variables

Definition 4.16 (Extended random variable)**.** An extended Borel function X on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ is called an *extended random variable.* The *distribution of* X is the push-forward measure $\mathbb{P}_X$ on $(\overline{\mathbb{R}},\mathcal{B})$ given by

$$\mathbb{P}_X := X_*\mathbb{P}=\mathbb{P}\circ X^{-1},$$

that is,

$$\mathbb{P}_X(A)=\mathbb{P}(X\in A)=\mathbb{P}(\{\omega : X(\omega)\in A\}),\quad A\in\mathcal{B}(\overline{\mathbb{R}}).$$

The function $F_X:\overline{\mathbb{R}}\to[0,1]$ given by

$$F_X(x):=\mathbb{P}_X\big([-\infty,x]\big)$$

is called *the cumulative distribution function of* X.

Remark 4.17. If X never take values $\pm\infty$, we say that X is a *random variable.* △

Remark 4.18. For a random variable X, the measure $\mathbb{P}_X$ on $(\mathbb{R},\mathcal{B})$ is the Lebesgue-Stieltjes measure induced by the function F_X, as defined in §3.2.4. △

Example 4.19. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and A some event in $\mathcal{F}$. Consider the random variable X given by

$$X(\omega)=\mathbb{1}_A(\omega),\ \omega\in\Omega,$$

with distribution

$$\mathbb{P}(X=1)=\mathbb{P}(A)=: p\in[0,1].$$

Then, X is called a *Bernoulli random variable with parameter* p. △

Example 4.20. Let us revisit the construction of Example 4.14. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and for $n\in\mathbb{N}$, let $A_1,\dots,A_n$ be events in $\mathcal{F}$ that partition Ω and $b_1,\dots,b_n$ distinct real numbers. Consider the random variable X defined by

$$X(\omega)=\sum_{i=1}^{n} b_i\mathbb{1}_{A_i}(\omega),\ \omega\in\Omega,$$

with distribution

$$\mathbb{P}(X = b_k) = \mathbb{P}(A_k) =: p_k \in [0,1],\ \ k = 1, \ldots, n.$$

Then, X is called a *simple random variable that takes value b_k with probability p_k*, and has cumulative distribution function

$$F_X(x) = \sum_{i=1}^{n} p_i \mathbb{1}_{\{(-\infty,x]\}}(b_i). \qquad \triangle$$

Example 4.21. Take $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$ and let $\mathbb{P} = m_{|_{[0,1]}}$ be the restriction of the Lebesgue measure to $[0,1]$. Let $X(\omega) = -\log|\omega|$. Then X is measurable (why?), hence an extended random variable. Moreover,

$$F_X(x) = m(\{\omega : -\log|\omega| \leqslant x\} \cap [0,1])) = \begin{cases} 1 - e^{-x}, & x \geqslant 0, \\ 0, & x \leqslant 0, \end{cases}$$

and X is said to follow the *exponential distribution* with unit parameter. $\triangle$

4.3 Operations on extended Borel functions

Given a measurable space $(\Omega, \mathcal{F})$, let $f, g : \Omega \to \overline{\mathbb{R}}$ be extended Borel functions. A property that we constantly use is that measurability is preserved under various operations. The properties studied in this section are so basic that in subsequent chapters we use them without making explicit reference.

Lemma 4.22 (Comparison of functions)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and $f, g : \Omega \to \overline{\mathbb{R}}$ be extended Borel functions. Then, the subsets of Ω given by $\{f < g\}$, $\{f \leqslant g\}$, $\{f = g\}$, and $\{f \neq g\}$ are in $\mathcal{F}$.*

Proof. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, for each $\omega \in \Omega$, $f(\omega) < g(\omega)$ if and only if there is $r \in \mathbb{Q}$ such that $f(\omega) < r < g(\omega)$. Thus, we can express $\{\omega : f(\omega) < g(\omega)\}$ as the union of $\{\omega : f(\omega) < r < g(\omega)\}$ over all $r \in \mathbb{Q}$, and $\mathbb{Q}$ is countable. Further, $\{f < r\}, \{r < g\} \in \mathcal{F}$, by Lemma 4.12 (b) and (d). Putting everything together yields

$$\{f < g\} = \bigcup_{r \in \mathbb{Q}} (\{f < r < g\}) = \bigcup_{r \in \mathbb{Q}} (\{f < r\} \cap \{r < g\}) \in \mathcal{F}. \tag{4.23}$$

Now, we can see that

$$\begin{aligned} &\{f \neq g\} = \{f < g\} \cup \{f > g\} \in \mathcal{F}, \\ &\{f = g\} = \{f \neq g\}^c \in \mathcal{F}, \\ &\{f \leqslant g\} = \{f < g\} \cup \{f = g\} \in \mathcal{F}. \end{aligned}$$

This proves the lemma. $\square$

As one would expect, sums and products of Borel functions f and g on some measurable space $(\Omega, \mathcal{F})$ are also Borel. Importantly, we can extend this to f and g being extended Borel functions, as we only need to impose restrictions on infinite values. In particular, $f+g$ is extended Borel if it is well-defined, that is, if there is no $\omega \in \Omega$ for which $f(\omega)+g(\omega) = \infty - \infty$.

Lemma 4.24 (Sums and products)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and $f, g : \Omega \to \overline{\mathbb{R}}$ be extended Borel functions. Then fg is an extended Borel function. If $f+g$ is well-defined for all ω, then it is an extended Borel function.*

Proof. First, we show that $af+b$ is an extended Borel function for every $a, b \in \mathbb{R}$. For $a = 0$, the proof is trivial, so we assume that $a \in \overline{\mathbb{R}} \setminus \{0\}$. We have that

$$\{af + b < c\} = \{af < c - b\} = \begin{cases} \{f < \frac{c-b}{a}\}, & a > 0 \\ \{f > \frac{c-b}{a}\}, & a < 0 \end{cases}$$

where $\{f < \frac{c-b}{a}\}, \{f > \frac{c-b}{a}\} \in \mathcal{F}$, by Lemma 4.12 (b) and (d).

Now suppose $f+g$ is well-defined for all ω. For every $b \in \mathbb{R}$, it follows that

$$\{f + g < b\} = \{f < b - g\} \in \mathcal{F},$$

since $b - g$ is extended Borel and by applying Lemma 4.22.

Next, we show that fg is extended Borel, where we treat the case of infinite values separately. First we show that f^2 is also extended Borel. We have that

$$\{f^2 > a\} = \begin{cases} \Omega, & a < 0 \\ \{f < -\sqrt{a}\} \cup \{f > \sqrt{a}\}, & a \geqslant 0 \end{cases}$$

where $\Omega \in \mathcal{F}$ and $\{f < -\sqrt{a}\} \cup \{f > \sqrt{a}\} \in \mathcal{F}$, by Lemma 4.12 (b) and (d).

Now suppose f and g are both finite Borel functions. Notice that we can write

$$fg = \frac{1}{4}[(f+g)^2 - (f-g)^2],$$

where $(f+g)^2$ and $(f-g)^2$ are both Borel, since $f+g$ and $f-g$ are, so we can conclude that fg is Borel.

Finally, suppose f and g are both extended Borel functions. Let $A = \{\omega : f(\omega) \in \mathbb{R}, g(\omega) \in \mathbb{R}\}$, and $B = A^c = \{\omega : f(\omega) = \pm\infty \text{ or } g(\omega) = \pm\infty\}$. Note that $A, B \in \mathcal{F}$. Define $f_1 = f\mathbb{1}_A$, $f_2 = f\mathbb{1}_B$, $g_1 = g\mathbb{1}_A$, $g_2 = g\mathbb{1}_B$. Then $f_2 g_2$ is measurable (why?), and by the previous case $f_1 g_1$ is measurable. Since $fg = f_1 g_1 + f_2 g_2$, we conclude that fg is measurable. ☐

Remark 4.25. If f is an extended Borel function on some measurable space $(\Omega, \mathcal{F})$, then any scalar multiple of f is also extended Borel, that is for any $a \in \overline{\mathbb{R}}$, af is an extended Borel function (why?). △

Lemma 4.26 (Maximum and minimum)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $f, g : \Omega \to \overline{\mathbb{R}}$ *be extended Borel functions. Then, we have that* $\max\{f, g\}$ *and* $\min\{f, g\}$ *are extended Borel functions.*

Proof. By taking the suitable inequalities for minimum and maximum, For every $a \in \mathbb{R}$, we have

$$\{\max\{f, g\} < a\} = \{f < a\} \cap \{g < a\} \in \mathcal{F},$$

since $\{f < a\}, \{g < a\} \in \mathcal{F}$ by Lemma 4.12 (b). Finally, use $\min\{f, g\} = -\max\{-f, -g\}$ and Remark 4.25 for the minimum. $\square$

Limiting operations also maintain measurability.

Lemma 4.27 (Suprema and limits of functions)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $(f_n)_{n \geqslant 1}$ *be a sequence of extended Borel functions. Then, the functions* $\sup_{n \geqslant 1} f_n$*,* $\inf_{n \geqslant 1} f_n$*,* $\limsup_{n \to \infty} f_n$*,* $\liminf_{n \to \infty} f_n$ *are extended Borel functions. Moreover, if* $\lim_{n \to \infty} f_n$ *exists, then*

$$\lim_{n \to \infty} f_n = \limsup_{n \to \infty} f_n = \liminf_{n \to \infty} f_n,$$

is an extended Borel function.

Proof. To prove that $\sup_{n \geqslant 1} f_n$ is extended Borel, the idea is similar to the proof of Lemma 4.26. For any $a \in \mathbb{R}$, we can write

$$\left\{\sup_{n \geqslant 1} f_n \leqslant a\right\} = \bigcap_{n=1}^{\infty} \{f_n \leqslant a\} \in \mathcal{F}.$$

where we used the fact that $\mathcal{F}$ is a σ-algebra, so it closed under countable intersections. The claim for $\inf_{n \geqslant 1} f_n$ follows by noticing that

$$\inf_{n \geqslant 1} f_n = -\sup_{n \geqslant 1}(-f_n),$$

and using Remark 4.25. Now, using the definition of limit superior and limit inferior,

$$\limsup_{n \to \infty} f_n = \inf_{n \geqslant 1} \sup_{m \geqslant n} f_m,$$
$$\liminf_{n \to \infty} f_n = \sup_{n \geqslant 1} \inf_{m \geqslant n} f_m,$$

which both are extended Borel, by applying the above properties twice for each limit. Finally, assuming that the limit of f_n exists, we have $\lim_{n \to \infty} f_n = \limsup_{n \to \infty} f_n$, which concludes the proof. $\square$

5 Integration and expectation

In Real Analysis, we first introduce the Riemann Integral, which approximates the area under a curve on a given interval via vertical rectangles and considering a sequence of step functions that converges to the associated curve. However this procedure is not the most suitable to analyse sequences of functions when taking limits. In contrast, the Lebesgue Integral considers horizontal shapes, which makes it more flexible. It is also more descriptive as to when it is possible to take limits. As an example, suppose we want to show that[6]

$$\lim_{n\to\infty} \int_0^\infty \frac{e^{-nx^2}}{\sqrt{x}}\,\mathrm{d}x = 0$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 e^{tx} f(x)\,\mathrm{d}x = \int_0^1 x e^{tx} f(x)\,\mathrm{d}x$$

for some bounded continuous f. Sure, we could do these by brute force, making estimate over estimate. However, these will become much easier if we use the Dominated Convergence Theorem, to be introduced shortly.

> The reason to abandon the Riemann integral and adopt the Lebesgue integral does not stem from the possibility of integrating exotic functions, but rather from the robust and streamlined way in which the Lebesgue integral commutes with limits, derivatives, Fourier series and other integrals.

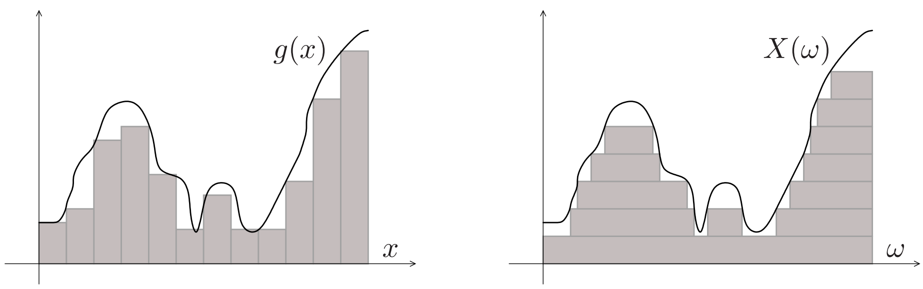


Figure 5.1: Comparison between the Riemann integral on $\mathbb{R}$ and the Lebesgue integral on a measure space.

Another reason to introduce the Lebesgue integral is that some measure spaces (such as probability spaces!) are not naturally partitioned into little intervals or little cubes, but we can still make measurements there. In order to define an integral in this case, instead of partitioning the domain and measuring the height of the graph of a function, we *partition* the co-domain and *measure* chunks of the domain, as shown in Figure 5.1.

[6]Example from [Bar95].

5.1 Axiomatic definition

In this section we define the integral of extended Borel functions by assuming existence of a linear unitary monotone σ-additive operator on non-negative functions.

5.1.1 Non-negative functions

Let $(\Omega, \mathcal{F}, \mu)$ denote a measure space. The integral of non-negative extended Borel functions against a measure μ is denoted by

$$\int_\Omega f \, \mathrm{d}\mu \quad \text{or} \quad \int_\Omega f(\omega) \, \mu(\mathrm{d}\omega)$$

and defined below.

The above two formulas refer to the same operation but are useful in different contexts. The first, rather dry formula, shows a function f on a sample space Ω being integrated against a measure μ. It is very useful in theory for its conciseness, but it only works for functions that can be referenced directly by symbols such as f, αf, $f+g$, $h \circ f$, etc. The second, more versatile, shows a sample space, a measure with a dummy variable as indicated in $\mu(\mathrm{d}\omega)$, and an expression which can be used to define a function by having ω as a free variable. When the expression is $f(\omega)$, this coincides with the first formula, but it also works with expressions such as $\frac{\sin \omega}{e^\omega + 1}$ as shown at the beginning of this chapter. When the measure is the Lebesgue measure, we write $\int f \, \mathrm{d}x$ or $\int f(x) \, \mathrm{d}x$.

> The Lebesgue integral is the *unique* operator with following properties:
>
> - Linear: $\int_\Omega (\alpha f + g) \, \mathrm{d}\mu = \alpha \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu$;
> - Unitary: $\int_\Omega \mathbb{1}_A \, \mathrm{d}\mu = \mu(A)$;
> - Monotone: $f \geqslant g \implies \int_\Omega f \, \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu$;
> - σ-additive: $\int_\Omega (\sum_n f_n) \, \mathrm{d}\mu = \sum_n \int_\Omega f_n \, \mathrm{d}\mu$;
>
> for $\alpha \in [0, +\infty]$, $A \in \mathcal{F}$ and non-negative extended Borel functions f, g, f_n.

We prove existence and uniqueness of this operator in §5.4. Before that, let us discuss properties that follow from the above four.

Exercise 5.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f \geqslant 0$ an extended Borel function. Show that $\int_\Omega f \, \mathrm{d}\mu = 0$ if and only if $\mu(\{f > 0\}) = 0$. △

Solution. Suppose $\mu(\{f > 0\}) > 0$. Then by continuity from below of measures, there exists $\varepsilon > 0$ and $n \in \mathbb{N}$ such that $\mu(\{f > \frac{1}{n}\}) > \varepsilon$ (otherwise we would have $\mu(\{f > 0\}) \leqslant \varepsilon$ for every ε). Let $g(\omega) = \frac{1}{n}$ if $f(\omega) > \frac{1}{n}$ and 0 otherwise. Then $0 \leqslant g \leqslant f$, hence $\int_\Omega f \, \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu > \frac{\varepsilon}{n} > 0$. Now suppose $\mu(\{f > 0\}) = 0$.

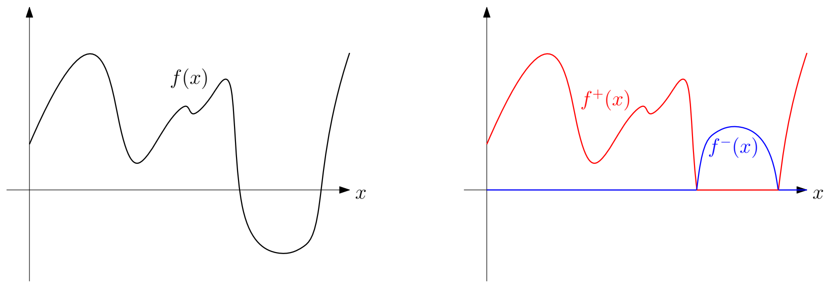


Figure 5.2: Separation of extended Borel function $f : \mathbb{R} \to \mathbb{R}$ by f^+ and f^- (coloured)

.

Define $g(\omega) = +\infty$ if $f(\omega) > 0$ and 0 otherwise. Then $0 \leqslant f \leqslant g$, hence $\int_\Omega f \, \mathrm{d}\mu \leqslant \int_\Omega g \, \mathrm{d}\mu = \infty \cdot 0 = 0$. $\square$

5.1.2 Extended Borel functions and random variables

We now allow f to take values on $[-\infty, +\infty]$, similarly to (1.11). Denote the positive and negative parts of numbers by decomposing

$$x = [x]^+ - [x]^-, \quad [x]^+ = \begin{cases} x, & x \geqslant 0, \\ 0, & x \leqslant 0, \end{cases} \quad [x]^- = \begin{cases} 0, & x \geqslant 0, \\ -x, & x \leqslant 0. \end{cases}$$

For a function f, denote its positive part f^+ and negative part f^- by

$$f^\pm(\omega) = [f(\omega)]^\pm,$$

see Figure 5.2 for an illustration of this decomposition. We then define

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f^+ \, \mathrm{d}\mu - \int_\Omega f^- \, \mathrm{d}\mu. \tag{5.2}$$

unless it gives “ $+\infty - \infty$ ”, in which case $\int_\Omega f \, \mathrm{d}\mu$ is *undefined*. When both integrals are finite, we say that f *is integrable.*

Exercise 5.3. Let f be an extended Borel function on $(\Omega, \mathcal{F}, \mu)$. Prove that f is integrable if and only if $\int_\Omega |f| \, \mathrm{d}\mu < \infty$. $\triangle$

This definition also results in a linear and monotone operator.

Theorem 5.4. *The Lebesgue integral has the following properties:*

- *Homogeneous:* $\int_\Omega (\alpha f) \, \mathrm{d}\mu = \alpha \int_\Omega f \, \mathrm{d}\mu$
- *Additive:* $\int_\Omega (f + g) \, \mathrm{d}\mu = (\int_\Omega f \, \mathrm{d}\mu) + (\int_\Omega g \, \mathrm{d}\mu)$
- *Monotone:* $f \geqslant g \implies \int_\Omega f \, \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu$.

for measurable f, g and $\alpha \in [-\infty, +\infty]$ as long as the integrals are defined.

These properties apply as follows. Homogeneity holds if both integrals are defined. Linearity is true if both integrals are defined and their sum does not result in "$+\infty - \infty$." Monotonicity means that, if $\int f^+ \, \mathrm{d}\mu < \infty$ or $\int g^- \, \mathrm{d}\mu < \infty$, then both integrals are defined and satisfy the inequality.

We prove these properties from (5.2) in §5.4.

Exercise 5.5. Let f be a non-negative extended Borel function on $(\Omega, \mathcal{F}, \mu)$. Show that $\int_\Omega f \, \mathrm{d}\mu$ is defined and non-negative. △

Exercise 5.6. Let f be an extended Borel function on $(\Omega, \mathcal{F}, \mu)$. Suppose that $\int_\Omega f \, \mathrm{d}\mu$ is defined. Show that $|\int_\Omega f \, \mathrm{d}\mu| \leqslant \int_\Omega |f| \, \mathrm{d}\mu$. △

Exercise 5.7. Let f, g be extended Borel functions on $(\Omega, \mathcal{F}, \mu)$. Suppose that $|f| \leqslant g$ and g is integrable. Show that f is integrable. △

Exercise 5.8. Let f be an extended Borel function on $(\Omega, \mathcal{F}, \mu)$ and $A \in \mathcal{F}$ be a measurable set. Suppose that $\int_\Omega f \, \mathrm{d}\mu$ is defined. Show that $\int_\Omega f \mathbb{1}_A \, \mathrm{d}\mu$ is defined. △

Exercise 5.9. Let f be an extended Borel function on $(\Omega, \mathcal{F}, \mu)$ and $A \in \mathcal{F}$ be a measurable set. Suppose that f is integrable. Show that $f \mathbb{1}_A$ is integrable. △

Exercise 5.10. Show that the family of integrable functions on a given measure space $(\Omega, \mathcal{F}, \mu)$ is a vector space over $\mathbb{R}$. △

We now give a precise and unified definition of expectation.

Definition 5.11 (Expectation)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For an extended random variable X, the *expectation of* X is defined as

$$\mathbb{E}X = \int_\Omega X \, \mathrm{d}\mathbb{P}. \tag{5.12}$$

Exercise 5.13. Let X be an extended random variable such that $a \leqslant X \leqslant b$ for all $\omega \in \Omega$. Prove that $a \leqslant \mathbb{E}X \leqslant b$. △

Exercise 5.14. Supposed that X is an integrable extended random variable. Prove that $\mathbb{E}[X - \mathbb{E}X] = 0$. △

Remark 5.15. Notation $\mathbb{E}X$ is used when the measure $\mathbb{P}$ is arbitrary or clear form the context. If more than one measure are being considered, one may use $\mathbf{E}$ and $\tilde{\mathbb{E}}$ for the integrals against $\mathbf{P}$ and $\tilde{\mathbb{P}}$, etc. If we need even more measures than fonts and tildes allow, it may be better to write $\int_\Omega X \, \mathrm{d}\mathbb{P}_\alpha$ instead. △

5.2 Main properties of the integral

In this section we present some general properties of the integral, some inequalities, conditions for an extended Borel function to be integrable, and the relation with the Riemann integral.

5.2.1 General properties

Given a measurable set $B \in \mathcal{F}$, the integral of f on B is given by the integral of the function $f\mathbb{1}_B$ which coincides with f on B and is zero outside B:

$$\int_B f \, \mathrm{d}\mu := \int_\Omega (f\mathbb{1}_B) \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}(\mu_{|_B}) \tag{5.16}$$

The first equality above is the definition of the integral on a subset whenever the second integral is defined. The second equality says that the integral of f on B coincides with the integral of f against the measure μ restricted to B, as defined in Example 2.33, and is valid whenever one of the two is defined.

Preview of proof. We assume that f only takes values on $\{0, 1\}$ (in §5.4 we will see a standard procedure to bootstrap arguments for this type of function to general functions). So $f = \mathbb{1}_A$ for some $A \in \mathcal{F}$, and

$$\int_\Omega (f\mathbb{1}_B) \, \mathrm{d}\mu = \int_\Omega \mathbb{1}_{A\cap B} \, \mathrm{d}\mu = \mu(A \cap B) = \mu_{|_B}(A) = \int_\Omega f \, \mathrm{d}(\mu_{|_B}) \qquad \square$$

All this sounds great but at some point we will want to actually *compute* an integral. There are many methods, but the most important one is the Fundamental Theorem of Calculus.

Theorem 5.17 (Fundamental Theorem of Calculus)**.** *If F is continuous on $[a, b]$ and $F' = f$ on (a, b), then*

$$\int_{[a,b]} f \, \mathrm{d}x = F(b) - F(a),$$

where $\mathrm{d}x$ *denotes the Lebesgue measure.*

Proof. It is is the same as the proof for the Riemann integral. The latter uses the fact that the integral is monotone, finitely additive, and that $\int_{[a,b]} 1\mathrm{d}x = b - a$ for all $a < b$. All these three properties are satisfied by the Lebesgue integral against the Lebesgue measure, hence the Fundamental Theorem of Calculus also holds in this context. $\square$

Example 5.18. $\int_0^\pi \sin x \, \mathrm{d}x = -\cos \pi + \cos 0 = 2.$ $\triangle$

Now, going one step further in the diagram of (4.7), if we have an extended Borel function defined on Ω_2, we can define its *pull-back* on Ω_1 by composition:

$$\begin{array}{ccccc} \Omega_1 & \xrightarrow{f} & \Omega_2 & \xrightarrow{g} & \mathbb{R} \\ \omega & \mapsto & \omega' & \mapsto & x \end{array}.$$

Defining the function $f^*g = g \circ f$ on Ω_1 and get

$$\begin{tikzcd}
\Omega_1 \arrow[r, dotted, no head] \arrow[d, "f"] & \sigma(f) \arrow[r, dotted, no head] & \mu \arrow[r, dotted, no head] \arrow[d, "\circ f^{-1}"] & f^*g \\
\Omega_2 \arrow[r, dotted, no head] & \mathcal{F}_2 \arrow[r, dotted, no head] \arrow[u, "f^{-1}"] & f_*\mu \arrow[r, dotted, no head] & g \arrow[u, "\circ f"]
\end{tikzcd}$$

In summary, from a function that takes points $\omega \in \Omega_1$ to points $\omega' \in \Omega_2$, we can pull a σ-algebra on Ω_2 back to a σ-algebra on Ω_1, then push a measure on Ω_1 forward to a measure on Ω_2, and pull an observable defined on Ω_2 back to an observable defined on Ω_1.

Theorem 5.19 (Change of variable)**.** *For measurable functions $f : \Omega_1 \to \Omega_2$ and $g : \Omega_2 \to \mathbb{R}$, we have*

$$\int_{\Omega_2} g \, \mathrm{d}(f_*\mu) = \int_{\Omega_1} (f^*g) \, \mathrm{d}\mu,$$

provided one of the two is defined.

Remark 5.20 (Expectation). In practice, we do not compute an expectation by (5.12) but instead

$$\mathbb{E}X = \int_{\mathbb{R}} x \, \mathbb{P}_X(\mathrm{d}x), \tag{5.21}$$

which is a particular case of the above theorem with $f = X$ and $g(x) = x$. △

Remark 5.22 (u substitution). A familiar instance of this principle is when f maps "x" to "u" and it becomes

$$\int_A g(f(x)) f'(x) \, \mathrm{d}x = \int_B g(u) \, \mathrm{d}u,$$

where $A = f^{-1}(B)$ and $\mathrm{d}u = f_*\mathrm{d}x = f'(x)\mathrm{d}x$ in a sense that will be made precise in §6. △

Preview of proof. We assume that g only takes values on $\{0, 1\}$ and leave the complete statement for §5.4. So $g = \mathbb{1}_C$ for some $C \in \mathcal{F}_2$, and $f^*g = \mathbb{1}_D$ where $D = f^{-1}(C) \in \mathcal{F}_1$. Substituting these identities, we get

$$\int_{\Omega_2} g \, \mathrm{d}(f_*\mu) = (f_*\mu)(C) = \mu(f^{-1}(C)) = \mu(D) = \int_{\Omega_1} (f^*g) \, \mathrm{d}\mu. \qquad \square$$

We say that a property holds μ-*almost everywhere*, or for μ-*almost every* ω, abbreviated by μ-a.e., if there is a set $A \in \mathcal{F}$ such that $\mu(A^c) = 0$ and such that this property holds for all $\omega \in A$. In case μ is clear from the context, we may simply say "a.e." omitting μ. In case of a probability measure, it is customary to say *a.s.* or *almost surely* instead.

Proposition 5.23 (a.e. equal functions)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g : \Omega \to \mathbb{R}$ extended Borel functions. If $f = g$ μ-a.e., then*

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu,$$

provided one of the two is defined.

Proof. Set $E := \{\omega : f(\omega) = g(\omega)\} \in \mathcal{F}$. Define $f_1 = f\mathbb{1}_E$ and $f_2 = f_1 + \infty \cdot \mathbb{1}_{E^c}$. Then

$$0 \leqslant f_1^+ \leqslant f^+ \leqslant f_2^+ \text{ and } 0 \leqslant f_1^+ \leqslant g^+ \leqslant f_2^+.$$

Since $\int_\Omega f_2 \,\mathrm{d}\mu = \int_\Omega f_1 \,\mathrm{d}\mu + \infty \cdot \mu(E^c) = \int_\Omega f_1 \,\mathrm{d}\mu$, by monotonicity we get

$$\int_\Omega f^+ \,\mathrm{d}\mu = \int_\Omega g^+ \,\mathrm{d}\mu.$$

By a similar argument, $\int_\Omega f^- \,\mathrm{d}\mu = \int_\Omega g^- \,\mathrm{d}\mu$. Assuming one of these two is finite, we get $\int_\Omega f \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu$. □

Remark 5.24. Most of measure theory is insensitive to whether two functions differ on a set of zero measure. In upcoming sections and chapters, the reader will see that two a.e. equal functions are basically the same function for (almost!) all practical purposes. △

Exercise 5.25. Let f be an extended Borel function on a measure space $(\Omega, \mathcal{F}, \mu)$ such that $\mu(f \neq 0) = 0$. Show that $\int_\Omega f \,\mathrm{d}\mu = 0$. △

5.2.2 Inequalities

The following inequality is fundamental for many arguments in Measure Theory and Probability.

Theorem 5.26 (Chebyshev's inequality)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \overline{\mathbb{R}}$ an extended Borel function. Then, for all $\varepsilon > 0$ and $\alpha > 0$,*

$$\mu(\{|f| \geqslant \varepsilon\}) \leqslant \frac{1}{\varepsilon^\alpha} \int_\Omega |f|^\alpha \,\mathrm{d}\mu.$$

Proof. Let $\varepsilon \geqslant 0$ and $\alpha \geqslant 0$. Then we have that,

$$\int_\Omega |f|^\alpha \,\mathrm{d}\mu \geqslant \int_{\{|f| \geqslant \varepsilon\}} |f|^\alpha \,\mathrm{d}\mu \geqslant \int_{\{|f| \geqslant \varepsilon\}} \varepsilon^\alpha \,\mathrm{d}\mu =$$

$$= \varepsilon^\alpha \int_\Omega \mathbb{1}_{\{|f| \geqslant \varepsilon\}} \,\mathrm{d}\mu = \varepsilon^\alpha \mu(\{|f| \geqslant \varepsilon\}),$$

and by rearranging, the inequality follows. ☐

Remark 5.27. In the setting of $(\Omega, \mathcal{F}, \mathbb{P})$ being a probability space and $X : \Omega \to \overline{\mathbb{R}}$ an extended random variable, $\forall \varepsilon \geqslant 0$ and $\alpha \geqslant 0$,

$$\mathbb{P}(|X| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^\alpha} \mathbb{E}|X|^\alpha,$$

which is sometimes referred to as *Markov's inequality*. We can, in turn, bound deviation from the mean using the variance of X by,

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} \mathbb{V}X. \tag{5.28}$$

This is sometimes also referred to as Chebyshev's inequality. △

Exercise 5.29. Let X be a random variable such that, almost surely, $a \leqslant X \leqslant b$. Suppose X is not almost-surely equal to a constant. Prove that $a < \mathbb{E}X < b$.

[Note: this exercise should have appeared in §5.1, I am keeping it here for now to avoid messing up with the numbers.] △

For the next inequality, we recall the notion of a *convex function*. For an open interval $I \subset \mathbb{R}$, a function $f : I \to \mathbb{R}$ is called *convex* if for every $x, y \in I$ and $a, b \in [0, 1]$ with $a + b = 1$, we have

$$f(ax + by) \leqslant af(x) + bf(y).$$

This inequality means that any line segment joining two points in the graph of f stay above the graph. In case f is twice differentiable, it is convex if and only if $f''(x) \geqslant 0$ for every $x \in I$. Examples of convex functions are $|x|$, e^x, $ax + b$, x^2 and $[x]^+$ on $\mathbb{R}$, as well as x^{-2}, $-\log x$, $-\arctan x$ and x^{-1} on $(0, +\infty)$.

Theorem 5.30 (Jensen's inequality)**.** *Let X be an integrable random variable taking values on an open interval I, and $f : I \to \mathbb{R}$ a convex function. Then,*

$$f(\mathbb{E}X) \leqslant \mathbb{E}f(X).$$

Proof. Since f is convex, by [Kle14, Corollary 7.8(ii)], for each $x_0 \in I$, there is $c \in \mathbb{R}$ such that

$$f(x) \geqslant f(x_0) + c(x - x_0), \ \forall x \in I.$$

Choosing $x = X$ and $x_0 = \mathbb{E}X$, we get

$$f(X) \geqslant f(\mathbb{E}X) + c \cdot (X - \mathbb{E}X), \ \text{a.s.},$$

and taking expectations on both sides we get the stated inequality.[7] ☐

[7]Based on [Shi16, §2.6.7].

Theorem 5.31 (Cauchy-Schwarz Inequality)**.** *If* $\mathbb{E}X^2 < \infty$ *and* $\mathbb{E}Y^2 < \infty$, *then* XY *is integrable and*

$$\mathbb{E}[XY] \leqslant \sqrt{\mathbb{E}X^2}\,\sqrt{\mathbb{E}Y^2}.$$

Proof. We can assume $\mathbb{E}X^2 = 1$ and $\mathbb{E}Y^2 = 1$ (why?). The inequality follows immediately from the trick

$$0 \leqslant \mathbb{E}(X-Y)^2 = 2 - 2\mathbb{E}[XY]. \qquad \square$$

Example 5.33. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X = \mathbb{1}_A, Y = \mathbb{1}_B$ be Bernoulli random variables with parameter p, where A, B are events in $\mathcal{F}$. Further, assume that A and B are independent, i.e.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = p^2.$$

Then, we can see that

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{1}_A \mathbb{1}_B] = \mathbb{E}[\mathbb{1}_{A\cap B}] = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = p^2.$$

On the other hand, we have

$$\sqrt{\mathbb{E}X^2}\,\sqrt{\mathbb{E}Y^2} = \sqrt{\mathbb{E}X}\,\sqrt{\mathbb{E}Y} = p,$$

and since $p \in (0,1)$ by construction, Cauchy-Schwarz Inequality is verified. $\triangle$

5.2.3 Integrability

Proposition 5.34. *Suppose there is a countable set* $A \subseteq [0, +\infty]$ *such that* X *only takes values in* A. *Then*

$$\mathbb{E}X = \sum_{x\in A} x\mathbb{P}(X = x).$$

Proof. Using σ-additivity of the expectation,

$$\mathbb{E}X = \mathbb{E}\Big[\sum_{x\in A} x \cdot \mathbb{1}_{[X=x]}\Big] = \sum_x x\mathbb{E}\mathbb{1}_{[X=x]} = \sum_x x\mathbb{P}(X = x). \qquad \square$$

Example 5.35. If $\mathbb{P}(X = n) = \frac{\lambda^n e^{-\lambda}}{n!}$ for $n \in \mathbb{N}_0$, then

$$\begin{aligned} EX = \sum_{n=0}^{\infty} n\frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} = \\ = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda, \end{aligned}$$

giving the familiar formula for the expectation of a Poisson variable. $\triangle$

Proposition 5.36. *If X only takes values on $\{0, 1, 2, 3, \dots, +\infty\}$, then*

$$\mathbb{E}X = \sum_{n \in \mathbb{N}} \mathbb{P}(X \geqslant n).$$

Proof. Let $A = \{0, 1, 2, 3, \dots, +\infty\}$. Replacing n by a dull sum over k,

$$\begin{aligned}\mathbb{E}X = \sum_{n\in A} n\mathbb{P}(X = n) &= \sum_{n\in A}\sum_{k\in\mathbb{N}} \mathbb{1}_{k\leqslant n}\mathbb{P}(X = n) = \\ &= \sum_{k\in\mathbb{N}}\sum_{n\in A} \mathbb{1}_{n\geqslant k}\mathbb{P}(X = n) = \sum_{k\in\mathbb{N}}\sum_{n\geqslant k} \mathbb{P}(X = n) = \sum_{k\in\mathbb{N}} \mathbb{P}(X \geqslant k),\end{aligned}$$

where Theorem 1.13 was used to interchange sums. □

Example 5.37. If $\mathbb{P}(X \in \mathbb{N}_0) = 1$ and $\mathbb{P}(X \geqslant n) = (1-p)^{n-1}$ for all $n \in \mathbb{N}$, where $0 < p < 1$, then

$$\mathbb{E}X = \sum_{n=1}^{\infty} P(X \geqslant n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \sum_{j=0}^{\infty} (1-p)^j = \frac{1}{1-(1-p)} = \frac{1}{p},$$

giving the familiar formula for the expectation of a Geometric variable. △

Remark 5.38. The above propositions and their proof also work for measure spaces with X replaced by f and $\mathbb{E}$ by $\int_\Omega \cdots \, \mathrm{d}\mu$. △

Proposition 5.39 (Criterion for integrability)**.** *A random variable X is integrable if and only if*

$$\sum_n \mathbb{P}(|X| \geqslant n) < \infty.$$

Above we do not specify whether $n = 0$ or $n = \infty$ are included in the sum, because all that matters is whether the sum is finite.

Proof. Let $Z = |X|$ and $Y = \lfloor Z \rfloor$. Observe that

$$\mathbb{E}Y = \sum_n \mathbb{P}(Y \geqslant n) = \sum_n \mathbb{P}(Z \geqslant n).$$

On the other hand, $Y \leqslant Z \leqslant Y+1$. By additivity, we have $\mathbb{E}[Y+1] = (\mathbb{E}Y)+1$. By monotonicity, $\mathbb{E}Y \leqslant \mathbb{E}Z \leqslant (\mathbb{E}Y)+1$. If one of these numbers is finite, all the others are finite, and if one is infinite, all the others are infinite (why?). This concludes the proof. □

Exercise 5.40. Can you find which parts of the proof break down if we work on a measure space instead of probability space? Can you find counter-examples? △

5.2.4 Relation to the Riemann integral

We conclude this section with a comparison with the Riemann integral.

Theorem 5.41. *Let $f : [a, b] \to \mathbb{R}$ be a Borel function. If f is Riemann integrable then it is Lebesgue integrable and the value of both integrals coincide.*

Proof. Using the definition from [Coh13, §2.5], the Riemann integral of f on $[a, b]$ equals $L \in \mathbb{R}$ if, for every $\varepsilon > 0$, there exist step functions g and h such that $g \leqslant f \leqslant h$ and $L - \varepsilon < \int_{[a,b]} g \, \mathrm{d}x \leqslant \int_{[a,b]} h \, \mathrm{d}x < L + \varepsilon$. But this implies that g and h (and hence f) are bounded, therefore $\int_{[a,b]} f \, \mathrm{d}x$ is defined and $| \int_{[a,b]} f \, \mathrm{d}x - L| < \varepsilon$. Since this holds for every $\varepsilon > 0$, we get $\int_{[a,b]} f \, \mathrm{d}x = L$. $\square$

Remark 5.42. The converse is false. A function can be Lebesgue integrable and not Riemann integrable. In case f is unbounded, it is not Riemann integrable even if it is continuous at almost every point.

Another popular example is $f : [0, 1] \to \mathbb{R}$ given by $f = \mathbb{1}_{\mathbb{Q}}$. Since $m(\mathbb{Q}) = 0$, we have $\int_{[0,1]} f \, \mathrm{d}x = 0$ but this function is not Riemann integrable.[8] $\triangle$

Remark 5.43. The reader may wonder what happens if f is Riemann integrable but not Borel measurable. We will never bump into such a function unless we are looking for it. In any case, if $f : [a, b] \to \mathbb{R}$ is Riemann integrable then it is measurable with respect to a σ-algebra larger than $\mathcal{B}$, denoted $\mathcal{F}^*$ in Lemma 3.12, it is Lebesgue integrable and again the value of both integrals coincide. See [Coh13, 2.5.4] for a proof. $\triangle$

5.2.5 The fallacy of improper integrals

It is often claimed that there are functions which are Riemann integrable but not Lebesgue integrable. The infamous example is

$$f(x) = \frac{\sin x}{x} \cdot \mathbb{1}_{[\pi,+\infty)}(x),$$

usually followed by the pompous claim that it is Riemann integrable.

The reason why this function causes some commotion is that f is not Lebesgue integrable (proof omitted) but nevertheless the limit

$$\lim_{z \to +\infty} \int_{[\pi,z]} f(x) \, \mathrm{d}x$$

[8] Let us prove that it is not Riemann integrable. Consider any partition of $[0, 1]$ into finitely many non-degenerate intervals, and notice that all intervals will contain both rational and irrational numbers. Thus, every step function $g \leqslant f$ must satisfy $g \leqslant 0$ a.e., and every step function $h \geqslant f$ must satisfy $h \geqslant 1$ a.e. Hence, $\int g \, \mathrm{d}x \leqslant 0$ and $\int h \, \mathrm{d}x \geqslant 1$ and taking $\varepsilon = \frac{1}{3}$ we see that there is no number L such that $L - \varepsilon < \int g \, \mathrm{d}x \leqslant \int h \, \mathrm{d}x < L + \varepsilon$.

is finite (proof omitted). Some would infer form this that f is Riemann integrable but not Lebesgue integrable, without realising that the above limit is the same regardless of whether the integral on $[\pi, z]$ is Lebesgue or Riemann.

Anyway, there are some serious problems with the *interpretation* of this limit. It is finite, so what? As far as Measure Theory and Probability Theory are concerned, this limit has no meaning. It takes *one particular* sequence of sets A_n satisfying $A_n \uparrow [\pi, +\infty)$. By picking another sequence with the same property, one can make this limit give *any* number in $\overline{\mathbb{R}}$ (proof omitted). There is no physical interpretation of this limit in terms of signed areas, centre of mass, etc. If this limit were to be understood as the value of the integral, Theorem 5.19 would not hold. If it were to correspond to the expectation of a random variable, the law of large numbers would not hold either.

5.3 Convergence of integral and applications

We can think of the region under the graph of a non-negative extended real function as having an area, volume or mass. If the function is $f(x) = n \cdot \mathbb{1}_{(0,\frac{1}{n}]}(x)$, or $g(x) = \mathbb{1}_{(n,n+1]}(x)$, this mass is always equals to 1 and, nevertheless, it disappears when we let $n \to \infty$. In both cases, we can say that the mass "escaped to infinity." In the former case it disappeared vertically and in the latter case, horizontally. The three properties studied in this section explain what can happen to this mass when we take limits.

5.3.1 The main theorems

The Monotone Convergence Theorem says that nothing strange can happen with the mass of an increasing sequence. More precisely, it says that extra mass cannot appear in the limit.

Theorem 5.44 (Monotone Convergence Theorem)**.** *Let $(f_n)_n$ be a sequence of non-negative extended Borel functions on $(\Omega, \mathcal{F}, \mu)$ such that $f_{n+1}(\omega) \geqslant f_n(\omega)$ for all $n \in \mathbb{N}$ and $\omega \in \Omega$. Then*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\mu(\mathrm{d}\omega) = \int_\Omega (\lim_{n\to\infty} f_n(\omega))\mu(\mathrm{d}\omega).$$

An easier way to remember is: $0 \leqslant f_n \uparrow f \implies \int f_n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu$.

Proof. We included σ-additivity as an axiom instead of monotone convergence for aesthetic reasons, so the proof is a bit dull. Write $f_n = g_1 + \cdots + g_n$. Then $\sum_n g_n = f$, and by σ-additivity we have

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega (\textstyle\sum_n g_n) \,\mathrm{d}\mu = \textstyle\sum_n \displaystyle\int_\Omega g_n \,\mathrm{d}\mu = \lim_n \int_\Omega f_n \,\mathrm{d}\mu.$$

In §5.4 we will re-prove this theorem in order to prove σ-additivity. $\square$

From this theorem, we derive another fundamental property relating integrals and limits. Fatou's lemma says that, even though the previous examples show that even though we can lose mass in a limit, we can never gain mass.

> **Theorem 5.45** (Fatou's lemma). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n\geqslant 1}$ a sequence of non-negative extended Borel functions. Then,*
>
> $$\int_\Omega (\liminf_{n\to\infty} f_n(\omega))\, \mu(\mathrm{d}\omega) \leqslant \liminf_{n\to\infty} \int_\Omega f_n(\omega)\, \mu(\mathrm{d}\omega).$$
>
> To remember the inequality, we consider functions $f_n = \mathbb{1}_{[n,\infty)}$.

Proof. Taking

$$g_n(\omega) = \inf_{k\geqslant n} f_k(\omega)$$

and defining

$$g(\omega) = \liminf_{n\to\infty} f_n(\omega),$$

we have that $0 \leqslant g_n \uparrow g$. By the Monotone Convergence Theorem,

$$\lim_{n\to\infty} \int_\Omega g_n \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu.$$

Since $g_n \leqslant f_n$, we get

$$\liminf_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \geqslant \liminf_{n\to\infty} \int_\Omega g_n \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu = \int_\Omega (\liminf_{n\to\infty} f_n) \,\mathrm{d}\mu,$$

concluding the proof. $\square$

The Dominated Convergence Theorem says that, if the graphs of a sequence f_n of functions are confined in a region of finite mass, then we cannot lose mass in the limit either. The reason is that the graph of f_n divides this region of finite mass in two parts, and the fact that each of the two parts cannot gain mass in the limit implies that the other one cannot lose it.

> **Theorem 5.46** (Dominated Convergence Theorem). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n\geqslant 1}$ a sequence of extended Borel functions such that $f_n(\omega) \to f(\omega)$ as $n \to \infty$ for a.e. ω. If there exists a non-negative extended Borel function g that is μ-integrable, such that $|f_n| \leqslant g$ for all $n \geqslant 1$, then f is μ-integrable, and*
>
> $$\lim_{n\to\infty} \int_\Omega f_n(\omega)\mu(\mathrm{d}\omega) = \int_\Omega \Big(\lim_{n\to\infty} f_n(\omega)\Big)\mu(\mathrm{d}\omega).$$

Proof. By changing f_n, f, g on a set of measure zero, we can assume that convergence holds for every ω. Notice that $f_n + g \geqslant 0$ for all $n \geqslant 1$, so by Fatou's lemma, we get that

$$\int_\Omega (f+g)\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_\Omega (f_n+g)\,\mathrm{d}\mu$$

and hence

$$\int_\Omega f\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu.$$

Similarly, we have that $-f_n + g \geqslant 0$ for all $n \geqslant 1$, which gives

$$\int_\Omega (-f+g)\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_\Omega (-f_n+g)\,\mathrm{d}\mu$$

and hence

$$\int_\Omega f\,\mathrm{d}\mu \geqslant \limsup_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu.$$

From these two inequalities, we get

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

5.3.2 Some corollaries and applications

Corollary 5.47 (Bounded Convergence Theorem)**.** *If $X_n \to X$ a.s. and there is $M \in \mathbb{R}$ such that $|X_n| \leq M$ a.s. for all n, then $\mathbb{E}X_n \to \mathbb{E}X$.*

Proof. Apply the Dominated Convergence Theorem with constant $g = M$. $\square$

Corollary 5.48. *Suppose $0 \leqslant f_n \uparrow f$, and there is $M \in \mathbb{R}$ such that $\int_\Omega f_n\,\mathrm{d}\mu < M$ for all n. Then $f(\omega) < \infty$ for a.e. ω.*

Proof. Exercise. $\square$

Proposition 5.49 (Moment generating function)**.** *Let $M(t) = \mathbb{E}e^{tX}$ and suppose there exists $\varepsilon > 0$ such that $M(t) < \infty$ for $-\varepsilon < t < \varepsilon$. Then*

$$M^{(k)}(0) = \mathbb{E}X^k$$

for $k = 0, 1, 2, 3, \dots$.

Proof. We will show a stronger statement, namely that

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t) = \mathbb{E}[X^k e^{tX}]$$

for every $t \in (-\varepsilon, +\varepsilon)$, by induction on k.

For $k = 0$ we have the identity $M_X(t) = \mathbb{E}e^{tX}$ which holds trivially. Suppose the identity holds for some $k \in \mathbb{N}$. Write $g_x(t) = x^k e^{tx}$ and $g_x'(t) = x^{k+1}e^{tx}$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}M_X^{(k)}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}g_X(t) = \lim_{h\to 0}\mathbb{E}\frac{g_X(t+h)-g_X(t)}{h}.$$

If we can commute the expectation and the limit, we get

$$M_X^{(k+1)}(t) = \mathbb{E}\lim_{h\to 0}\frac{g_X(t+h)-g_X(t)}{h} = \mathbb{E}g_X'(t) = \mathbb{E}[X^{k+1}e^{tX}].$$

Fix a sequence $h_n \to 0$. In order to apply the Dominated Convergence Theorem, it is enough to bound the random term $|\frac{g_X(t+h)-g_X(t)}{h}|$ by an integrable extended random variable. By the Mean Value Theorem,

$$\frac{g_x(t+h)-g_x(t)}{h} = g_x'(\theta),$$

where $\theta \in [t, t+h]$ depends on x, t and h. Taking $\delta = \frac{\varepsilon - |t|}{3}$, for $|h| < \delta$ we have

$$|g_x'(\theta)| \leq |x|^{k+1}e^{(\varepsilon-2\delta)|x|} \leqslant Ce^{(\varepsilon-\delta)|x|} \quad \text{for every } x \in \mathbb{R},$$

where C depends on ε and t. It follows from the assumption that $\mathbb{E}e^{(\varepsilon-\delta)|X|} < \infty$, which concludes the proof. □

5.4 Construction of the integral

In this last section we prove the existence and uniqueness of the operator $f \mapsto \int_\Omega f\,\mathrm{d}\mu$ defined on non-negative extended Borel functions f. In this process, we will show that non-negative functions can be approximated by simple functions, which allows us to bootstrap properties proved for binary functions (such as Theorem 5.19) to general measurable functions.

5.4.1 Simple functions

We say that f is a *simple function* if it is measurable and only takes finitely many values. In case f is also non-negative, we define its integral as

$$\int_\Omega f\,\mathrm{d}\mu := \sum_x x \cdot \mu(f = x),$$

where "$\mu(f = x)$" means $\mu(\{\omega \in \Omega : f(\omega) = x\})$. Let $a_1, \ldots, a_n$ denote the distinct values attained by f, and $A_1, \ldots, A_n \in \mathcal{F}$ the sets where f takes these values. Then the sets $A_1, \ldots, A_n$ form a partition of Ω and

$$f = \sum_{j=1}^n a_j \mathbb{1}_{A_j}.$$

This representation is unique except for permutation of the indices $\{1, \dots, n\}$. In this representation, we have

$$\int_\Omega f \, \mathrm{d}\mu = \sum_{j=1}^{n} a_j \mu(A_j) \tag{5.50}$$

for every f non-negative and simple.

Lemma 5.51. *This definition is unitary, monotone, and linear.*

It is unitary by construction. Monotonicity follows directly from linearity. Indeed, if $f \geqslant g \geqslant 0$ are simple, then $h := f - g$ is non-negative and simple, and $\int_\Omega f \, \mathrm{d}\mu = \int_\Omega g \, \mathrm{d}\mu + \int_\Omega h \, \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu$. So it remains to prove linearity.

Lemma 5.52. *If $A_1, \dots, A_n \in \mathcal{F}$ form a partition of Ω and $f = \sum_{j=1}^{n} a_j \mathbb{1}_{A_j}$, then $\int_\Omega f \, \mathrm{d}\mu = \sum_{j=1}^{n} a_j \mu(A_j)$, even if $a_1, \dots, a_n \in [0, +\infty]$ are not distinct.*

Idea of proof. Group the sets A corresponding to same value a. □

Proof of linearity. Write $f = \sum_{j=1}^{n} a_j \mathbb{1}_{A_j}$ and $g = \sum_{i=1}^{k} b_i \mathbb{1}_{B_i}$, where both $A_1, \dots, A_n$ and $B_1, \dots, B_k$ form partitions of Ω. After expanding and grouping,

$$af + bg = \cdots = \sum_{j=1}^{n} \sum_{i=1}^{k} (a a_j + b b_i) \mathbb{1}_{A_j \cap B_i}.$$

But the collection $\{A_j \cap B_i\}_{j=1,\dots,n; i=1,\dots,k}$ forms a partition of Ω, so using the previous lemma we get

$$\int_\Omega (af + bg) \, \mathrm{d}\mu = \sum_{j=1}^{n} \sum_{i=1}^{k} (a a_j + b b_i) \mu(A_j \cap B_i).$$

Splitting the terms in the sum and using that $B_1, \dots, B_k$ as well as $A_1, \dots, A_n$ are partitions, one can eventually gets

$$\int_\Omega (af + bg) \, \mathrm{d}\mu = \sum_{j=1}^{n} a a_j \mu(A_j) + \sum_{i=1}^{k} b b_i \mu(B_i) = a \int_\Omega f \, \mathrm{d}\mu + b \int_\Omega g \, \mathrm{d}\mu. \quad \square$$

Remark 5.53. We really had no choice for how to define the integral of simple functions, as (5.50) follows from the integral being unitary and linear. △

5.4.2 Non-negative extended Borel functions

We define the integral of a non-negative extended Borel function f by

$$\int_\Omega f \, \mathrm{d}\mu = \sup \Big\{ \textstyle\int_\Omega g \, \mathrm{d}\mu : 0 \leqslant g \leqslant f \text{ and } g \text{ is simple} \Big\}. \tag{5.54}$$

As promised in §5.1, we will now prove the following.

Theorem 5.55. *This definition is linear, unitary, monotone and σ-additive.*

It is to be noted that for simple non-negative functions, this definition coincides with the previous one, so it really is only extending it. In particular, this integral is unitary by construction. Monotonicity follows from inclusion: the larger f, the more simple functions are allowed in the above supremum.

To prove linearity and σ-additivity, we will use monotone convergence:

$$0 \leqslant f_n \uparrow f \implies \int_\Omega f_n \,\mathrm{d}\mu \to \int_\Omega f \,\mathrm{d}\mu \tag{5.56}$$

for every non-decreasing sequence of non-negative extended Borel functions. This was already "proved" using σ-additivity, but in order to avoid a circular argument we now prove it directly from the above definition.

Proof of (5.56). By monotonicity, the sequence $\int_\Omega f_n \,\mathrm{d}\mu$ is non-decreasing and bounded by $\int_\Omega f \,\mathrm{d}\mu$. Hence, it converges, and

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \leqslant \int_\Omega f \,\mathrm{d}\mu.$$

It remains to show the opposite inequality. Let $0 \leqslant g \leqslant f$ be simple, taking positive values $a_1, \dots, a_k$. Let $0 < \alpha < 1$ and $A_n = \{\omega : f_n(\omega) > \alpha g(\omega)\}$. Since $f \geqslant g$ and $0 \leqslant f_n \uparrow f$, for each ω there is n such that $f_n(\omega) \geqslant \alpha g(\omega)$, so $A_n \uparrow \Omega$ and

$$\begin{aligned}\int_\Omega f_n \,\mathrm{d}\mu \geqslant \int_\Omega (f_n \mathbb{1}_{A_n}) \,\mathrm{d}\mu \geqslant \int_\Omega (\alpha g \mathbb{1}_{A_n}) \,\mathrm{d}\mu = \\ = \alpha \sum_{j=1}^k a_j \mu\big(\{\omega \in A_n : g(\omega) = a_j\}\big) \to \alpha \sum_{j=1}^k a_j \mu\big(g = a_j\big) = \alpha \int_\Omega g \,\mathrm{d}\mu.\end{aligned}$$

Hence,

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \geqslant \alpha \int_\Omega g \,\mathrm{d}\mu.$$

Since this is true for every $0 < \alpha < 1$, we get

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \geqslant \int_\Omega g \,\mathrm{d}\mu.$$

Since this is true for every simple g such that $0 \leqslant g \leqslant f$,

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \geqslant \int_\Omega f \,\mathrm{d}\mu. \qquad \square$$

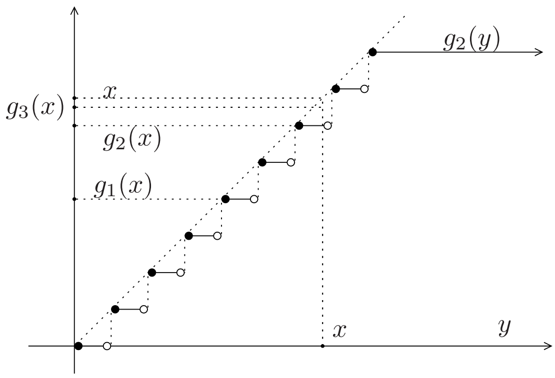


Figure 5.3: Graph of $g_2(y)$ and approximation $g_k(x) \uparrow x$ for some fixed x.

5.4.3 Approximation by simple functions

In order to bootstrap properties already established for simple functions to the case of non-negative functions, we start by showing that the latter can be approximated by the former in a monotone way, and finally apply (5.56).

We define a sequence of functions that approximate every non-negative extended number in a monotone way. For $n \in \mathbb{N}$, let $g_n : [0, +\infty] \to \mathbb{R}_+$ be defined by

$$g_n(x) = 2^{-n} \cdot \max \left\{ j \in \{0, 1, \dots, 2^n n\} \, : \, 2^{-n} j \leqslant x \right\},$$

illustrated in Figure 5.3. Note that, each g_n assumes finitely many values and, for every $x \in [0, +\infty]$, $g_n(x) \uparrow x$ as $n \to \infty$. For a non-negative extended Borel function f defined on Ω, the Borel functions $f_n := g_n \circ f$ are simple and satisfy $f_n \uparrow f$, see Figure 5.4.

Remark 5.57. The above construction shows that, for every non-negative extended Borel function f, there exists a sequence of simple functions $(f_n)_n$ such that $0 \leqslant f_n \uparrow f$. △

Remark 5.58. Combining the previous remark with the Monotone Convergence Theorem and Remark 5.53, we see that any other definition of integral of non-negative extended Borel functions that is unitary, monotone, linear and σ-additive, would be equivalent to this one. In other words, the integral is unique. △

5.4.4 Proofs of the axioms and other general properties

Proof of linearity. Suppose $f \geqslant 0$ and $h \geqslant 0$ are extended Borel functions. For homogeneity, given $\alpha \geqslant 0$ and we get $\int_\Omega (\alpha f) \, \mathrm{d}\mu = \alpha \int_\Omega f \, \mathrm{d}\mu$ directly from (5.54) since we can multiply each g in the supremum by α. For additivity, consider simple functions $f_n \uparrow f$ and $h_n \uparrow h$. Using (5.56) three times and linearity for the integral of simple functions,

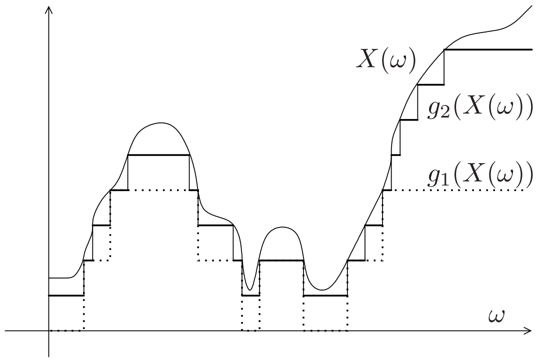


Figure 5.4: Approximation of an extended Borel function X by $g_1(X)$ e $g_2(X)$. Notice the similarity with Figure 5.1.

$$\int_\Omega (f+g)\,\mathrm{d}\mu = \lim_{n\to\infty}\int_\Omega (f_n+g_n)\,\mathrm{d}\mu = \lim_{n\to\infty}\Big(\int_\Omega f_n\,\mathrm{d}\mu + \int_\Omega g_n\,\mathrm{d}\mu\Big) =$$
$$= \lim_{n\to\infty}\int_\Omega f_n\,\mathrm{d}\mu + \lim_{n\to\infty}\int_\Omega g_n\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu.$$

This concludes the proof. □

Proof of σ-additivity. Let $(h_n)_n$ be a sequence of non-negative extended Borel functions. Define $f_n = h_1 + \cdots + h_n$. Then $f_n \uparrow f := \sum_j h_j$, and by (5.56) we have

$$\sum_{j=1}^{\infty}\int_\Omega h_j\,\mathrm{d}\mu = \lim_{n\to\infty}\sum_{j=1}^{n}\int_\Omega h_j\,\mathrm{d}\mu = \lim_{n\to\infty}\int_\Omega \sum_{j=1}^{n} h_j\,\mathrm{d}\mu =$$
$$= \lim_{n\to\infty}\int_\Omega f_n\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu = \int_\Omega \sum_{j=1}^{\infty} h_j\,\mathrm{d}\mu,$$

concluding the proof. □

Proof of (5.16). We have already proved this identity when f is an indicator function. By linearity, it also holds when f is a non-negative simple function. By (5.56), it holds for any non-negative extended Borel function f. Finally, if f is an extended Borel function,

$$\int_\Omega (f\mathbb{1}_B)\,\mathrm{d}\mu = \int_\Omega (f^+\mathbb{1}_B)\,\mathrm{d}\mu - \int_\Omega (f^-\mathbb{1}_B)\,\mathrm{d}\mu =$$
$$= \int_\Omega f^+\,\mathrm{d}(\mu_{|B}) - \int_\Omega f^-\,\mathrm{d}(\mu_{|B}) = \int_\Omega f\,\mathrm{d}(\mu_{|B}),$$

and when one of the two is defined, these expressions do not entail a "$\infty - \infty$." □

Proof of Theorem 5.19. We have already proved this identity when g is an indicator function. Just like in the previous proof, by linearity and (5.56), the identity holds for any non-negative extended Borel function g. Assuming g is an extended Borel function,

$$\int_{\Omega_2} g \,\mathrm{d}(f_*\mu) = \int_{\Omega_2} g^+ \,\mathrm{d}(f_*\mu) - \int_{\Omega_2} g^- \,\mathrm{d}(f_*\mu) = \\ = \int_{\Omega_1} (f^*g)^+ \,\mathrm{d}\mu - \int_{\Omega_1} (f^*g)^- \,\mathrm{d}\mu = \int_{\Omega_1} (f^*g) \,\mathrm{d}\mu,$$

and when one of the two is defined, these expressions do not entail a "$\infty - \infty$." □

We finally give the belated proof of Theorem 5.4 using Theorem 5.55.

Proof. We begin with monotonicity. Suppose $f \geqslant g$ and $\int_\Omega f^+ \,\mathrm{d}\mu < \infty$. Then $g^+ \leqslant f^+$ and $f^- \leqslant g^-$, hence $\int_\Omega g^+ \,\mathrm{d}\mu \leqslant \int_\Omega f^+ \,\mathrm{d}\mu < \infty$ and $-\int_\Omega g^- \,\mathrm{d}\mu \leqslant -\int_\Omega f^- \,\mathrm{d}\mu$. Adding the two last inequalities we get $\int_\Omega g \,\mathrm{d}\mu \leqslant \int_\Omega f \,\mathrm{d}\mu$ without incurring any "$\infty - \infty$" operation. The case $\int g^- \,\mathrm{d}\mu < \infty$ is identical.

We now move to homogeneity. Suppose both integrals are defined. Consider the case $-\infty \leqslant \alpha \leqslant 0$ first. Then

$$\int_\Omega (\alpha f) \,\mathrm{d}\mu = \int_\Omega (-\alpha f^-) \,\mathrm{d}\mu - \int_\Omega (-\alpha f^+) \,\mathrm{d}\mu = \\ = (-\alpha) \int_\Omega f^- \,\mathrm{d}\mu - (-\alpha) \int_\Omega f^+ \,\mathrm{d}\mu = \\ = \alpha \Big(\int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu \Big) = \alpha \int_\Omega f \,\mathrm{d}\mu.$$

These expressions cannot contain "$\infty - \infty$" because both integrals are defined. The case $\alpha \geqslant 0$ is analogous.

We conclude with additivity. Note that

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

whence

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

These are all non-negative extended Borel functions. By additivity in this case,

$$\int_\Omega (f+g)^+ \,\mathrm{d}\mu + \int_\Omega f^- \,\mathrm{d}\mu + \int_\Omega g^- \,\mathrm{d}\mu = \int_\Omega (f+g)^- \,\mathrm{d}\mu + \int_\Omega f^+ \,\mathrm{d}\mu + \int_\Omega g^+ \,\mathrm{d}\mu.$$

We are supposing that $\int_\Omega f \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu$ is defined, which implies that $\int_\Omega f^- \,\mathrm{d}\mu + \int_\Omega g^- \,\mathrm{d}\mu < \infty$ or $\int_\Omega f^+ \,\mathrm{d}\mu + \int_\Omega g^+ \,\mathrm{d}\mu < \infty$. Assume without loss of generality the former case. Since $(f+g)^- \leqslant f^- + g^-$, we have $\int_\Omega (f+g)^- \,\mathrm{d}\mu < \infty$, too, so we can subtract these three terms from both sides, getting

$$\int_\Omega (f+g)^+ \,\mathrm{d}\mu - \int_\Omega (f+g)^- \,\mathrm{d}\mu = \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu + \int_\Omega g^+ \,\mathrm{d}\mu - \int_\Omega g^- \,\mathrm{d}\mu.$$

This concludes the proof. □

6 Density and Radon-Nikodým Theorem

There are many situations where we want to define or study a measure in terms of another. Some of the most common uses is to tilt a probability measure and obtain a new one, to get probability distributions on $\mathbb{R}$ by specifying some density, and to define conditional expectations.

We explore these three cases, while keeping in mind that these are rather general concepts and tools. In particular, Proposition 6.17 we prove the good old formula for the expectation of an absolutely continuous random variable.

6.1 Density of measures and continuous variables

In this section we explore the concept of density and how it applies to absolutely continuous random variables.

6.1.1 Measures defined by a density

Given a measure space and a non-negative extended Borel function, it is possible to construct a new measure via the Lebesgue integral.

Proposition 6.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and f a non-negative extended Borel function. Then,*

$$\nu(A) := \int_A f \,\mathrm{d}\mu = \int_\Omega f \mathbb{1}_A \,\mathrm{d}\mu, \ \forall A \in \mathcal{F},$$

defines a measure on $(\Omega, \mathcal{F})$.

Exercise 6.2. Prove the above proposition. △

Example 6.3. Consider the measure space $(\mathbb{R}, \mathcal{B}, m)$ and the non-negative Borel function $g(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$ for $x \in \mathbb{R}$, where $\lambda > 0$. Then,

$$\mathbb{P}_\lambda(B) := \int_B g \,\mathrm{d}m = \int_\mathbb{R} g \mathbb{1}_B \,\mathrm{d}m, \ \forall B \in \mathcal{B},$$

defines a measure on $(\mathbb{R}, \mathcal{B})$, by Proposition 6.1. In fact, $\mathbb{P}_\lambda$ is a probability measure, since

$$\mathbb{P}_\lambda(\mathbb{R}) = \int_\mathbb{R} \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x) \,\mathrm{d}x = \int_0^\infty \lambda e^{-\lambda x} \,\mathrm{d}x = 1. \qquad △$$

A probability mass function assigns mass to elements of a discrete set, which can be seen as changing the weights assigned by the counting measure. In the above example we changed the weights assigned by the Lebesgue measure instead.

Definition 6.5 (Radon-Nikodým derivative)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to [0, +\infty]$ a non-negative extended Borel function. If ν is a measure on $(\Omega, \mathcal{F})$ such that

$$\nu(A) := \int_A f \, \mathrm{d}\mu, \quad \forall A \in \mathcal{F}, \tag{6.6}$$

we say that f is *the Radon-Nikodým derivative* (or *density*) *of* ν *with respect to* μ, and we write $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Remark 6.7. Given two measures μ, ν on $(\Omega, \mathcal{F})$, we say that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ *exists* if there is a non-negative extended Borel function f such that (6.6) holds. △

We say *the* Radon-Nikodým derivative because it is unique, in the following sense.

Proposition 6.8. *Given a σ-finite measure μ, and two non-negative extended Borel functions f and g such that*

$$\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu$$

for all $A \in \mathcal{F}$, we have $f = g$ for μ-a.e. ω.

Proof. Take $B_n \uparrow \Omega$ such that $\mu(B_n) < \infty$ for every n. Define $A_n := \{\omega \in B_n : g(\omega) \leqslant n, g(\omega) < f(\omega)\}$. Since $\mu(A_n) \leqslant \mu(B_n) < \infty$, we have that $g\mathbb{1}_{A_n}$ is finite and integrable. So $\int_{A_n} (f - g) \mathrm{d}\mu = \int_{A_n} f \mathrm{d}\mu - \int_{A_n} g \mathrm{d}\mu = 0$, since the second integral is finite and both are equal by assumption. By Exercise 5.1, $\mu(A_n) = 0$. On the other hand, by Proposition 2.35(ii), $\mu(g < f) = \lim_n \mu(A_n) = 0$. Analogous argument shows that $\mu(f < g) = 0$, concluding the proof.[9] □

Remark 6.9. Given a measure μ and a non-negative extended Borel function f, by Proposition 6.1 we can define a unique ν such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f$ for μ-a.e. ω. △

Example 6.10. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}(A) = \frac{\#A}{6}$. Take the function $g(n) = \frac{2}{7} n$. Then by Proposition 6.1 the function $\mathbf{P}$ given by $\mathbf{P}(A) = \int_\Omega (g\mathbb{1}_A) \mathrm{d}\mathbb{P}$ defines another measure on $(\Omega, \mathcal{F})$. △

Exercise 6.11. Show that the above $\mathbf{P}$ is a probability measure. △

Proposition 6.12 (Chain rule of the Radon-Nikodým derivative)**.** *Let ν, μ be measures on a given measurable space $(\Omega, \mathcal{F})$ such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ exists, and let $g : \Omega \to [0, +\infty]$ be a non-negative extended Borel function. Then,*

$$\int_\Omega g \, \mathrm{d}\nu = \int_\Omega g \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mu.$$

[9] Note that this is false without assuming that μ is σ-finite: taking $\Omega = \{1\}$ and $\mu(\{1\}) = \infty$, $f(1) = 1$ and $g(1) = 2$ we have $\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu$ for all $A \subseteq \Omega$ even though $f \neq g$ everywhere.

Proof. First, suppose that $g = \mathbb{1}_A$ for some $A \in \mathcal{F}$. In this case, we have

$$\int_\Omega \mathbb{1}_A \,\mathrm{d}\nu = \nu(A) = \int_A \tfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\mu = \int_\Omega \mathbb{1}_A \tfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\mu$$

as claimed. By linearity, it also holds when g is a non-negative simple function. By (5.56), it holds for any non-negative extended Borel function g. □

6.1.2 Absolutely continuous random variables

We now give the proper definition of a very familiar object.

Definition 6.13 (Absolutely continuous variables)**.** Given a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that X is an *absolutely continuous random variable* if $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}x}$ exists, where $\mathrm{d}x$ denotes the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$. The Radon-Nikodým derivative $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}x}$ is called the *probability density function* of X and usually denoted $f_X : \mathbb{R} \to \mathbb{R}$. Note that f_X is determined by X and $\mathbb{P}$ but the latter is omitted in the notation.

The probability measure of Example 6.3 is known *the Exponential distribution with parameter* λ, and g is called *the density* of this distribution.

Example 6.14. The unique probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B})$ such that

$$\tfrac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}x}(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

is called *the standard normal distribution on* $\mathbb{R}$. △

Example 6.15. The unique probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B})$ such that

$$\tfrac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}x}(x) = \mathbb{1}_{[0,1]}(x)$$

is called *the uniform distribution on* $[0, 1]$. △

Remark 6.16. The density of an absolutely continuous random variable is unique up to almost-everywhere equality. For instance, $\mathbb{1}_{(0,1)}(x)$ is also the density of the uniform distribution, as it differs from $\mathbb{1}_{[0,1]}(x)$ on a set of zero Lebesgue measure. △

The identity below is sometimes referred to as the law of the unconscious statistician, maybe because some people find it so obvious that they use it without realising it needs a proof.

Proposition 6.17. *If X is an absolutely continuous random variable, then, for every non-negative extended Borel function g, we have*

$$\mathbb{E}g(X) = \int_\mathbb{R} g(x) f_X(x) \,\mathrm{d}x.$$

Proof. First, notice that Theorem 5.19 gives

$$\mathbb{E}g(X) = \int_\Omega g(X(\omega))\,\mathbb{P}(\mathrm{d}\omega) = \int_\mathbb{R} g(x)\,\mathbb{P}_X(\mathrm{d}x) = \int_\mathbb{R} g\,\mathrm{d}\mathbb{P}_X$$

Now, since X is absolutely continuous, we can apply Proposition 6.12 to get

$$\int_\mathbb{R} g\,\mathrm{d}\mathbb{P}_X = \int_\mathbb{R} g\tfrac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}x}\,\mathrm{d}x = \int_\mathbb{R} g(x) f_X(x)\,\mathrm{d}x,$$

which concludes the proof. □

Example 6.18. If $X \sim \mathrm{Exp}(\lambda)$, then $\mathbb{E}X = \frac{1}{\lambda}$. Indeed, the density of $\mathbb{P}_X$ with respect to the Lebesgue measure is $\lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$. By the above proposition, we have $\mathbb{E}X = \int_\mathbb{R} x\lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)\,\mathrm{d}x = \cdots = \frac{1}{\lambda}$. △

6.1.3 Tilting a distribution

Below we see how to change the distribution of a random variable without changing the variable itself.

Example 6.19 (Tilting a probability measure). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \sim \mathcal{N}(0,1)$, i.e. X is a standard Normal random variable with probability distribution, that is,

$$\mathbb{P}[X \in A] = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x,\ \forall A \in \mathcal{B}(\mathbb{R}).$$

Fix some $\theta \in \mathbb{R}$. We want to define another measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$, such that

$$X \sim \mathcal{N}(\theta, 1),\ \text{under } \mathbf{P}.$$

Take $\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbb{P}}(\omega) = \frac{e^{\theta X(\omega)}}{e^{\frac{\theta^2}{2}}}$ and by Proposition 6.12, we get

$$\begin{aligned}\mathbf{E}[e^{t(X-\theta)}] &= \mathbb{E}[e^{t(X-\theta)}e^{\theta X - \frac{\theta^2}{2}}] \\ &= e^{-\theta t - \frac{\theta^2}{2}}\mathbb{E}[e^{(t+\theta)X}] \\ &= e^{-\theta t - \frac{\theta^2}{2}} e^{\frac{(t+\theta)^2}{2}} = e^{\frac{t^2}{2}},\ t \in \mathbb{R}.\end{aligned}$$

First, taking $t = 0$ we see that $\mathbf{P}(\Omega) = \mathbf{E}[1] = \mathbf{E}[e^0] = 1$, so $\mathbf{P}$ is a indeed a probability measure. Also, by uniqueness of moment generating functions, $X - \theta \sim \mathcal{N}(0,1)$ under $\mathbf{P}$. △

Remark 6.20. The fact that $\mathbf{P}$ is a probability measure comes from our careful choice of Radon-Nikodým derivative where the denominator equals the $\mathbb{P}$-expectation of the numerator. In general, if g is a non-negative extended Borel function and $\mathbb{E}g(X) < \infty$, then we can define $\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbb{P}} = \frac{g(X)}{\mathbb{E}g(X)}$, and the distribution of X under $\mathbf{P}$ equals the distribution of X under $\mathbb{P}$ *biased* by the function g.

The resulting measure is a probability measure because, taking $A = \Omega$ and $f = \frac{g(X)}{\mathbb{E}g(X)}$ in (6.6) will give

$$\mathbf{P}(\Omega) = \int_{\Omega} f \, \mathrm{d}\mathbb{P} = \frac{1}{\mathbb{E}g(X)} \int_{\Omega} g(X) \mathrm{d}\mathbb{P} = 1. \qquad \triangle$$

Example 6.21 (Size-biasing). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose $X \sim \text{Poisson}(\mu)$. Defining $\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbb{P}} = \frac{X}{\mu}$, the measure $\mathbf{P}$ is a probability measure and the distribution of $X - 1$ under $\mathbf{P}$ is Poisson with the same parameter. $\triangle$

Example 6.22 (Size-biasing). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose $Y \sim \text{Exp}(\lambda)$. Defining $\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbb{P}} = \lambda Y$, the measure $\mathbf{P}$ is a probability measure and the distribution of Y under $\mathbf{P}$ is $\Gamma(2, \lambda)$. $\triangle$

Example 6.23 (Size-biasing). In Example 6.10, we have the experiment of rolling a die, where $\mathbb{P}$ corresponds to sampling one of the six faces with equal probability. The measure $\mathbf{P}$ corresponds to sampling one of the 21 dots with equal probability, and observing the number of dots in its face. $\triangle$

6.2 The Radon-Nikodým Theorem

In this section we study when a measure ν can be expressed in terms of another measure μ, deformed by a variable factor $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ that we call Radon-Nikodým derivative.

6.2.1 Absolute continuity and the Radon-Nikodým Theorem

The following is a useful property for comparing measures.

Definition 6.24 (Absolute continuity between measures)**.** Let ν, μ be measures on a given measurable space $(\Omega, \mathcal{F})$. We say that ν *is absolutely continuous with respect to* μ if, for all $A \in \mathcal{F}$

$$\mu(A) = 0 \implies \nu(A) = 0.$$

We denote this relation by $\nu \ll \mu$.

Example 6.25. Consider the measurable space $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, and recall the Dirac measure δ_k at a point $k \in \mathbb{N}$, defined in Example 2.27. Denote by μ_C the counting measure on $\mathbb{N}$, as defined in Example 2.29. If a set $A \subseteq \mathbb{Z}$ has μ_C measure 0, then it has to have δ_k measure 0. Hence, δ_k is absolutely continuous with respect to μ_C, i.e. $\delta_k \ll \mu_C$. $\triangle$

Exercise 6.26. In the above example, can you think of a function $f : \mathbb{N} \to [0, \infty]$ that would be the Radon-Nikodým derivative $\frac{\mathrm{d}\delta_k}{\mathrm{d}\mu_C}$ of δ_k with respect to μ_C? $\triangle$

Example 6.27. Let μ_C be the counting measure on $(\mathbb{R}, \mathcal{B})$, so that:

- If $\mu_C(A) = \infty$, then A is an infinite Borel set.
- If $\mu_C(A) = N < \infty$, then $A = \{a_1, ..., a_N\}$ for some $a_1, ..., a_N \in \mathbb{R}$.
- If $\mu_C(A) = 0$, then $A = \emptyset$.

From the above cases, if $\mu_C(A) = 0$, then $m(A) = 0$, where m denotes the Lebesgue measure. Therefore, $m \ll \mu_C$. △

Exercise 6.28. In the above example, show that the converse is not true, that is, μ_C is not absolutely continuous with respect to m. △

Let μ and ν be measures on $(\Omega, \mathcal{F})$ and suppose the Radon-Nikodým derivative f of ν with respect to μ exists, that is, $\nu(A) = \int_A f \,\mathrm{d}\mu$ for ever $A \in \mathcal{F}$. By Exercise 5.25 we have $\nu \ll \mu$. We wonder whether the converse implication holds, that is, if $\nu \ll \mu$, does $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ exist?

Theorem 6.29 (The Radon-Nikodým Theorem)**.** *Let ν and μ be σ-finite measures on a given measurable space $(\Omega, \mathcal{F})$. Then, $\nu \ll \mu$ if and only if there exists a non-negative extended Borel function f, such that*

$$\nu(A) = \int_A f \,\mathrm{d}\mu, \ \forall A \in \mathcal{F}.$$

Proof. Given on page 96. □

We say that a random variable X is *absolutely continuous* if its distribution $\mathbb{P}_X$ assigns measure zero to Borel sets of Lebesgue measure zero, that is $\mathbb{P}_X \ll m$. From the above theorem, we see that any absolutely continuous random variable X must have a density!

6.2.2 Continuous random variables and densities

We say that a random variable X is *continuous* if its distribution assigns measure zero to every fixed number $x \in \mathbb{R}$. Since finite sets have Lebesgue measure zero, every absolutely continuous variable is also continuous (showing that the adverb "absolutely" has been used properly, unlike outer measures and finitely additive measures which are not measures).

One may wonder about what could be a random variable which is continuous but not absolutely continuous: in fact it is possible to construct such variables, and their distributions assign positive probability to a set which is uncountable but has zero Lebesgue measure.

If we are given a non-negative function f and we can check that

$$\mathbb{P}_X\big((-\infty, x]\big) = \int_{-\infty}^{x} f(s) \,\mathrm{d}s$$

for all $x \in \mathbb{R}$, then we can conclude that X is absolutely continuous and has density f. Indeed, the measure $\mathbf{P}(B) = \int_B f \, \mathrm{d}x$ coincides with $\mathbb{P}_X$ on the π-system $\{(-\infty, x]\}_{x\in\mathbb{R}}$ which generates $\mathcal{B}$, and the claim follows from the π-λ Theorem.

The converse is more delicate, and there is a theory that decomposes $\mathbb{P}_X$ into absolutely continuous and singularly continuous parts. We will bypass the theory by taking a more modest approach that is very useful in practice.

Suppose we are given a distribution function F and want to test if it is absolutely continuous. We take f to be its almost-everywhere derivative, and check whether

$$F(x) = \int_{-\infty}^{x} f(s) \, \mathrm{d}s$$

for every $x \in \mathbb{R}$. If this is the case, we can conclude that f is the density of F with respect to the Lebesgue measure.

Exercise 6.30. Suppose U has the uniform distribution on $[0, 1]$ and we take $X = e^U$. Find the density of X. (answer in footnote[10]) △

6.2.3 Conditional expectation

The Radon-Nikodým Theorem is a fundamental theorem in Measure Theory and Probability Theory. We now show one of the its many non-trivial applications.

Definition 6.31 (Conditional expectation)**.** Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, X is an extended random variable, and $\mathcal{G} \subseteq \mathcal{F}$ is another σ-algebra. We say that an extended random variable Z is the *conditional expectation of X given $\mathcal{G}$* if

$$Z : \Omega \to \overline{\mathbb{R}} \quad \text{is} \quad \mathcal{G}\text{-measurable} \tag{6.32}$$

and

$$\int_A Z \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P} \qquad \text{for all } A \in \mathcal{G}. \tag{6.33}$$

The above identity means whenever one of the integrals is defined, the other one is also defined and they coincide.

Exercise 6.34. Suppose that two extended random variables Z and W satisfy (6.32) and (6.33). Prove that $\mathbb{P}(Z = W) = 1$. △

Remark 6.35 (Uniqueness). The above definition says *the* conditional expectation because, if there is such a function, it is unique in the a.s. sense. △

Proposition 6.36. *Suppose $\mathbb{E}X$ is defined and there are two extended random variables Z and $\tilde{Z}$ that satisfy* (6.32) *and* (6.33)*. Then* $\mathbb{P}(Z = \tilde{Z}) = 1$.

[10] $f_X(x) = x^{-1}\mathbb{1}_{[1,e]}(x)$

Proof. We can assume without loss of generality that $\mathbb{E}X \neq -\infty$, so Z^- and $\tilde{Z}^-$ are integrable. By changing Z and $\tilde{Z}$ on a set of measure zero, we can assume that Z^- and $\tilde{Z}^-$ are finite for all ω. By linearity of the integral, $\int_A (Z + Z^- + \tilde{Z}^-)\,\mathrm{d}\mathbb{P} = \int_A (\tilde{Z} + Z^- + \tilde{Z}^-)\,\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{G}$. By Proposition 6.8, $Z + Z^- + \tilde{Z}^- = \tilde{Z} + Z^- + \tilde{Z}^-$ a.s., which proves the claim. □

Theorem 6.37. *In the setup of Definition 6.31, if X is a.s. finite and $\mathbb{E}X$ is defined, then the conditional expectation of X given $\mathcal{G}$ is defined.*

Proof (non-negative integrable case). Suppose that X is integrable, and define the measure ν on $(\Omega, \mathcal{G})$ by

$$\nu(A) = \int_A X\,\mathrm{d}\mathbb{P}$$

for $A \in \mathcal{G}$. Then $\nu(\Omega) = \mathbb{E}X < \infty$, so ν is σ-finite. By the Radon-Nikodým Theorem, $\frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}}$ exists as a non-negative extended Borel function on $(\Omega, \mathcal{G})$. Define $Z := \frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}}$ and note that it satisfies both (6.32) and (6.33). □

Proof (non-negative case). Now suppose that X is a non-negative finite random variable, not necessarily integrable. Then X can be written as $X = \sum_n X_n$ where each X_n is an integrable random variable. From the previous case, each X_n has a conditional expectation, which we denote Z_n.

We take $Z := \sum_n Z_n$. Note that Z is also non-negative and $\mathcal{G}$-measurable. By the previous case and σ-additivity of the integral, for every $A \in \mathcal{G}$,

$$\int_A Z\,\mathrm{d}\mathbb{P} = \sum_n \int_A Z_n\,\mathrm{d}\mathbb{P} = \sum_n \int_A X_n\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P},$$

which concludes the proof. □

Proof (general case). We can assume that $\mathbb{E}X^- < \infty$. From the previous case, there exist two $\mathcal{G}$-measurable functions $Z^\pm$ such that $\int_A Z^\pm\,\mathrm{d}\mathbb{P} = \int_A X^\pm\,\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{G}$. In particular, $\int_\Omega Z^-\,\mathrm{d}\mathbb{P} = \int_\Omega X^-\,\mathrm{d}\mathbb{P} < \infty$, so by Z^- is a.s. finite, and we can assume that $Z^-(\omega) < \infty$ for every $\omega \in \Omega$. Define $Z = Z^+ - Z^-$, which is well-defined because Z^- is always finite. Then for all $A \in \mathcal{G}$, we have

$$\int_A Z\,\mathrm{d}\mathbb{P} = \int_A Z^+\,\mathrm{d}\mathbb{P} - \int_A Z^-\,\mathrm{d}\mathbb{P} = \int_A X^+\,\mathrm{d}\mathbb{P} - \int_A X^-\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P}$$

which are well-defined because Z^- and X^- are integrable. □

7 Product measures and independence

Product measures are a fundamental tool in Measure Theory. Among its many applications, this topic is useful for computing high-dimensional integrals one variable at at time, computing the expectation of an integral as the integral of the expected value, and to construct new probability spaces from old ones by running two experiments (or even two copies of the same experiment) independently.

7.1 Product σ-algebra and product measure

7.1.1 Product σ-algebras

Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, and let $\Omega_1 \times \Omega_2$ be the Cartesian product of Ω_1 and Ω_2. We say that a subset of $\Omega_1 \times \Omega_2$ is a *rectangle with measurable sides* if it is of the form $A \times B$, for some $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$.

Definition 7.1 (Product of σ-algebras)**.** We define *the product of the σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$*, denoted as $\mathcal{F}_1 \otimes \mathcal{F}_2$, as the σ-algebra generated by the class

$$\mathcal{F}_1 \times \mathcal{F}_2 = \{A \times B \subseteq \Omega_1 \times \Omega_2 : A \in \mathcal{F}_1,\ B \in \mathcal{F}_2\}$$

on the sample space $\Omega_1 \times \Omega_2$.

In other words, $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the σ-algebra generated by the collection of all rectangles with measurable sides.[11] It is important to note that $\mathcal{F}_1 \otimes \mathcal{F}_2$ is much larger than $\mathcal{F}_1 \times \mathcal{F}_2$. Let us discuss the case of $\mathbb{R}^2$.

Example 7.2. Consider the sample space $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. The product of Borel σ-algebras $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ equals the Borel σ-algebra $\mathcal{B}(\mathbb{R}^2)$ – see page 99. △

Remark 7.3. To see that $\mathcal{B}(\mathbb{R}^2) \neq \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$, notice that the set

$$\{(x, y) : 0 \leqslant x^2 + y^2 < 1\}$$

lies in $\mathcal{B}(\mathbb{R}^2)$, but not in $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. Literally, the circle is not a rectangle! △

Given a subset or a function on $\Omega_1 \times \Omega_2$, we need some notation involving the separate parts on Ω_1 and Ω_2.

Definition 7.4 (Sections of subsets and functions on $\Omega_1 \times \Omega_2$)**.** Let $\Omega_1 \times \Omega_2$ be the Cartesian product of given sample spaces Ω_1 and Ω_2, E a subset of $\Omega_1 \times \Omega_2$ and f a function on $\Omega_1 \times \Omega_2$. Fix $x \in \Omega_1$ and $y \in \Omega_2$.

[11] There is an abuse of notation when using the symbol "$\times$" for collections of sets, as we are using $\mathcal{F}_1 \times \mathcal{F}_2$ to denote a collection of rectangles $A \times B$ rather than pairs (A, B).

(i) The *sections* E_x *and* E^y *of* E are the subsets of Ω_2 and Ω_1, respectively, given by

$$E_x = \{\tilde{y} \in \Omega_2 : (x, \tilde{y}) \in E\} \quad \text{and} \quad E^y = \{\tilde{x} \in \Omega_1 : (\tilde{x}, y) \in E\}.$$

(ii) The *sections* f_x *and* f^y *of* f are the functions on Ω_2 and Ω_1, respectively, given by

$$f_x(\tilde{y}) = f(x, \tilde{y}) \quad \text{and} \quad f^y(\tilde{x}) = f(\tilde{x}, y).$$

Naturally, we want to study subsets and functions that are measurable with respect to a product of σ-algebras.

Lemma 7.5. *Let* $(\Omega_1, \mathcal{F}_1)$ *and* $(\Omega_2, \mathcal{F}_2)$ *be measurable spaces.*

(i) *If* E *is* $\mathcal{F}_1 \otimes \mathcal{F}_2$*-measurable, then for each* $x \in \Omega_1$ *and* $y \in \Omega_2$*, the sections* E_x *and* E^y *are* $\mathcal{F}_2$*-measurable and* $\mathcal{F}_1$*-measurable, respectively.*

(ii) *If* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *is* $\mathcal{F}_1 \otimes \mathcal{F}_2$*-measurable, then for each* $x \in \Omega_1$ *and* $y \in \Omega_2$*, the sections* f_x *and* f^y *are* $\mathcal{F}_2$*-measurable and* $\mathcal{F}_1$*-measurable, respectively.*

Proof. Given on page 100. □

7.1.2 Product measure

We now establish existence and uniqueness of the product measure defined by σ-finite measure spaces. We start with a basic lemma.

Lemma 7.6. *Let* $(\Omega_1, \mathcal{F}_1, \mu)$ *and* $(\Omega_2, \mathcal{F}_2, \nu)$ *be two* σ*-finite measure spaces. If* E *is* $\mathcal{F}_1 \otimes \mathcal{F}_2$*-measurable, then the functions* $x \mapsto \nu(E_x)$ *and* $y \mapsto \mu(E^y)$ *are* $\mathcal{F}_1$*-measurable and* $\mathcal{F}_2$*-measurable, respectively.*

Proof. Given on page 100. □

Theorem 7.7 (Product measure)**.** *Let* $(\Omega_1, \mathcal{F}_1, \mu)$ *and* $(\Omega_2, \mathcal{F}_2, \nu)$ *be two* σ*-finite measure spaces. Then, there is a unique measure* $\mu \otimes \nu$ *on the* σ*-algebra* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *such that*

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B), \ \forall A \in \mathcal{F}_1, B \in \mathcal{F}_2.$$

It is given by

$$(\mu \otimes \nu)(E) = \int_{\Omega_1} \nu(E_x)\mu(\mathrm{d}x) = \int_{\Omega_2} \mu(E^y)\nu(\mathrm{d}y), \ \forall E \in \mathcal{F}_1 \otimes \mathcal{F}_2.$$

The measure $\mu \otimes \nu$ *is called the product of* μ *and* ν*.*

Proof. Given on page 101. $\square$

Example 7.8. We return to the case of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. We know from Example 7.2 that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Let m_1 and m_2 denote the Lebesgue measure on $\mathbb{R}$ and $\mathbb{R}^2$, respectively. For each rectangle of the form

$$(a_1, b_1] \times (a_2, b_2], \ -\infty < a_i \leqslant b_i < +\infty, \ i = 1, 2,$$

its measure is given by

$$\begin{aligned} m_2((a_1, b_1] \times (a_2, b_2]) &= (b_1 - a_1)(b_2 - a_2) = \\ &= m_1((a_1, b_1]) m_1((a_2, b_2]) = (m_1 \otimes m_1)((a_1, b_1] \times (a_2, b_2]). \end{aligned} \tag{7.9}$$

Therefore, we can deduce that

$$m_2 = m_1 \otimes m_1 \text{ on } \mathcal{B}(\mathbb{R}^2).$$

This is precisely the motivation for the construction of the Lebesgue measure m_2 on $\mathbb{R}^2$.[12] $\triangle$

7.2 Fubini and Tonelli Theorems

7.2.1 Fubini and Tonelli Theorems

Having constructed product measures, we move on to the fundamental theorems that allow us to integrate a measurable function via iterated integrals, which we can compute.

Theorem 7.10 (Tonelli Theorem)**.** *Let $(\Omega_1, \mathcal{F}_1, \mu)$ and $(\Omega_2, \mathcal{F}_2, \nu)$ be two σ-finite measure spaces, and let $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ be a non-negative $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function. Then, the function $x \mapsto \int_{\Omega_2} f_x \, \mathrm{d}\nu$ is $\mathcal{F}_1$-measurable, the function $y \mapsto \int_{\Omega_1} f^y \, \mathrm{d}\mu$ is $\mathcal{F}_2$-measurable, and*

$$\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}(\mu \otimes \nu) = \int_{\Omega_1} \Big(\int_{\Omega_2} f_x \, \mathrm{d}\nu \Big) \mu(\mathrm{d}x) = \int_{\Omega_2} \Big(\int_{\Omega_1} f^y \, \mathrm{d}\mu \Big) \nu(\mathrm{d}y).$$

Note that, in practice, the most useful form of the last equality is

$$\int_{\Omega_1} \Big(\int_{\Omega_2} f(x, y) \, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) = \int_{\Omega_2} \Big(\int_{\Omega_1} f(x, y) \, \mu(\mathrm{d}x) \Big) \nu(\mathrm{d}y), \tag{7.11}$$

and a particular case is, for $h(x, y) = f(x) g(y)$,

$$\int_{\Omega_1 \times \Omega_2} h \, \mathrm{d}(\mu \otimes \nu) = \Big(\int_{\Omega_1} f \, \mathrm{d}\mu \Big) \Big(\int_{\Omega_2} g \, \mathrm{d}\nu \Big). \tag{7.12}$$

[12] Expanded from [Coh13, 5.1.5].

Proof. We show this for the case of $f = \mathbb{1}_E$, for some $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, as linearity of the integral extends this to any non-negative simple function, and applying the Monotone Convergence Theorem generalises the theorem to any non-negative measurable function.

Fix $x \in \Omega_1$ and $y \in \Omega_2$, and notice that

$$f_x = (\mathbb{1}_E)_x = \mathbb{1}_{E_x} \quad \text{and} \quad f^y = (\mathbb{1}_E)^y = \mathbb{1}_{E^y}.$$

Thus, we get

$$\int_{\Omega_2} f_x \, \mathrm{d}\nu = \nu(E_x) \quad \text{and} \quad \int_{\Omega_1} f^y \, \mathrm{d}\mu = \mu(E^y).$$

Note that measurability of the functions $x \mapsto \int_{\Omega_2} f_x \, \mathrm{d}\nu$ and $y \mapsto \int_{\Omega_1} f^y \, \mathrm{d}\mu$ follows from Lemma 7.6. Equality of the three integrals in the statement follows from Theorem 7.7.[13] $\square$

Theorem 7.13 (Fubini Theorem). *Let $(\Omega_1, \mathcal{F}_1, \mu)$ and $(\Omega_2, \mathcal{F}_2, \nu)$ be two σ-finite measure spaces, and let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be an $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable and $\mu \otimes \nu$-integrable function. Then, the section f_x is ν-integrable for μ-a.e. x and the section f^y is μ-integrable for ν-a.e. y. In addition,*

$$\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}(\mu \otimes \nu) = \int_{\Omega_1} \Big(\int_{\Omega_2} f_x \, \mathrm{d}\nu \Big) \mu(\mathrm{d}x) = \int_{\Omega_2} \Big(\int_{\Omega_1} f^y \, \mathrm{d}\mu \Big) \nu(\mathrm{d}y).$$

In the above iterated integrals, we can take the integrand to be zero on the (zero-measure) set where the inner integral is undefined.

As before, the most useful form of the last equality is (7.11). However, here the innermost integral may be undefined on a set of points that has measure zero (with respect to the second measure). In order to check whether f is integrable with respect to $\mu \otimes \nu$, we can compute (or estimate!) one of the iterated integrals in (7.11) with $|f|$ instead of f, to check if it is finite, and use Tonelli Theorem to conclude that f is integrable or not. This powerful combination of both theorems is sometimes called the Fubini-Tonelli Theorem.

Proof. Let f^+ and f^- denote the positive and negative part of f, respectively. From Lemma 7.5 (ii), we know that each f_x is $\mathcal{F}_2$-measurable, and so is each $(f^+)_x$ and $(f^-)_x$. Theorem 7.10 and integrability of f (i.e. integrability of f^+ and f^-) imply that the functions $x \mapsto \int_{\Omega_2} (f^+)_x \, \mathrm{d}\nu$ and $x \mapsto \int_{\Omega_2} (f^-)_x \, \mathrm{d}\nu$ are $\mathcal{F}_1$-measurable and μ-integrable, and so finite μ-a.e. Thus, $x \mapsto \int_{\Omega_2} f_x \, \mathrm{d}\nu$ is finite μ-a.e., i.e. f_x is ν-integrable μ-a.e. Now, consider the set N given by

$$N = \{x \in \Omega_1 : \int_{\Omega_2} |f_x| \, \mathrm{d}\nu = \infty\}.$$

[13] Based on [Coh13, 5.2.1].

As argued above, $N \in \mathcal{F}_1$ and $\mu(N) = 0$. By Theorem 7.10 and linearity of the integral, we get

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f\,\mathrm{d}(\mu\otimes\nu) &= \int_{\Omega_1\times\Omega_2} f^+\,\mathrm{d}(\mu\otimes\nu) - \int_{\Omega_1\times\Omega_2} f^-\,\mathrm{d}(\mu\otimes\nu)\\ &= \int_{\Omega_1}\Big(\int_{\Omega_2}(f^+)_x\,\mathrm{d}\nu\Big)\mu(\mathrm{d}x) - \int_{\Omega_1}\Big(\int_{\Omega_2}(f^-)_x\,\mathrm{d}\nu\Big)\mu(\mathrm{d}x)\\ &= \int_{N^c}\Big(\int_{\Omega_2}(f^+)_x\,\mathrm{d}\nu\Big)\mu(\mathrm{d}x) - \int_{N^c}\Big(\int_{\Omega_2}(f^-)_x\,\mathrm{d}\nu\Big)\mu(\mathrm{d}x)\\ &= \int_{N^c}\Big(\int_{\Omega_2} f_x\,\mathrm{d}\nu\Big)\mu(\mathrm{d}x) = \int_{\Omega_1}\Big(\int_{\Omega_2} f_x\,\mathrm{d}\nu\Big)\mu(\mathrm{d}x),\end{aligned}$$

where we use that the integral on a set of measure zero is zero.

The proof that $\int_{\Omega_1\times\Omega_2} f\,\mathrm{d}(\mu\otimes\nu) = \int_{\Omega_2}\Big(\int_{\Omega_1} f^y\,\mathrm{d}\mu\Big)\nu(\mathrm{d}y)$ is analogous.[14] □

7.2.2 Applications of Fubini and Tonelli Theorems

Example 7.14. Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space, and let m be the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Consider a non-negative $\mathcal{F}$-measurable function $f : \Omega \to [0, \infty]$, and let E be the set given by

$$E = \{(\omega, y) \in (\Omega, \mathbb{R}) : 0 \leqslant y < f(\omega)\}.$$

Observe that E encapsulates the area under f. Moreover, E lies in the product σ-algebra $\mathcal{F} \otimes \mathcal{B}(\mathbb{R})$, since f is $\mathcal{F}$-measurable. Then, we can compute that

$$\begin{aligned}(\mu\otimes m)(E) = \int_\Omega m(E_\omega)\mu(\mathrm{d}\omega) = &\\ = \int_\Omega m(\{y\in\mathbb{R} : 0\leqslant y < f(\omega)\})\mu(\mathrm{d}\omega) &= \int_\Omega f(\omega)\mu(\mathrm{d}\omega).\end{aligned}$$

On the other hand, we have

$$(\mu\otimes m)(E) = \int_{\mathbb{R}} \mu(E^y)\,\mathrm{d}y = \int_0^\infty \mu(\{\omega\in\Omega : 0\leqslant y < f(\omega)\})\,\mathrm{d}y,$$

which gives the equality

$$\int_\Omega f(\omega)\mu(\mathrm{d}\omega) = \int_0^\infty \mu(\{f(\omega) > y\})\,\mathrm{d}y.$$

This relation tells us that we can integrate a non-negative measurable function f over a σ-finite measure μ by instead looking at the function $y \mapsto \mu(E^y)$ (with E as defined above), whose integral can sometimes be easier to compute.[15] △

[14] Based on [Coh13, 5.2.2].
[15] Based on [Coh13, 5.3.1].

To grasp more intuition on Theorems 7.10 and 7.13, we revisit two important properties of doubly-indexed sequences.

Proof of Theorem 1.13. Let μ denote the counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, where μ is σ-finite, and consider the function $f : \mathbb{N} \times \mathbb{N} \to [0, \infty]$ given by

$$f(m, n) = x_{m,n}, \ \ m, n \in \mathbb{N}.$$

Since $\mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N}) = \mathcal{P}(\mathbb{N} \times \mathbb{N})$, the function f is $\mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N})$-measurable. Moreover,

$$\int_{\mathbb{N}\times\mathbb{N}} f \, \mathrm{d}(\mu \otimes \mu) = \sum_{(m,n)\in\mathbb{N}^2} x_{m,n},$$

$$\int_{\mathbb{N}} \Big(\int_{\mathbb{N}} f_m \, \mathrm{d}\mu \Big) \mu(\mathrm{d}m) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x_{m,n},$$

and

$$\int_{\mathbb{N}} \Big(\int_{\mathbb{N}} f^n \, \mathrm{d}\mu \Big) \mu(\mathrm{d}n) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} x_{m,n},$$

which are all equal by Theorem 7.10, completing the proof.[16] □

Proof of Theorem 1.14. Let μ denote the counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, and consider the function $f : \mathbb{N} \times \mathbb{N} \to \overline{\mathbb{R}}$ given by

$$f(m, n) = x_{m,n}, \ \ m, n \in \mathbb{N}.$$

Since $\mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N}) = \mathcal{P}(\mathbb{N} \times \mathbb{N})$, the function f is $\mathcal{P}(\mathbb{N}) \otimes \mathcal{P}(\mathbb{N})$-measurable. Moreover, the sum $\sum_{(m,n)\in\mathbb{N}^2} x_{m,n}$ converges absolutely if and only if f is $\mu \otimes \mu$-integrable. Thus, by Theorem 7.13, we can conclude as for Theorem 1.13, in that

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x_{m,n} = \sum_{(m,n)\in\mathbb{N}^2} x_{m,n} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} x_{m,n}.$$

In other words, for absolutely convergent (not necessarily non-negative) doubly-indexed series, the order of summation can be reversed.[17] □

7.3 Independence

In this section, we introduce various concepts of independence and how they relate to each other.

[16] Expanded from [Coh13, 5.3.2].
[17] Expanded from [Coh13, 5.3.2].

7.3.1 Independence of σ-algebras

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The σ-algebra $\mathcal{F}$ on Ω is a collection of events, so that an event is an $\mathcal{F}$-measurable subset of Ω.

We know that two events A and B are said to be independent if their joint probability equals the product of their probabilities, that is

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \tag{7.15}$$

and two random variables X and Y are said to be *independent* if

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D), \ \forall C, D \in \mathcal{B}(\mathbb{R}). \tag{7.16}$$

We also know it is possible to have A_1 independent of A_2, A_2 independent of A_3, and A_3 independent of A_1, without $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$. This notion of pairwise independence can be useful at times (for instance, the variance of the sum is the sum of the variances, etc.), but here we are interested in a stronger notion of independence.

For A_1, A_2, A_3 to be independent, we require that

$$\mathbb{P}(B_1 \cap B_2 \cap B_3) = \mathbb{P}(B_1)\mathbb{P}(B_2)\mathbb{P}(B_3),$$

for all choices of $B_j \in \{\emptyset, A_j, A_j^c, \Omega\}$. This motivates the following definition.

Definition 7.17 (Independent σ-algebras)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}_1, \ldots, \mathcal{G}_n$ a family of sub-σ-algebras of $\mathcal{F}$. We say that $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are *independent* if

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i), \ \forall A_i \in \mathcal{G}_i, \ i = 1, \ldots, n.$$

The abstract formulation of σ-algebras turns out to be very powerful and versatile, as we will see now.

Definition 7.18 (Independent random variables)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ a family of random variables. We say that $X_1, \ldots, X_n$ are *independent* if the sub-σ-algebras of $\mathcal{F}$ given by

$$\sigma(X_1), \ldots, \sigma(X_n)$$

are independent.

Remark 7.19. The random variables $X_1, \ldots, X_n$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if and only if

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i), \ \forall B_i \in \mathcal{B}(\mathbb{R}), \ i = 1, \ldots, n.$$

Indeed, the events in $\sigma(X_j)$ are exactly of the form $\{X_j \in B_j\}$ for $B_j \in \mathcal{B}$. $\triangle$

Exercise 7.20. Suppose X_1 and X_2 are independent random variables. Suppose also that g_1 and $g_2 : \mathbb{R} \to \mathbb{R}$ are Borel-measurable functions. Show that $g_1(X_1)$ and $g_2(X_2)$ are independent random variables. $\triangle$

Finally, the definition of independent events is conveniently formulated using the definition of independent random variables.

> **Definition 7.21** (Independent events)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, \dots, A_n$ a collection of events. We say that $A_1, \dots, A_n$ are *independent* if the random variables given by
>
> $$\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$$
>
> are independent.

Remark 7.22. Events $A_1, \dots, A_n$ on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if and only if the sub-σ-algebras of $\mathcal{F}$ given by

$$\mathcal{G}_1, \dots, \mathcal{G}_n$$

are independent, where $\mathcal{G}_i = \sigma(\mathbb{1}_{A_i}) = \{\emptyset, \Omega, A_i, A_i^c\},\ i = 1, \dots, n.$ $\triangle$

7.3.2 Independence and product measures

We now turn to product measures as discussed in previous sections.

Remark 7.23 (Longer products). Given σ-finite measure spaces $(\Omega_j, \mathcal{F}_j, \mu_j)$ for $j = 1, \dots, n$, we can define $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n := \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$. We can also define $\mu_1 \otimes \cdots \otimes \mu_n$ as the unique measure ν such that $\nu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n)$. The theory studied in previous sections is a particular case when $n = 2$. Uniqueness follows from the π-λ Theorem as before, and existence follows by taking $(\mu_1 \otimes \cdots \otimes \mu_{n-1}) \otimes \mu_n$. $\triangle$

We define the *distribution* of a family of random variables $X_1, \dots, X_n$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ by

$$\mathbb{P}_{X_1,\dots,X_n}(A) := \mathbb{P}\big[(X_1, \dots, X_n) \in A\big],\ A \in \mathcal{B}(\mathbb{R}^n).$$

From Remarks 7.19 and 7.23, $X_1, \dots, X_n$ are independent if and only if their distribution satisfies

$$\mathbb{P}_{X_1,\dots,X_n} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}. \tag{7.24}$$

We now establish a useful and well-known criterion that shows independence of random variables via their cumulative distribution functions.

Proposition 7.25 (Criterion for independent random variables)**.** *Let $X_1, \dots, X_n$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the variables $X_1, \dots, X_n$ are independent if and only if*

$$\mathbb{P}(X_1 \leqslant a_1, \dots, X_n \leqslant a_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leqslant a_i), \ \forall a_1, \dots, a_n \in \mathbb{R}.$$

Proof. The direct implication is immediate. Now suppose the above equality holds for all $a_1, \dots, a_n \in \mathbb{R}$. Define $A_{a_1,\dots,a_n} := (-\infty, a_1] \times \cdots \times (-\infty, a_n]$ and $\mathcal{E} := \{A_{a_1,\dots,a_n}\}_{a_1,\dots,a_n \in \mathbb{R}}$. The above equality says that $\mathbb{P}_{X_1,\dots,X_n}$ and $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ coincide on $\mathcal{E}$. Now note that $\mathcal{E}$ is a π-system and $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R}^n)$. By the π-λ Theorem, $\mathbb{P}_{X_1,\dots,X_n} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$, completing the proof. □

Proposition 7.26. *Suppose X and Y are independent non-negative extended random variables. Then $\mathbb{E}[XY] = (\mathbb{E}X)(\mathbb{E}Y)$.*

Proof. The proof is through the following chain:

$$\begin{aligned}
\mathbb{E}[XY] &= \int_\Omega (X(\omega)Y(\omega))\mathbb{P}(\mathrm{d}\omega) \\
&= \int_{\mathbb{R}\times\mathbb{R}} xy \, \mathbb{P}_{X,Y}(\mathrm{d}x, \mathrm{d}y) \\
&= \int_{\mathbb{R}\times\mathbb{R}} xy \, (\mathbb{P}_X \otimes \mathbb{P}_Y)(\mathrm{d}x, \mathrm{d}y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} xy \, \mathbb{P}_X(\mathrm{d}x)\mathbb{P}_Y(\mathrm{d}y) \\
&= \int_{\mathbb{R}} y \Big(\int_{\mathbb{R}} x \, \mathbb{P}_X(\mathrm{d}x) \Big) \, \mathbb{P}_Y(\mathrm{d}y) \\
&= \Big(\int_{\mathbb{R}} x \, \mathbb{P}_X(\mathrm{d}x) \Big) \Big(\int_{\mathbb{R}} y \, \mathbb{P}_Y(\mathrm{d}y) \Big) \\
&= \mathbb{E}X \cdot \mathbb{E}Y,
\end{aligned}$$

which gives $\mathbb{E}[XY] = (\mathbb{E}X)(\mathbb{E}Y)$. □

Exercise 7.27. Justify each step in the above chain. △

Proposition 7.28. *Suppose X and Y are independent integrable random variables. Then XY is an integrable random variable and $\mathbb{E}[XY] = (\mathbb{E}X)(\mathbb{E}Y)$.*

Proof. We separate positive from negative parts, expand and regroup:

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-] \\
&= \mathbb{E}X^+ \cdot \mathbb{E}Y^+ - \mathbb{E}X^+ \cdot \mathbb{E}Y^- - \mathbb{E}X^- \cdot \mathbb{E}Y^+ + \mathbb{E}X^- \cdot \mathbb{E}Y^- \\
&= (\mathbb{E}X^+ - \mathbb{E}X^-)(\mathbb{E}Y^+ - \mathbb{E}Y^-) = \mathbb{E}X \cdot \mathbb{E}Y.
\end{aligned}$$

□

Exercise 7.29. Suppose $X_1, \dots, X_n$ are independent and non-negative (or integrable) extended random variables. Prove that $\mathbb{E}\big[\prod_j X_j\big] = \prod_j \mathbb{E}X_j$. △

8 Convergence of measurable functions

Convergence is related to approximation, and as such it is at the core of Measure Theory and Probability Theory. To say that a number x is close to a fixed number y has a unique meaning: their difference is small. More precisely, $x_n \to y$ if $|x_n - y|$ can be made as small as possible by requiring n to be large.

The situation becomes considerably more involved when we replace numbers by functions. If we are considering Borel functions f_n and g on a measure space $(\Omega, \mathcal{F}, \mu)$, or random variables X_n and Y that can be observed on a given probabilistic model, there are several different notions of convergence, each one with their own meaning and implications.

8.1 Modes of convergence

8.1.1 Convergence of Borel functions

Definition 8.1 (a.e. convergence)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n \geqslant 1}, f$ be Borel functions. We say that $(f_n)_{n \geqslant 1}$ *converges to* f μ*-a.e.* if

$$\mu\{\omega : f_n(\omega) \not\to f(\omega)\} = 0.$$

Then, we write $f_n \xrightarrow{\text{a.e.}} f$. The measure μ is implicit in the notation.

Let us consider another closely-related definition and then move to some examples to see how they compare.

Definition 8.2 (Convergence in measure)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n \geqslant 1}, f$ be Borel functions. We say that $(f_n)_{n \geqslant 1}$ *converges to* f *in measure* if for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \mu\{|f_n - f| > \varepsilon\} = 0.$$

Then, we write $f_n \xrightarrow{\mu} f$.

Example 8.3. Consider the measure space $([0,1], \mathcal{B}([0,1]), m_{|_{[0,1]}})$ and $(f_n)_{n \geqslant 1}$ a sequence of Borel functions given by

$$f_n(x) = \mathbb{1}_{[0, \frac{1}{n})}, \; x \in [0,1], \; n \geqslant 1.$$

Then, we can see that

$$m_{|_{[0,1]}}(f_n \not\to 0) = m_{|_{[0,1]}}(\{0\}) = 0,$$

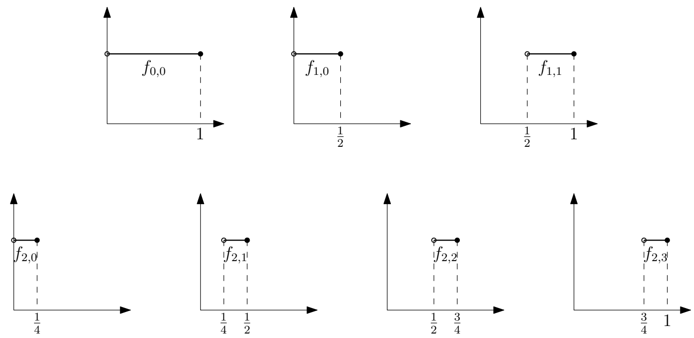


Figure 8.1: First elements of the 'dancing wave.'

and so, $f_n \xrightarrow{\text{a.e.}} 0$. Moreover, for any $0 < \varepsilon < 1$,

$$m\{|f_n| > \varepsilon\} = \frac{1}{n} \to 0,$$

and for $\varepsilon \geqslant 1$ this measure is zero for all n, so $f_n \xrightarrow{m} 0$. △

Example 8.4 (Travelling wave). Consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, and consider the sequence of Borel functions $(f_n)_{n \geqslant 1}$, where $f_n = \mathbb{1}_{[n,n+1)}$, which resembles a travelling wave. Notice that $f_n(x) \to 0$ for every x, in particular $f_n(x) \to 0$ for a.e. x. However, we have that

$$m\{|f_n| \geqslant \varepsilon\} = 1 \not\to 0$$

for $\varepsilon = \frac{1}{2}$, which implies that f_n does not converge to 0 in measure. △

Example 8.5 (Dancing wave). Consider the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m_{|_{[0,1]}})$, and consider the sequence of Borel functions $(f_{n,k})_{n \geqslant 0}$, where $f_{n,k} = \mathbb{1}_{(\frac{k}{2^n}, \frac{k+1}{2^n}]}$, $k = 0, \dots, 2^n - 1$, which resembles a dancing wave. Figure 8.1 illustrates the construction of $f_{n,k}$. As n gets larger, each sub-interval gets shorter, so that for any $0 < \varepsilon < 1$ and any $p \geqslant 1$, we have

$$m\{(\tfrac{k}{2^n}, \tfrac{k+1}{2^n}]\} = 2^{-n} \to 0$$

On the other hand, given $n \geqslant 0$, letting k range from 0 to $2^n - 1$ ensures that $f_{n,k}$ still oscillates between 0 and 1 for every $x \in (0, 1]$. We now order the pairs (n, k) sequentially as follows. For each $j \geqslant 1$, there is a unique pair $n \geqslant 0$ and $0 \leqslant k < 2^n$ such that $j = 2^n + k$. Write g_j for the corresponding $f_{n,k}$. From the above observations, $g_j \to 0$ in measure as $j \to \infty$, but the set of x such that $g_j(x)$ converges does not contain points in $(0, 1]$. △

Definition 8.6 (Convergence in the vector space $\mathcal{L}^p$)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. For $p \geqslant 1$, we define $\mathcal{L}^p(\mu)$ as the space of Borel functions

f such that $\int_\Omega |f|^p \,\mathrm{d}\mu < \infty$. Given f and $(f_n)_{n\geqslant 1}$ in $\mathcal{L}^p(\mu)$, we say that $(f_n)_{n\geqslant 1}$ *converges to f in $\mathcal{L}^p$* if

$$\lim_{n\to\infty} \int_\Omega |f_n - f|^p \,\mathrm{d}\mu = 0.$$

In this case, we write $f_n \xrightarrow{\mathcal{L}^p} f$.

Remark 8.7. To see that $\mathcal{L}^p$ is a vector space, use $|f+g|^p \leqslant |2f|^p + |2g|^p$. $\triangle$

Example 8.8 (Dancing wave revisited). Consider the 'growing dancing wave' given by $\tilde{f}_{n,k} = 2^{n/2} f_{n,k}$, and define $\tilde{g}_j$ as the corresponding sequence. Then for $p = 1$ we have

$$\int_{[0,1]} |\tilde{f}_{n,k}| \,\mathrm{d}m = 2^{-n/2} \to 0,$$

so $\tilde{g}_j \to 0$ in $\mathcal{L}^1$ as $j \to \infty$. On the other hand,

$$\int_{[0,1]} |\tilde{f}_{n,k}|^2 \,\mathrm{d}m = 1 \not\to 0,$$

so $\tilde{g}_j \not\to 0$ in $\mathcal{L}^2$ as $j \to \infty$. $\triangle$

Example 8.9 (Integrating the tail). Let $f(x) = e^{-|x|}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, and take the sequence $f_n(x) = \frac{1}{n} f(\frac{x}{n})$. We can think of f_n as 'spreading' f by a factor of n. Then $f_n(x) \to 0$ for every $x \in \mathbb{R}$, and

$$\int_{\mathbb{R}} |f_n(x)|^2 \,\mathrm{d}m = \frac{1}{n} \int_{\mathbb{R}} |f(x)|^2 \,\mathrm{d}m \to 0,$$

so $f_n \to 0$ in $\mathcal{L}^2$ as $n \to \infty$. On the other hand,

$$\int_{\mathbb{R}} |f_n(x)| \,\mathrm{d}m = \int_{\mathbb{R}} |f(x)| \,\mathrm{d}m = 2 \not\to 0,$$

so $f_n \not\to 0$ in $\mathcal{L}^1$ as $n \to \infty$. Note that for every $t \in \mathbb{R}$, $0 \leqslant f_n(t) \leqslant \frac{1}{n}$, so $f_n \xrightarrow{a.e.} 0$. $\triangle$

8.1.2 Convergence of random variables

When the measure spaces are probability spaces, we use terminology more suitable to random variables. To this end, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\geqslant 1}, X$ be random variables.

(i) We say that $(X_n)_{n\geqslant 1}$ *converges to X $\mathbb{P}$-a.s.* if

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

In this case, we write $X_n \xrightarrow{\text{a.s.}} X$ (the measure $\mathbb{P}$ is implicit).

(ii) We say that $(X_n)_{n\geqslant 1}$ *converges to X in probability* if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

In this case, we write $X_n \xrightarrow{\mathbb{P}} X$.

(iii) Finally, we say that $(X_n)_{n\geqslant 1}$ *converges to X in $\mathcal{L}^p$* (or in p-th moment) if $X \in \mathcal{L}^p(\mathbb{P}),\ p \in [1,\infty)$, i.e. $\mathbb{E}|X|^p < \infty$, and

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0.$$

In this case, we write $X_n \xrightarrow{\mathcal{L}^p} X$.

Remark 8.10. For a probability space, if $X_n \xrightarrow{\mathcal{L}^p} X$, then $\mathbb{E}X_n \to \mathbb{E}X$. △

Proof. Using Exercise 5.6 and Jensen's inequality:

$$\big|\mathbb{E}X_n - \mathbb{E}X\big|^p \leqslant \big(\mathbb{E}|X_n - X|\big)^p \leqslant \mathbb{E}|X_n - X|^p \to 0. \qquad \square$$

Recall that a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a cumulative distribution function $F_X(x) = \mathbb{P}(X \leqslant x),\ x \in \mathbb{R}$. In particular, we can discuss convergence of a sequence of random variables via their distribution functions.

Definition 8.11 (Convergence in distribution)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\geqslant 1}$ and X be random variables with $(F_{X_n})_{n\geqslant 1}$ and F_X their corresponding distribution functions. We say that $(X_n)_{n\geqslant 1}$ *converges to X in distribution* if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for every point $x \in \mathbb{R}$ at which F_X is continuous. Then, we write $X_n \xrightarrow{d} X$.

Remark 8.12. Convergence in distribution is only concerned with the associated distribution functions. In particular, a sequence of random variables $(X_n)_{n\geqslant 1}$ converging in distribution to some random variable X might even be defined on a different probability space from that of X. △

Example 8.13. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\geqslant 1}$ a sequence of random variables, each Geometric distributed with probability of success p_n, that is,

$$F_{X_n}(x) = 1 - (1-p)^x,\ x \in \mathbb{N},\ n \geqslant 1.$$

Further, assume that $np_n \to \lambda > 0$, as $n \to \infty$. We would like to show that $(\frac{X_n}{n})_{n\geqslant 1}$ converges in distribution to the exponential distribution with parameter λ. First, for $x \in [0,\infty)$ and $n \geqslant 1$, we can write

$$F_{\frac{X_n}{n}}(x) = \mathbb{P}(X_n \leqslant nx) = \mathbb{P}(X_n \leqslant \lfloor nx \rfloor) =$$
$$= 1 - (1-p_n)^{\lfloor nx \rfloor} = 1 - (1 - \frac{np_n}{n})^{nx}(1 - \frac{np_n}{n})^{\lfloor nx \rfloor - nx}.$$

Now, by observing that

$$(1 - \frac{np_n}{n})^{nx} \xrightarrow{n\to\infty} e^{-\lambda x} \quad \text{and} \quad (1 - \frac{np_n}{n})^{\lfloor nx \rfloor - nx} \xrightarrow{n\to\infty} 1,$$

we get

$$F_{\frac{X_n}{n}}(x) \xrightarrow{n\to\infty} 1 - e^{\lambda x},$$

where the limiting distribution is precisely that of the exponential distribution with parameter λ, and so we are done. △

Example 8.14. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\geqslant 1}$ a sequence of random variables, each with $\frac{1}{n}$ and 1 being equiprobable outcomes, that is,

$$\mathbb{P}(X_n = \frac{1}{n}) = \mathbb{P}(X_n = 1) = \frac{1}{2},\ n \geqslant 1.$$

Under this construction, we can expect that $(X_n)_{n\geqslant 1}$ converges in distribution to a Bernoulli random variable X with parameter $\frac{1}{2}$. Indeed, for each $x \in \mathbb{R}$, we have

$$F_{X_n}(x) = \begin{cases} 0, & x < \frac{1}{n} \\ \frac{1}{2}, & x \in [\frac{1}{n}, 1) \\ 1, & x \geqslant 1 \end{cases} \xrightarrow[n\to\infty]{} \begin{cases} 0, & x \leqslant 0 \\ \frac{1}{2}, & x \in (0,1) \\ 1, & x \geqslant 1 \end{cases}$$

where the right hand side is precisely the distribution function of X, except at $x = 0$. But this is a discontinuity point of F_X, so we can still deduce that $X_n \xrightarrow{d} X$. △

Example 8.15. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\geqslant 1}$ be a sequence of random variables, each Cauchy distributed with location parameter 0 and scale parameter $\frac{1}{n}$, that is,

$$f_{X_n}(x) = \frac{1}{\frac{1}{n}\pi(1 + n^2x^2)} = \frac{n}{\pi(1+n^2x^2)},\ x \in \mathbb{R},\ n \geqslant 1.$$

From Figure 8.2, we can see that as n gets larger, $f_{X_n}(x)$ becomes closer to $f(x) = 0$, so we claim that $(X_n)_{n\geqslant 1}$ converges in distribution to 0. For $x \in \mathbb{R}$ and $n \geqslant 1$, we have

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) = \int_{-\infty}^{x} \frac{n}{\pi(1+n^2y^2)}\,\mathrm{d}y =$$
$$= \left[\frac{\arctan(ny)}{\pi}\right]_{-\infty}^{x} = \frac{1}{2} + \frac{\arctan(nx)}{\pi}$$

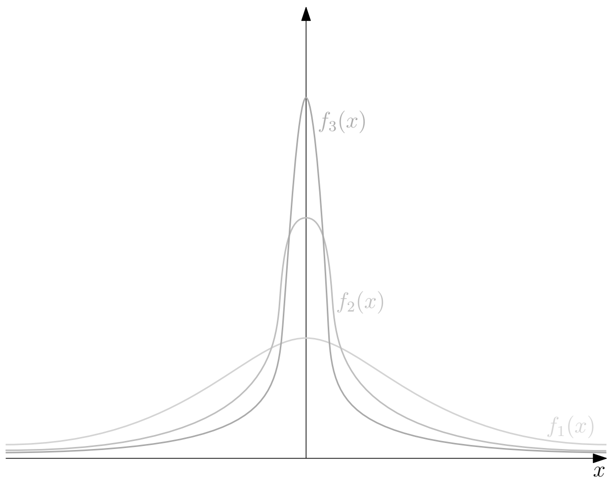


Figure 8.2: Probability density function of Cauchy distribution with location parameter 0 and scale parameter $n = 1, \frac{1}{2}, \frac{1}{3}$.

Now, taking limit as $n \to \infty$ gives

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & x = 0 \\ 1, & x > 0 \end{cases}$$

where the limiting distribution is that of a constant $X = 0$ random variable, except at $x = 0$, which is a point of discontinuity of F_X, and hence we can still deduce that $X_n \xrightarrow{d} 0$. In addition, we show convergence in probability. For each $\varepsilon > 0$, we have

$$\mathbb{P}(|X_n| > \varepsilon) = 2\int_{\varepsilon}^{\infty} \frac{n}{\pi(1+n^2x^2)}\,\mathrm{d}x = 2\left[\frac{\arctan(nx)}{\pi}\right]_{\varepsilon}^{\infty} = 1 - 2\frac{\arctan(n\varepsilon)}{\pi}.$$

Taking limit as $n \to \infty$ yields

$$\lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0,$$

thus, $X_n \xrightarrow{\mathbb{P}} 0$. △

Remark 8.16. In the above example, convergence in distribution follows from Proposition 8.42 that we will see later on. △

Theorem 8.17 (Convergence in distribution in terms of test functions)**.** *For a family of random variables X and $(X_n)_n$, we have $X_n \xrightarrow{d} X$ if and only if $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.*

Proof. See [Kle14, Theorem 13.23]. □

We shall later see that convergence in distribution is the weakest among these modes of convergence. However, it is quite often useful in practice.

8.2 Almost sure convergence and Borel-Cantelli

To get to convergence of real-valued random variables, for "most" or "almost all" points $\omega \in \Omega$, we need a systematic way of representing convergence $X_n \to X$ in terms of certain events that are amenable to study.

8.2.1 Infinitely often and eventually

Definition 8.18 (Infinitely often and eventually)**.** Suppose $(\Omega, \mathcal{F})$ is a measurable space and $(A_n)_{n\geqslant 1}$ is a sequence of events (i.e. measurable sets). We define the event "A_n infinitely often," abbreviated "A_n i.o.," also denoted $\limsup_{n\to\infty} A_n$, as

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

This event is the set of all ω such that $\omega \in A_n$ for infinitely many A_n's, that is, the set of ω such that, for all $n \geqslant 1$ there exists $m \geqslant n$ such that $\omega \in A_m$. Occurrence of the event $\limsup_n A_n$ means that infinitely many events among $\{A_n\}_n$ occur.

We define the event "A_n eventually," also denoted $\liminf_{n\to\infty} A_n$, as

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

This event is the set of all ω such that $\omega \in A_n$ for all but finitely many A_n's, that is, the set of ω such that, for some $n \geqslant 1$ and all $m \geqslant n$ one has $\omega \in A_m$. Occurrence of the event $\limsup_n A_n$ means that all but finitely many events among $\{A_n\}_n$ occur.

Remark 8.19. By De Morgan's laws, we have

$$\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ eventually}\} \quad \text{and} \quad \{A_n^c \text{ i.o.}\} = \{A_n \text{ eventually}\}^c.$$

Exercise 8.20. Prove the above identities. △

Let us point out already how the notions of eventually and infinitely often are related to convergence. A sequence of numbers $(x_n)_n$ converges to a number x if, for every $\varepsilon > 0$, we have $|x_n - x| < \varepsilon$ for all n sufficiently large, which means $|x_n - x| < \varepsilon$ for all but finitely many n's, which means $|x_n - x| < \varepsilon$ eventually. Likewise, $x_n \not\to x$ if for some $\varepsilon > 0$ we have $|x_n - x| \geqslant \varepsilon$ infinitely often.

Proposition 8.21 (Criterion for a.e. convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n\geqslant 1}, f$ be Borel functions. Then, $(f_n)_{n\geqslant 1}$ converges* *to f μ-a.e., if and only if*

$$\mu\big(\{|f_n - f| \geqslant \varepsilon \text{ i.o.}\}\big) = 0 \text{ for all } \varepsilon > 0.$$

Proof. We have the following chain of equivalent properties:

$$\begin{aligned}
\mu\{f_n \not\to f\} &= 0 \\
&\Updownarrow \text{ definition of limit} \\
\mu\{\exists \varepsilon > 0, |f_n - f| \geqslant \varepsilon \text{ i.o.}\} &= 0 \\
&\Updownarrow \text{ take } \tfrac{1}{k+1} < \varepsilon \leqslant \tfrac{1}{k} \\
\mu\{\exists k \in \mathbb{N}, |f_n - f| \geqslant \tfrac{1}{k} \text{ i.o.}\} &= 0 \\
&\Updownarrow \text{ continuity from below of } \mu \\
\lim_{k\to\infty} \mu\{|f_n - f| \geqslant \tfrac{1}{k} \text{ i.o.}\} &= 0 \\
&\Updownarrow \text{ non-decreasing numbers} \\
\forall k \in \mathbb{N}, \mu\{|f_n - f| \geqslant \tfrac{1}{k} \text{ i.o.}\} &= 0 \\
&\Updownarrow \text{ take } \tfrac{1}{k+1} < \varepsilon \leqslant \tfrac{1}{k} \\
\forall \varepsilon > 0, \mu\{|f_n - f| \geqslant \varepsilon \text{ i.o.}\} &= 0.
\end{aligned}$$

This completes the proof. □

8.2.2 Borel-Cantelli lemmas

From the above proposition, we have enough motivation to try and estimate the probability that a sequence of events occur infinitely often.

Lemma 8.22 (Borel-Cantelli Lemma I)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(A_n)_{n\geqslant 1}$ a sequence of measurable sets. If*

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty,$$

then

$$\mu\big(\limsup_n A_n\big) = 0.$$

Proof. The idea is to set appropriate upper bounds that converge to 0, so as make the relevant conclusions. Using continuity from above and countable subadditivity of μ, we have

$$\mu(\{\limsup_n A_n\}) = \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right)$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \mu\left(\bigcup_{m=n}^{\infty} A_m\right) \\
&\leqslant \lim_{n\to\infty} \sum_{m=n}^{\infty} \mu(A_m) = 0.
\end{aligned}
$$

The last limit is zero because it is the tail of a convergent series. □

Lemma 8.23 (Borel-Cantelli Lemma II)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n)_{n\geqslant 1}$ a sequence of independent events. If*

$$
\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty,
$$

then

$$
\mathbb{P}\big(\limsup_n A_n\big) = 1.
$$

Proof. We look instead at $\{A_n \text{ i.o.}\}^c$. By Remark 8.19, continuity from below and from above, and independence of $(A_n^c)_{n\geqslant 1}$, we get

$$
\begin{aligned}
\mathbb{P}(\{A_n \text{ i.o.}\}^c) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^c\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) \\
&= \lim_{n\to\infty}\lim_{k\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{k} A_m^c\right) \\
&= \lim_{n\to\infty}\lim_{k\to\infty} \prod_{m=n}^{k} (1 - \mathbb{P}(A_m)) \\
&\leqslant \lim_{n\to\infty}\lim_{k\to\infty} \prod_{m=n}^{k} e^{-\mathbb{P}(A_m)} \\
&= \lim_{n\to\infty}\lim_{k\to\infty} e^{-\sum_{m=n}^{k} \mathbb{P}(A_m)} = e^{-\infty} = 0,
\end{aligned}
$$

where the upper bound follows from inequality $1 - x \leqslant e^{-x}$ for $x \geqslant 0$. □

Remark 8.24. In practice, we can apply Borel-Cantelli Lemma I to deduce almost everywhere convergence, that is, if we can show that for every $\varepsilon > 0$,

$$
\sum_{n=1}^{\infty} \mu\big(|f_n - f| \geqslant \varepsilon\big) < \infty
$$

then by Proposition 8.21 we can conclude that $f_n \xrightarrow{\text{a.e.}} f$. △

Example 8.25. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n \geqslant 1}$ be a sequence of independent random variables, each uniformly distributed on $[0, 1]$. Consider the sequence of random variables $(Y_n)_{n \geqslant 1}$ given by

$$Y_n = \min\{X_1, \ldots, X_n\}, \ n \geqslant 1.$$

We show that $Y_n \xrightarrow{\text{a.s.}} 0$. Given $n \in \mathbb{N}$, independence and construction of $X_1, \ldots, X_n$ for each $\varepsilon \in (0, 1)$ yields

$$\begin{aligned} \mathbb{P}(|Y_n - Y| \geqslant \varepsilon) = \mathbb{P}(|Y_n| > \varepsilon) &= \mathbb{P}(X_1 > \varepsilon, \ldots, X_n > \varepsilon) \\ &= \mathbb{P}(X_1 > \varepsilon)^n \\ &= (1 - \varepsilon)^n, \end{aligned}$$

and consequently, we get that

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n| \geqslant \varepsilon) = \sum_{n=1}^{\infty} (1 - \varepsilon)^n = \frac{1 - \varepsilon}{\varepsilon} < \infty.$$

Hence, for each $\varepsilon \in (0, 1)$, by Borel-Cantelli Lemma I, it follows that

$$\mathbb{P}(\{|Y_n| > \varepsilon \text{ i.o.}\}) = 0,$$

so by Proposition 8.21, we have $Y_n \xrightarrow{\text{a.s.}} 0$. △

Exercise 8.26. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n \geqslant 1}$ be a sequence of random variables with probability distribution given by

$$\mathbb{P}(X_n = a_n) = p_n \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - p_n, \quad n \geqslant 1,$$

for sequences $(a_n)_n$ and $(p_n)_n$ with $\sum_{n=1}^{\infty} p_n < \infty$. Show that $X_n \xrightarrow{\text{a.s.}} 0$. △

Example 8.27. Suppose $a_n \uparrow \infty$ and $p_n \downarrow 0$ with $\sum_n p_n = \infty$, and that $(X_n)_n$ is a sequence of independent random variables distributed as

$$\mathbb{P}(X_n = a_n) = p_n \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - p_n, \quad n \geqslant 1.$$

Then $\mathbb{P}(X_n \to 0) = 0$ but $X_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. △

Remark 8.28. When applying Borel-Cantelli Lemma II to argue that almost sure convergence does not hold, we must be careful to check independence of the concerned random variables. △

Example 8.29. Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}, m_{|_{[0,1]}})$, $X_n(\omega) = \mathbb{1}_{[0, \frac{1}{n}]}(\omega)$ and $X(\omega) = 0 \ \forall \omega$. Then $\mathbb{P}(X_n \to 0) = 1$ since $X_n(\omega) \to 0$ for all $\omega \neq 0$. However, $\mathbb{P}(X_n = 1) = \frac{1}{n}$ which is non-summable. If we used the same reasoning as in the previous example, without noticing that the variables $(X_n)_n$ are dependent, we would conclude that $\mathbb{P}(X_n \to 0) = 0$, which is false. △

Proposition 8.30. *If* $\mathbb{P}(A_n \textit{ i.o.}) = 0$ *then* $\mathbb{P}(A_n) \to 0$ *as* $n \to \infty$.

Proof. Suppose $\mathbb{P}(A_n \text{ i.o.}) = 0$, and take $B_k = \cup_{n \geqslant k} A_n$. Since $B_k \downarrow [A_n \text{ i.o.}]$ as $k \to \infty$, we have $\mathbb{P}(B_k) \to \mathbb{P}(A_n \text{ i.o.}) = 0$. On the other hand, $B_k \supseteq A_k$, thus $\mathbb{P}(A_k) \leqslant \mathbb{P}(B_k)$, and therefore $\mathbb{P}(A_n) \to 0$ as $n \to \infty$. □

8.3 Laws of Large Numbers

The topic of Laws of Large number dispenses introductions.

Let $X_1, X_2, X_3, \dots$ be a sequence of random variables. Define

$$S_n := X_1 + \cdots + X_n.$$

We say that the sequence $(X_n)_n$ satisfies the law of large numbers if

$$\frac{S_n - \mathbb{E}S_n}{n} \to 0$$

as $n \to \infty$. We say that $(X_n)_n$ satisfies weak LLN if the above convergence is in probability, and the strong LLN in case convergence is almost-sure.

8.3.1 Laws of Large Numbers with second and fourth moments

Theorem 8.31 (Chebyshev's Weak Law of Large Numbers)**.** *Let $(X_n)_n$ be a sequence of random variables such that $\mathrm{Cov}(X_i, X_j) = 0$ for every $i \neq j$ and $\sup_n \mathbb{V}X_n < \infty$. Then $(X_n)_n$ satisfies the weak law of large numbers.*

Proof. By Chebyshev's inequality (5.28),

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{V}(\frac{S_n}{n})}{\varepsilon^2} = \frac{\mathbb{V}S_n}{\varepsilon^2 n^2} = \frac{\sum_{i=1}^n \mathbb{V}X_i}{\varepsilon^2 n^2} \leqslant \frac{n \cdot \sup_k \mathbb{V}X_k}{\varepsilon^2 n^2} \to 0,$$

concluding the proof. □

Theorem 8.32 (Cantelli's Strong Law of Large Numbers)**.** *Let $X_1, X_2, \dots$ be independent random variables such that $\sup_n \mathbb{E}X_n^4 < \infty$. Then $(X_n)_n$ satisfies the strong law of large numbers.*

Proof. We can suppose that $\mathbb{E}X_k = 0$ for all k (otherwise consider $X - \mathbb{E}X$ and use inequalities $|x| \leqslant 1 + x^4$ and $|x - c|^4 \leqslant |2x|^4 + |2c|^4$). Expand S_n^4 as

$$S_n^4 = (X_1 + \cdots + X_n)^4 = \sum_{i,j,k,l} X_i X_j X_k X_l = \sum_i X_i^4 + \frac{4!}{2!2!} \sum_{i<j} X_i^2 X_j^2 + \\ + \frac{4!}{3!} \sum_{i \neq k} X_i^3 X_k + \frac{4!}{2!} \sum_{\substack{j<k \\ i \neq j,k}} X_i^2 X_j X_k + 4! \sum_{i<j<k<l} X_i X_j X_k X_l.$$

By independence, we have

$$\begin{aligned}\mathbb{E}S_n^4 = \sum_i \mathbb{E}X_i^4 + 6\sum_{i<j}\mathbb{E}[X_i^2X_j^2] + \\ + \sum_k \mathbb{E}\left[4\sum X_i^3 + 12\sum X_i^2X_j + 24\sum X_iX_jX_l\right]\mathbb{E}X_k.\end{aligned}$$

Since $\mathbb{E}X_k = 0$, the second row is zero. Write $M := \sup_k \mathbb{E}X_k^4 < \infty$. By the Cauchy-Schwarz inequality, $\mathbb{E}(X_i^2X_j^2) \leqslant \sqrt{M}\sqrt{M} = M$. Hence,

$$\mathbb{E}S_n^4 \leqslant nM + 6\tbinom{n}{2}M = (3n^2 - 2n)M \leqslant 3n^2M.$$

By Markov's inequality, we have

$$\mathbb{P}\left(\left|\tfrac{S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{E}S_n^4}{\varepsilon^4 n^4} \leqslant \frac{3M}{\varepsilon^4 n^2},$$

and by Borel-Cantelli Lemma I we get $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$. □

8.3.2 Laws of Large Numbers with first moment

Theorem 8.33 (Kolmogorov's Strong Law of Large numbers)**.** *Let $X_1, X_2, \dots$ be pairwise independent and identically distributed integrable random variables. Then $(X_n)_n$ satisfies the strong law of large numbers.*

Proof. See [Kle14, Theorem 5.17] for the proof found by Etemadi in 1981. □

It should be emphasised that, when $(X_n)_n$ are i.i.d., the condition of X_1 being integrable is the weakest possible. It is in fact equivalent to $\frac{X_1+\cdots+X_n}{n} \overset{a.s.}{\to} \mu$ for some $\mu \in \mathbb{R}$, as shown by the following proposition.

Proposition 8.34. *Suppose $(X_n)_n$ are i.i.d. and not integrable. Then, a.s., $\limsup_n \frac{|S_n|}{n} = +\infty$ and $\frac{S_n}{n}$ does not converge to any real number.*

Proof. For every fixed $k \in \mathbb{N}$, the variable $\frac{X_1}{k}$ is not integrable. Since $(X_n)_n$ are identically distributed, by Proposition 5.39 we have

$$\sum_n \mathbb{P}\left(\tfrac{|X_n|}{k} \geqslant n\right) = \sum_n \mathbb{P}\left(\tfrac{|X_1|}{k} \geqslant n\right) = \infty.$$

By Borel-Cantelli Lemma II, almost surely the event $|X_n| \geqslant kn$ occurs for infinitely many values of n. Taking the complement and union over k, we get

$$\mathbb{P}\big(\text{for every } k,\ |X_n| \geqslant kn \text{ for infinitely many values of } n\big) = 1.$$

We now show that, for a sequence $(x_n)_n$ satisfying the above condition, we have $\limsup_n |\frac{s_n}{n}| = +\infty$, where $s_n := x_1 + \cdots + x_n$. This implies the proposition.

So we have to show that, for every $r \in \mathbb{N}$, there exists m such that $|\frac{s_m}{m}| \geqslant r$. Suppose $(x_n)_n$ satisfies the above condition and let $r \in \mathbb{N}$ be given. Take $k = 2r$ and n such that $|x_n| \geqslant kn$. If $|\frac{s_{n-1}}{n-1}| \geqslant r$, we take $m = n-1$ and we are done. So suppose $|\frac{s_{n-1}}{n-1}| \leqslant r$. Then

$$\begin{aligned}\frac{|x_1 + \cdots + x_n|}{n} \geqslant \frac{|x_n|}{n} - \frac{|x_1 + \cdots + x_{n-1}|}{n} \geqslant \\ \geqslant \frac{|x_n|}{n} - \frac{|x_1 + \cdots + x_{n-1}|}{n-1} \geqslant \frac{|x_n|}{n} - r \geqslant k - r = r,\end{aligned}$$

and we can take $m = n$. This concludes the proof. ☐

8.4 Relation between modes of convergence

8.4.1 Almost-sure convergence and subsequences

Proposition 8.35 (Convergence in measure implies a.e. in a subsequence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n\geqslant 1}, f$ Borel functions with $f_n \xrightarrow{\mu} f$. Then, there exists a subsequence such that $f_{n_k} \xrightarrow{a.e.} f$ as $k \to \infty$.*

Proof. First for each $n \geqslant 1$ and $k \in \mathbb{N}$, set $A_{n,k} := \{|f_n - f| > \frac{1}{k}\}$. Since $f_n \xrightarrow{\mu} f$, then for each $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mu(A_{n,k}) = \lim_{n\to\infty} \mu(|f_n - f| > \tfrac{1}{k}) = 0.$$

In other words, for each $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ such that for any $n \geqslant n_k$,

$$\mu(A_{n,k}) = \mu(|f_n - f| > \tfrac{1}{k}) < \tfrac{1}{k^2}.$$

We can also assume that $n_k > n_{k-1}$ for $k \geqslant 2$. Set $B_k := A_{n_k,k}$, and notice that

$$\sum_{k=1}^{\infty} \mu(B_k) < \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

and so $\mu(B_k \text{ i.o.}) = 0$ by Borel-Cantelli Lemma I, thus $f_{n_k} \xrightarrow{\text{a.e.}} f$, as $k \to \infty$. ☐

Convergence a.e. does not imply convergence in measure, as shown in Example 8.4. However, this is the case in finite measure spaces.

Proposition 8.36 (a.s. convergence implies convergence in probability)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n\geqslant 1}, X$ be random variables. If $X_n \to X$ a.s. as $n \to \infty$, then $X_n \to X$ in probability as $n \to \infty$.*

Proof. Since $X_n \xrightarrow{\text{a.s.}} X$, by Proposition 8.21, we have

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon \text{ i.o.}) = 0, \ \forall \varepsilon > 0.$$

Then, we can apply Proposition 8.30 to get

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geqslant \varepsilon) = 0, \ \forall \varepsilon > 0,$$

so that $X_n \xrightarrow{\mathbb{P}} X$, as required. □

8.4.2 Convergence in $\mathcal{L}^p$ and in expectation

Proposition 8.37 (Convergence in $\mathcal{L}^p$ implies convergence in measure)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(f_n)_{n\geqslant 1}, f$ be Borel functions. If $f_n \to f$ in $\mathcal{L}^p$ as $n \to \infty$, then $f_n \to f$ in measure as $n \to \infty$.*

Proof. Since $f_n \xrightarrow{\mathcal{L}^p} f$, for some $p \geqslant 1$, then

$$\lim_{n\to\infty} \int_\Omega |f_n - f|^p \, \mathrm{d}\mu = 0.$$

By Chebyshev's inequality (Theorem 5.26), for all $\varepsilon > 0$, it follows that

$$\mu(|f_n - f| > \varepsilon) \leqslant \frac{1}{\varepsilon^p} \int_\Omega |f_n - f|^p \, \mathrm{d}\mu \to 0$$

as $n \to \infty$, thus f_n converges to f in measure, i.e. $f_n \xrightarrow{\mu} f$. □

Proposition 8.38 (Comparison of $\mathcal{L}^p$ and $\mathcal{L}^q$)**.** *For a probability space, if $q \geqslant p \geqslant 1$ and $X_n \xrightarrow{\mathcal{L}^q} X$, then $X_n \xrightarrow{\mathcal{L}^p} X$.*

Proof. The proof goes as follows:

$$(\mathbb{E}|X_n - X|^p)^{1/p} \leqslant (\mathbb{E}|X_n - X|^q)^{1/q} \to 0,$$

so we only need to justify the first inequality. Let Y be a non-negative random variable. Take $(Z_n)_n$ as integrable random variables such that $0 \leqslant Z_n \uparrow Y^p$

(for instance, $Z_n = \min\{Y,n\}^p$). The function $f : [0,\infty) \to [0,\infty)$ given by $f(x) = x^{q/p}$ is convex. Hence, by Jensen's inequality, we have

$$(\mathbb{E}Z_n)^{q/p} \leqslant \mathbb{E}Z_n^{q/p} \leqslant \mathbb{E}Y^q.$$

Now, we can conclude that $(\mathbb{E}Y^p)^{q/p} \leqslant \mathbb{E}Y^q$ by the Monotone Convergence Theorem. This completes the proof. □

Example 8.39. Let $(\Omega,\mathcal{F},\mathbb{P}) = ([0,1],\mathcal{B},m)$, and take $X_n = n\mathbb{1}_{[0,\frac{1}{n})}$. Then $X_n \to 0$ a.s., but $\int_\Omega |X_n|\,\mathrm{d}\mathbb{P} = 1$ so $X_n \not\to 0$ in $\mathcal{L}^p$. △

Examples 8.8 and 8.39 show that neither a.s. convergence implies convergence in $\mathcal{L}^p$, nor convergence in $\mathcal{L}^p$ implies a.s. convergence. Combined with Propositions 8.36 and 8.37, convergence in probability implies neither of them.

Theorem 8.40 (Dominated Convergence Theorem revisited)**.** *Let $(\Omega,\mathcal{F},\mu)$ be a measure space, and $(f_n)_{n\geqslant 1}$ a sequence of Borel functions such that $f_n \xrightarrow{\mu} f$. If there exists a non-negative Borel function $g \in \mathcal{L}^p(\mu)$, $p \in [1,\infty)$, such that $|f_n| \leqslant g$ for all $n \geqslant 1$, then $f_n \xrightarrow{\mathcal{L}^p} f$.*

If moreover $p = 1$ or μ is a probability measure, then

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

Proof. Convergence of the integral follows from Remark 8.10 if $\mu(\Omega) = 1$, and from Exercise 5.6 if $p = 1$, so we only need to prove that $f_n \to f$ in $\mathcal{L}^p$.

Let us first suppose that $f_n \xrightarrow{\text{a.e.}} f$. By noticing that

$$|f_n(\omega) - f(\omega)|^p \leqslant 2^p|g(\omega)|^p,\ \omega \in \Omega,$$

we can use Theorem 5.46 to deduce that

$$\lim_{n\to\infty} \int_\Omega |f_n - f|^p\,\mathrm{d}\mu = 0. \tag{8.41}$$

Now, let us assume that $f_n \xrightarrow{\mu} f$. Consider a subsequence $(f_{n_k})_k$. Then $f_{n_k} \xrightarrow{\mu} f$ as $k \to \infty$. By Proposition 8.35, there is a sub-subsequence $(f_{n_{k_j}})_j$ such that $f_{n_{k_j}} \xrightarrow{\text{a.e.}} f$ as $j \to \infty$. By the previous case, we have $\int_\Omega |f_{n_{k_j}} - f|^p \to 0$ as $j \to \infty$. Using Lemma A.1, we get (8.41), concluding the proof. □

8.4.3 Convergence in distribution

Convergence in distribution is actually the weakest concept, and it is implied by all modes of convergence.

Proposition 8.42 (Convergence in probability implies convergence in distribution)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_n)_{n \geqslant 1}, X$ be random variables. If $X_n \to X$ in probability, then $X_n \to X$ in distribution as $n \to \infty$.*

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. By Theorem 8.17, it is enough to show that

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \tag{8.43}$$

as $n \to \infty$. First, suppose that $X_n \xrightarrow{\text{a.s.}} X$. Recall that a real function is continuous if and only if $f(x_n) \to f(x)$ whenever $x_n \to x$. Thus,

$$\{\omega : X_n(\omega) \to X(\omega)\} \subseteq \{\omega : f(X_n(\omega)) \to f(X(\omega))\},$$

and hence $f(X_n) \xrightarrow{\text{a.s.}} f(X)$. Since f is bounded, $|f(X_n)| \leqslant \sup_{s \in \mathbb{R}} |f(s)| < \infty$ for all $n \geqslant 1$. Applying Theorem 8.40 with constant g, we get (8.43).

If instead we suppose $X_n \xrightarrow{\mathbb{P}} X$, we can get (8.43) by taking sub-subsequences and invoking Lemma A.1, as we did in the proof of Theorem 8.40. □

Example 8.44. Let $\Omega = \{0, 1\}$, $\mathbb{P} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, $X_n(\omega) = \omega$ and $Y(\omega) = 1 - \omega$. Then, as $n \to \infty$, $X_n \to Y$ in distribution but not in probability. △

So convergence in distribution does not imply convergence in probability in general, but there is one situation where it does.

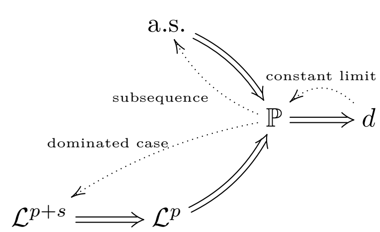
Proposition 8.45 (Convergence in distribution to constant limit)**.** *Suppose $X_n \xrightarrow{d} X$ and $X = c$ a.s. for some $c \in \mathbb{R}$. Then $X_n \xrightarrow{\mathbb{P}} c$.*

Proof. First, notice that by assumption $F_{X_n}(x) \to 0$ if $x < c$ and $F_{X_n}(x) \to 1$ if $x > c$. Let $\varepsilon > 0$. Then

$$\begin{aligned}\mathbb{P}(|X_n - c| > \varepsilon) \leqslant \mathbb{P}(X_n \leqslant c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon) = \\ = F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon) \to 0 + 1 - 1 = 0,\end{aligned}$$

so $X_n \xrightarrow{\mathbb{P}} c$, as claimed. □

In summary, for probability spaces, the relationship between the four different notions of convergence are represented below:



The general implications shown in the diagram are Propositions 8.36, 8.37, 8.38 and 8.42. No other implication holds in general, as shown by Examples 8.8, 8.39 and 8.44. The dotted lines refer to implications that hold in particular cases, which are Propositions 8.35 and 8.45, and Theorem 8.40.

A Preliminaries

Countable sets

See [Coh13, §A.6].

Extended real numbers

See §1.3.1 and [Coh13, §B.4].

Supremum and infimum

See [Coh13, §B.5].

liminf, limsup, limit

See [Coh13, §B.6].

One criterion of convergence is the following.

Lemma A.1. *Let $(a_n)_n$ be a sequence of real numbers, and $a \in \mathbb{R}$. Then $\lim_n a_n = a$ if and only if, for every subsequence $(a_{n_k})_k$, there is a further subsequence $(a_{n_{k_j}})_j$ such that $\lim_j a_{n_{k_j}} = a$.*

Sketch of proof. Suppose $a_n \to a$ and let a_{n_k} be a subsequence. For each $\varepsilon > 0$, $\{n : |a_n - a| \geqslant \varepsilon\}$ is finite, so $\{k : |a_{n_k} - a| \geqslant \varepsilon\}$ is also finite, hence $a_{n_k} \to a$. Now suppose $a_n \not\to a$, so there is $\varepsilon > 0$ such that $\{n : |a_n - a| \geqslant \varepsilon\}$ is infinite. Taking $(n_k)_k$ from this set, no sub-subsequence $(a_{n_{k_j}})$ may converge to a. □

Open and closed sets

Given a metric space (Ω, d) and a subset $B \subseteq \Omega$, the *interior* of B is the set of points $x \in \Omega$ for which there is a ball centred at x and contained in B. A set $B \subseteq \mathbb{R}$ is called *open* if the interior of B equals B. A set $B \subseteq \mathbb{R}$ is called *closed* if B^c is open. The sets Ω and $\emptyset$ are both opened and closed.

When $\Omega = \mathbb{R}$, the interior of $B \subseteq \mathbb{R}$ is the set $\{x \in B : \exists \varepsilon > 0, (x - \varepsilon, x + \varepsilon) \subseteq B\}$. Also, open intervals are open and closed intervals are closed. The only sets that are open and closed at the same time are $\mathbb{R}$ and $\emptyset$. Every open set is given by the union of countably many disjoint open intervals.

Given two metric spaces (Ω, d) and (Ω', d') and a function $f : \Omega \to \Omega'$, we say that f is *continuous* if the preimage of every open set is open.

Riemann integral - Darboux's definition

See [Coh13, §p.xv].

B Postponed proofs

B.0.1 The π-λ Theorem

Lemma B.1. *An algebra closed under countable disjoint unions is a σ-algebra.*

Proof. Exercise (see the proof or Proposition 2.35). $\square$

Lemma B.2. *A λ-system closed under intersections is a σ-algebra.*

Proof. Exercise (use the previous lemma). $\square$

Proof of Theorem 3.6. We define $\mathcal{G}$ as the smallest λ-system[18] that contains the π-system $\mathcal{C}$, and show that $\mathcal{G}$ is also a π-system. By Lemma B.2, $\mathcal{G}$ is a σ-algebra, hence $\mathcal{C} \subseteq \sigma(\mathcal{C}) \subseteq \mathcal{G} \subseteq \mathcal{D}$, proving the theorem.

The key idea is to consider, for each $B \in \mathcal{G}$, the collection

$$\mathcal{F}_B := \{A \in \mathcal{G} : A \cap B \in \mathcal{G}\}.$$

We can see that $\mathcal{F}_B$ is a λ-system using $A^c \cap B = ((A \cap B) \cup B^c)^c$.

Now, let $B \in \mathcal{C}$. Since $\mathcal{C}$ is a π-system, $\mathcal{C} \subseteq \mathcal{F}_B$. But $\mathcal{G}$ is the smallest λ-system that contains $\mathcal{C}$, so $\mathcal{F}_B = \mathcal{G}$. Finally, let $D \in \mathcal{G}$. For every $B \in \mathcal{C}$, we have $D \in \mathcal{F}_B$, which is equivalent to $B \in \mathcal{F}_D$. Hence, $\mathcal{C} \subseteq \mathcal{F}_D$. But $\mathcal{G}$ is the smallest λ-system that contains $\mathcal{C}$, so $\mathcal{F}_D = \mathcal{G}$. Since this is true for every $D \in \mathcal{G}$, we conclude that $\mathcal{G}$ is a π-system, which completes the proof.[19] $\square$

B.0.2 Carathéodory Extension Theorem

Proof of Lemma 3.11. That $\mu^*(E) \leqslant \mu^*(F)$ when $E \subseteq F$ follows from $\mathcal{C}_F \subseteq \mathcal{C}_E$.

We now prove that $\mu^*(\cup_n E_n) \leqslant \sum_n \mu^*(E_n)$. Let $\varepsilon > 0$, for each $n \in \mathbb{N}$, by definition of $\mu^*(E_n)$ we can take $(A_{n,k})_k$ in $\mathcal{A}$ such that $E_n \subseteq \cup_k A_{n,k}$ and $\sum_k \mu(A_{n,k}) \leqslant \mu^*(E_n) + \varepsilon 2^{-n}$. Since $\cup_n E_n \subseteq \cup_{n,k} A_{n,k}$, by definition of $\mu^*(\cup_n E_n)$ we have $\mu^*(\cup_n E_n) \leqslant \sum_{n,k} \mu(A_{n,k}) = \sum_n \sum_k \mu(A_{n,k}) \leqslant \sum_n \mu^*(E_n) + \varepsilon 2^{-n} = \sum_n \mu^*(E_n) + \varepsilon$. Since ε is arbitrary, the inequality follows.

We finally prove that $\mu^*(A) = \mu(A)$ for $A \in \mathcal{A}$. That $\mu^*(A) \leqslant \mu(A)$ follows immediately by taking $(A, \emptyset, \emptyset, \emptyset \dots) \in \mathcal{C}_A$. We show that $\mu^*(A) \geqslant \mu(A)$ using the assumption that μ is σ-additive on $\mathcal{A}$. Let $(A_n)_n \in \mathcal{C}_A$. Take $B_n = A \cap A_n \cap (\cap_{k=1}^{n-1} A_k^c)$, so they form a partition of A. Hence, $\mu(A) = \sum_n \mu(B_n) \leqslant \sum_n \mu(A_n)$. Since this holds for any $(A_n)_n \in \mathcal{C}_A$, we have $\mu(A) \leqslant \mu^*(A)$, concluding the proof.[20] $\square$

[18] This is defined in the same way the smallest σ-algebra that contains a given class.
[19] Based on https://almostsuremath.com/2019/10/06/the-monotone-class-theorem/.
[20] Adapted from [Kub15, Proposition 8.3].

Proof of Lemma 3.12. We make extensive use of Lemma 3.11. Let $\mathcal{F} = \mathcal{F}^*$.

We first prove that $\mathcal{A} \subseteq \mathcal{F}$. Let $A \in \mathcal{A}$ and $E \subseteq \Omega$. For $\varepsilon > 0$, by definition of μ^* there are $A_1, A_2, \dots \in \mathcal{A}$ with $E \subseteq \cup_n A_n$ and $\sum_n \mu(A_n) \leqslant \mu^*(E) + \varepsilon$. Then

$$\mu^*(E) \leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c) \leqslant \sum_n \mu^*(A_n \cap A) + \sum_n \mu^*(A_n \cap A^c) =$$
$$= \sum_n \mu(A_n \cap A) + \sum_n \mu(A_n \cap A^c) = \sum_n \mu(A_n) \leqslant \mu^*(E) + \varepsilon.$$

We used sub-additivity of μ^* twice, the fact that $\mu^* = \mu$ on $\mathcal{A}$, and finite additivity of μ. Since ε is arbitrary, we have $A \in \mathcal{F}$. Hence, $\mathcal{A} \subseteq \mathcal{F}$.

We now prove that $\mathcal{F}$ is a σ-algebra and μ^* is σ-additive on $\mathcal{F}$.

Step 1. The class $\mathcal{F}$ is an algebra.

Trivially, $\Omega \in \mathcal{F}$ and $A^c \in \mathcal{F}$ for all $A \in \mathcal{F}$. Now for $A, B \in \mathcal{F}$, using sub-additivity we get $\mu^*(E) \leqslant \mu^*(E \cap (A \cap B)) + \mu^*(A \cap (A \cap B)^c) \leqslant \mu^*(E \cap A \cap B) + \mu(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) = \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E)$, so $A \cap B \in \mathcal{F}$. Hence, $\mathcal{F}$ is an algebra.

Step 2. For disjoint $A_1, \dots, A_n \in \mathcal{F}$,

$$\mu^*\big(E \cap (\cup_{k=1}^n A_k)\big) = \sum_{k=1}^n \mu^*(E \cap A_k).$$

Indeed, noting that $(\cup_{k=1}^n A_k) \cap A_1 = A_1$ and $(\cup_{k=1}^n A_k) \cap A_1^c = \cup_{k=2}^n A_k$, using the definition of $A_1 \in \mathcal{F}$ with $E \cap (\cup_{k=1}^n A_k)$ instead of E, we get the identity $\mu^*\big(E \cap (\cup_{k=1}^n A_k)\big) = \mu^*\big(E \cap A_1\big) + \mu^*\big(E \cap (\cup_{k=2}^n A_k)\big)$. Applying the same argument to A_2, then A_3, and so on, we get the above identity.

Step 3. For $A_1, A_2, \dots \in \mathcal{F}$ disjoint, $\cup_n A_n \in \mathcal{F}$ and $\mu^*(\cup_n A_n) = \sum_n \mu^*(A_n)$.

These are both proved by showing

$$\mu^*(E) \leqslant \mu^*(E \cap G) + \mu^*(E \cap G^c) \leqslant \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap G^c) \leqslant \mu^*(E),$$

where $G = \cup_n A_n$. The above chain readily implies that $\cup_n A_n \in \mathcal{F}$, and $\mu^*(\cup_n A_n) = \sum_n \mu^*(A_n)$ follows by taking $E = \cup_n A_n$. The first two inequalities are sub-additivity. For the third inequality, writing $G_n := \cup_{k=1}^n A_k$, we have $\mu^*(E) = \mu^*(E \cap G_n) + \mu^*(E \cap G_n^c) = \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap G_n^c) \geqslant \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap G^c) \to \sum_{k=1}^\infty \mu^*(E \cap A_k) + \mu^*(E \cap G^c)$, where the first equality holds since $G_n \in \mathcal{F}$ by Step 1, and and second equality follows from Step 2.

By Lemma B.1, this concludes the proof.[21] $\square$

[21] Adapted from [Kub15, Theorem 8.4].

Proof of Lemma 3.15. Suppose $(C_n)_n$ is a sequence of disjoint sets in $\mathcal{A}$ such that $C_0 := \cup_n C_n \in \mathcal{A}$. We have to prove that $m(C_0) = \sum_n m(C_n)$. By writing $C_n^k = C_n \cap (k, k+1]$, we have $m(C_n) = \sum_{k\in\mathbb{Z}} m(C_n^k)$. Indeed, when C_n is bounded this follows from finite additivity of m on $\mathcal{A}$ and when C_n is unbounded both sides are infinite. Hence, by Theorem 1.13, it is enough to show $m(C_0^k) = \sum_n m(C_n^k)$ for each k. For simplicity we pick $k = 0$, that is, we can assume that $C_n \subseteq (0, 1]$ for all n. By taking $A_n = C_0 \setminus (C_1 \cup \cdots \cup C_n)$, it is enough to prove that for every decreasing sequence A_n of sets in $\mathcal{A}$ such that A_1 is bounded, if $A_n \downarrow \emptyset$ then $m(A_n) \to 0$.

We will prove the contrapositive, that is, if $m(A_n) \not\to 0$ then $\cap_n A_n \neq \emptyset$. So suppose for some $\varepsilon > 0$, we have $m(A_n) \geqslant 2\varepsilon$ for all n. For each $k \in \mathbb{N}$, we can find $B_k \in \mathcal{A}$ whose closure satisfies $\overline{B_k} \subseteq A_k$ and $m(A_k \setminus B_k) \leqslant \varepsilon 2^{-k}$ (just make them slightly shorter on the left than A_k). Then for each n, $m(\cup_{k=1}^n (A_k \setminus B_k) \leqslant \varepsilon$ by additivity of m on $\mathcal{A}$. On the other hand, $A_n \setminus \cap_{k=1}^n B_k = \cup_{k=1}^n A_n \setminus B_k \subseteq \cup_{k=1}^n A_k \setminus B_k$, so $m(A_n \setminus \cap_{k=1}^n B_k) \leqslant \varepsilon$, and since $m(A_n) \geqslant 2\varepsilon$, we have $m(\cap_{k=1}^n B_k) \geqslant \varepsilon$. In particular, $K_n := \cap_{k=1}^n \overline{B_k} \neq \emptyset$. Since K_n are compact sets by Heine-Borel Theorem, $\cap_{n=1}^\infty K_n \neq \emptyset$. Hence, $\emptyset \neq \cap_{n=1}^\infty K_n = \cap_{n=1}^\infty \overline{B_n} \subseteq \cap_{n=1}^\infty A_n$, concluding the proof.[22] $\square$

B.0.3 Hahn Decomposition Theorem

We are going to use the concept of a signed measure on our way towards proving the Radon-Nikodým Theorem. A *signed measure* on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to \overline{\mathbb{R}}$ such that $\mu(\emptyset) = 0$ and, for disjoint $(A_n)_n$ in $\mathcal{F}$, $\sum_n \mu(A_n)$ is defined and equals $\mu(\cup_n A_n)$. That is, from the definition of measure we drop the requirement that $\mu(A) \geqslant 0$.

If μ_1 is a measure and μ_2 is a finite measure, then $\nu = \mu_1 - \mu_2$ is a signed measure. In general, if μ_1 and μ_2 are two measures, $\nu = \mu_1 - \mu_2$ might not be a signed measure. First, it is possible that $\mu_1(A) - \mu_2(A)$ is undefined for some $A \in \mathcal{F}$. Second, even if $\nu(A)$ is defined for all A, we could have $\sum_n \nu(A_n)$ undefined for some family of disjoint sets $\{A_n\}_n$. In the same direction, a signed measure cannot take both $+\infty$ and $-\infty$ values. Indeed, if $\mu(A) = -\infty$ for some $A \in \mathcal{F}$, then $\mu(\Omega) = \mu(A) + \mu(A^c) = -\infty$; if $\mu(B) = +\infty$ for some $B \in \mathcal{F}$ then $\mu(\Omega) = +\infty$, so both cannot hold at the same time.

By the Hahn Decomposition Theorem below, every signed measure ν is the difference between a measure and a finite measure $\mu_1 - \mu_2$ or vice-versa. Moreover, this decomposition is unique if we require that these measures "live" on disjoint sets, that is, if $\mu_1(N) = \mu_2(N^c) = 0$ for some $N \in \mathcal{F}$.

We say that $N \in \mathcal{F}$ is a *negative set* for μ if $\mu(A \cap N) \leqslant 0$ for every $A \in \mathcal{F}$. Likewise, P a *positive set* for μ if $\mu(A \cap P) \geqslant 0$ for every $A \in \mathcal{F}$. Note that the restrictions $\mu_{|_P}$ and $-\mu_{|_N}$ are measures rather than just signed measures.

Lemma B.3. *Let μ be a signed measure on $(\Omega, \mathcal{F})$. For every $D \in \mathcal{F}$, there*

[22] Adapted from [Dem20, Lemma 1.1.31].

exits $A \in \mathcal{F}$ such that A is negative, $A \subseteq D$, and $\mu(A) \leqslant \mu(D)$.

Proof. We can assume that $\mu(D) < 0$, for otherwise we could take $A = \emptyset$. Let $A_0 := D$. We are going to define a decreasing sequence $(A_n)_n$ such that $A_n \downarrow A$ by removing at each step a fraction of any positive measure parts of A_n. Suppose A_n has been defined. Let $t_n := \sup\{\mu(B) : B \in \mathcal{F}, B \subseteq A_n\}$. Note that $t_n \geqslant \mu(\emptyset) = 0$. Let $\delta_n := \min\{\frac{t_n}{2}, 1\}$, so that $0 \leqslant \delta_n < t_n$ unless $t_n = 0$ (we cannot take $\delta_n = \frac{t_n}{2}$ because t_n could be infinite giving $\frac{t_n}{2} = t_n$). Take $B_n \in \mathcal{F}$ such that $B_n \subseteq A_n$ and $\mu(B_n) \geqslant \delta_n$, which is possible by the definition of δ_n and t_n. Define $A_{n+1} := A_n \setminus B_n$, so that $A_n \downarrow A$ for some $A \in \mathcal{F}$.

Since the B_n's are disjoint, we have $\sum_n \mu(B_n) = \mu(\cup_n B_n)$. Now, since $A = D \setminus (\cup_n B_n)$, we have $\mu(D) = \mu(A) + \sum_n \mu(B_n) \geqslant \mu(A) + \sum_n \delta_n \geqslant \mu(A)$.

It remains to show that A is a negative set. Note that $\sum_n \mu(B_n) < +\infty$, for otherwise we would have $\mu(D) = +\infty$. Thus, $0 \leqslant \mu(B_n) \to 0$, whence $\delta_n \to 0$ and therefore $t_n \to 0$ as well. Finally, let $B \in \mathcal{F}$ be such that $B \subseteq A$. For every $n \in \mathbb{N}$, we have $B \subseteq A_n$, so $\mu(B) \leqslant t_n$, and thus $\mu(B) \leqslant 0$. [23] □

Theorem B.4 (Hahn Decomposition Theorem)**.** *Let μ be a signed measure on $(\Omega, \mathcal{F})$. Then there is $N \in \mathcal{F}$ such that N is negative and N^c is positive.*

Proof. We can assume that $\mu(A) \neq -\infty$ for every $A \in \mathcal{F}$, for otherwise we would have $\mu(\Omega) = -\infty$ and we could work with $-\mu$ instead of μ.

Let $\alpha := \inf\{\mu(A) : A \in \mathcal{F}$ is a negative set for $\mu\}$, and note that $\alpha \leqslant \mu(\emptyset) = 0$. Take A_n negative such that $\mu(A_n) \to \alpha$, and let $N := \cup_n A_n$. Now, a countable union of negative sets is negative (why?), so N is a negative set. Note that $\alpha \leqslant \mu(N) = \mu(A_n) + \mu(N \setminus A_n) \leqslant \mu(A_n) \to \alpha$, so $\alpha = \mu(N) \neq -\infty$.

It remains to show that $\mu(B) \geqslant 0$ for any given measurable set $B \subseteq N^c$. By the previous lemma, there exists a negative set $A \in \mathcal{F}$ such that $A \subseteq B$ and $\mu(A) \leqslant \mu(B)$. On the other hand, $\alpha \leqslant \mu(A \cup N) = \mu(A) + \mu(N) = \mu(A) + \alpha$, so $\mu(A) \geqslant 0$, completing the proof. □

With N given by this theorem, the decomposition that we mentioned above is given by $\mu = \mu_{|_{N^c}} - (-\mu_{|_N})$. Restriction measures were defined in Example 2.33.

B.0.4 Radon-Nikodým Theorem

Proof of Theorem 6.29. We split the proof into 3 main steps.

Step 1. We can assume that μ and ν are finite.

Indeed, suppose we had proved the theorem for finite measures. Since μ and ν are σ-finite, there is a measurable partition $(A_n)_{n \in \mathbb{N}}$ of Ω such that $\mu_n := \mu_{|_{A_n}}$

[23] Adapted from Wikipedia.

and $\nu_n := \nu_{|A_n}$ are finite and $\nu_n \ll \mu_n$. Take the corresponding $f_n = \frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}$, and define $f = \sum_n f_n \mathbb{1}_{A_n}$. Checking that $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is straightforward:

$$\nu(A) = \sum_n \nu_n(A) = \sum_n \int_A f_n \, \mathrm{d}\mu_n = \sum_n \int_A f_n \mathbb{1}_{A_n} \, \mathrm{d}\mu = \int_A \sum_n f_n \mathbb{1}_{A_n} \, \mathrm{d}\mu.$$

Step 2. Construct the candidate for the Radon-Nikodým derivative.

We start by defining the family

$$\mathscr{C} := \{g \text{ Borel function} : \textstyle\int_A g \, \mathrm{d}\mu \leqslant \nu(A) \ \forall A \in \mathcal{F}\},$$

which contains $g \equiv 0$. The candidate is a Borel function $f \in \mathscr{C}$ satisfying

$$\int_\Omega f \, \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu \quad \forall g \in \mathscr{C}.$$

It is tempting to just define f as the supremum of all $g \in \mathscr{C}$. However, to make this idea work we need to take a tedious technical detour. As usual, the problem is that many points of measure zero can altogether have positive measure. In particular, at every $\omega \in \Omega$ such that $\mu(\{\omega\}) = 0$, we would have $\sup\{g(\omega) : g \in \mathscr{C}\} = +\infty$ and this would clearly not work.

First note that for $g_1, g_2 \in \mathscr{C}$, $g := \max\{g_1, g_2\} \in \mathscr{C}$. Indeed,

$$\int_A g \, \mathrm{d}\mu = \int_{A \cap \{g_1 \leqslant g_2\}} g_2 \, \mathrm{d}\mu + \int_{A \cap \{g_1 > g_2\}} g_1 \, \mathrm{d}\mu \leqslant$$
$$\leqslant \nu(A \cap \{g_1 \leqslant g_2\}) + \nu(A \cap \{g_1 > g_2\}) = \nu(A).$$

Now define

$$\alpha := \sup\{\textstyle\int_\Omega g \, \mathrm{d}\mu : g \in \mathscr{C}\} \leqslant \nu(\Omega) < \infty$$

and take $(g_n)_n$ in $\mathscr{C}$ such that $\int_\Omega g_n \, \mathrm{d}\mu \to \alpha$. Define $f_n := \max\{g_1, \dots, g_n\} \in \mathscr{C}$. Note that $\alpha \geqslant \int_\Omega f_n \, \mathrm{d}\mu \geqslant \int_\Omega g_n \, \mathrm{d}\mu \to \alpha$, thus $\int_\Omega f_n \, \mathrm{d}\mu \to \alpha$.

Finally, define $f := \lim_n f_n$. By the Monotone Convergence Theorem, $\int_\Omega f \, \mathrm{d}\mu = \lim_n \int_\Omega f_n \, \mathrm{d}\mu = \alpha$. So f is integrable and we can assume it is finite a.e. Again by the Monotone Convergence Theorem, $\int_A f \, \mathrm{d}\mu = \lim_n \int_A f_n \, \mathrm{d}\mu \leqslant \nu(A)$, for all $A \in \mathcal{F}$ because $f_n \in \mathscr{C}$. Therefore $f \in \mathscr{C}$ and its integral over Ω is maximal among all $g \in \mathscr{C}$, which are the properties we claimed above.

Step 3. Check that f is the Radon-Nikodým derivative.

Define the signed measure

$$\nu_0(A) := \nu(A) - \int_A f \, \mathrm{d}\mu.$$

By definition of $f \in \mathscr{C}$, the function ν_0 is actually a measure. Hence, to prove the theorem it suffices to show that $\nu_0(\Omega) = 0$.

Let $\varepsilon > 0$. We are going to show that the signed measure

$$\nu^\varepsilon := \varepsilon\mu - \nu_0$$

is actually a measure. This will give $\nu_0(\Omega) = \varepsilon\mu(\Omega) - \nu^\varepsilon(\Omega) \leqslant \varepsilon\mu(\Omega)$ and, since ε is arbitrary, $\nu_0(\Omega) = 0$, which is what we are trying to establish.

By the Hahn Decomposition Theorem, we can partition $\Omega = N \cup P$ and write ν^ε as the difference between two measures: $\nu^\varepsilon = \nu^\varepsilon_{|P} - (-\nu^\varepsilon_{|N})$. So it is enough to prove that $\nu^\varepsilon(N) = 0$.

Define

$$g = f + \varepsilon \mathbb{1}_N.$$

We now check that $g \in \mathscr{C}$. Indeed, expanding and bounding,

$$\begin{aligned}\int_A g \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu + \varepsilon\mu(A \cap N) &= \int_A f \,\mathrm{d}\mu + \nu^\varepsilon(A \cap N) + \nu_0(A \cap N) \\ &\leqslant \int_A f \,\mathrm{d}\mu + \nu_0(A \cap N) \leqslant \int_A f \,\mathrm{d}\mu + \nu_0(A) = \nu(A).\end{aligned}$$

On the other hand, by maximality of f on $\mathscr{C}$, we have

$$\int_\Omega f \,\mathrm{d}\mu \geqslant \int_\Omega g \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu + \varepsilon\mu(N),$$

hence $\mu(N) = 0$. Since $\nu \ll \mu$, we have $\nu(N) = 0$. In particular, $0 \leqslant \nu_0(N) \leqslant \nu(N) = 0$, so $\nu^\varepsilon(N) = \varepsilon\mu(N) - \nu_0(N) = 0$ concluding the proof. [24] □

B.0.5 Product measures

We finally give the postponed proofs from Chapter 7.

Proof of Example 7.2. First, we show $\mathcal{B}(\mathbb{R}^2) \subseteq \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Let

$$\mathcal{E} := \{(a, b] \cap \mathbb{R} : -\infty \leqslant a \leqslant b \leqslant +\infty\},$$

$\mathcal{E}^2 := \mathcal{E} \times \mathcal{E}$, and recall that $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{E}^2)$. Since $\mathcal{E}^2 \subseteq \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$, we have $\mathcal{B}(\mathbb{R}^2) \subseteq \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

We now show the other direction. Note that

$$\{A \times B : A \in \sigma(\mathcal{E}), B = \mathbb{R}\} = \sigma(\{A \times B : A \in \mathcal{E}, B = \mathbb{R}\})$$

as the set B is being pinned to $B = \mathbb{R}$ plays no role in any of the theoretic operations involved. Likewise,

$$\{A \times B : A = \mathbb{R}, B \in \sigma(\mathcal{E})\} = \sigma(\{A \times B : A = \mathbb{R}, B \in \mathcal{E}\}).$$

[24] Adapted from [Coh13].

In particular, for $A, B \in \mathcal{B}(\mathbb{R}) = \sigma(\mathcal{E})$, we have

$$A \times B = (A \times \mathbb{R}) \cap (\mathbb{R} \times B) \in \sigma(\mathcal{E} \times \mathcal{E}).$$

Thus, $\sigma(\mathcal{E}) \times \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E} \times \mathcal{E})$ and hence

$$\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) = \sigma(\mathcal{E}) \times \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E} \times \mathcal{E}) = \sigma(\mathcal{E}^2) = \mathcal{B}(\mathbb{R}^2),$$

Therefore, $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})) \subseteq \mathcal{B}(\mathbb{R}^2)$. □

Proof of Lemma 7.5. For the first statement, we show that each section E_x is $\mathcal{F}_2$-measurable. A symmetric reasoning can be used to show that each section E^y is $\mathcal{F}_1$-measurable. Fix $x \in \Omega_1$, and define the collection $\mathcal{G}$ by

$$\mathcal{G} = \{E \subseteq \Omega_1 \times \Omega_2 : E_x \in \mathcal{F}_2\}.$$

By construction, $\mathcal{G}$ contains $\mathcal{F}_1 \times \mathcal{F}_2$. It remains to show that $\mathcal{G}$ is a σ-algebra on $\Omega_1 \times \Omega_2$, so that it contains $\sigma(\mathcal{F}_1 \times \mathcal{F}_2) = \mathcal{F}_1 \otimes \mathcal{F}_2$. Indeed, we have

(i) $\Omega_1 \times \Omega_2 \in \mathcal{G}$, since $(\Omega_1 \times \Omega_2)_x = \Omega_2 \in \mathcal{F}_2$.

(ii) Let $E \in \mathcal{G}$. Since $E_x \in \mathcal{F}_2$, then $(E^c)_x = (E_x)^c \in \mathcal{F}_2$, so that $E^c \in \mathcal{G}$.

(iii) Let $E_1, E_2, \dots \in \mathcal{G}$. Then, $(E_j)_x \in \mathcal{F}_2$ for each $j \in \mathbb{N}$. This in turn gives

$$\Big(\bigcup_{i=1}^{\infty} E_i\Big)_x = \bigcup_{i=1}^{\infty} (E_i)_x \in \mathcal{F}_2,$$

so that $\bigcup_{i=1}^{\infty} E_i \in \mathcal{G}$.

where we use the identities $(E^c)_x = (E_x)^c$ and $(\bigcup_{i=1}^{\infty} E_i)_x = \bigcup_{i=1}^{\infty} (E_i)_x$.

We now prove the second statement. Again, we only prove that each f_x is $\mathcal{F}_2$-measurable. Note that $(f_x)^{-1}(D) = (f^{-1}(D))_x$ for any $D \in \mathcal{B}(\mathbb{R})$. Combined with the first statement, this implies that $(f_x)^{-1}(D) \in \mathcal{F}_2$ for every $D \in \mathcal{B}$, concluding the proof.[25] □

Proof of Lemma 7.6. We show that $x \mapsto \nu(E_x)$ is $\mathcal{F}_1$-measurable, the other statement is analogous.

First, suppose that the measure ν is finite. Define the collection $\mathcal{G}$ by

$$\mathcal{G} = \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2 : x \mapsto \nu(E_x) \text{ is } \mathcal{F}_1\text{-measurable}\}.$$

We will show that $\mathcal{G} = \mathcal{F}_1 \otimes \mathcal{F}_2$.

Note from Lemma 7.5 (i) that, given $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, each section E_x lies in $\mathcal{F}_2$, so that the function $x \mapsto \nu(E_x)$ is well-defined. If $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, we have

$$\nu((A \times B)_x) = \nu(B)\mathbb{1}_A(x), \tag{B.5}$$

which implies that $\mathcal{G}$ contains $\mathcal{F}_1 \times \mathcal{F}_2$. Note that $\mathcal{F}_1 \times \mathcal{F}_2$ defines a π-system on $\Omega_1 \times \Omega_2$. We now show that $\mathcal{G}$ is a λ-system. Indeed, we have

[25] Based on [Coh13, 5.1.2].

(i) $\Omega_1 \times \Omega_2 \in \mathcal{G}$.

(ii) For $E \in \mathcal{G}$, we can write

$$\nu((E^c)_x) = \nu((E_x)^c) = \nu(\Omega_2) - \nu(E_x),$$

and so $E^c \in \mathcal{G}$.

(iii) Let $E_1, E_2, \cdots \in \mathcal{G}$, where $E_i \cap E_j = \emptyset$ for $i \neq j$. Then, by countable additivity of ν, we get

$$\nu\Big(\Big(\bigcup_{i=1}^{\infty} E_i\Big)_x\Big) = \nu\Big(\bigcup_{i=1}^{\infty} (E_i)_x\Big) = \sum_{i=1}^{\infty} \nu((E_i)_x),$$

so that $\bigcup_{i=1}^{\infty} E_i \in \mathcal{G}$.

Therefore, the π-λ Theorem gives $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{F}_1 \times \mathcal{F}_2) \subseteq \mathcal{G} \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$, concluding the proof.

Now, suppose that ν is σ-finite, and let $\{D_i\}_{i=1}^{\infty}$ be a sequence of disjoint events in $\mathcal{F}_2$ with $\nu(D_j) < \infty$, $\forall j \in N$, and $\bigcup_{i=1}^{\infty} D_i = \Omega_2$. Define the finite measures $\nu_1, \nu_2, \dots$ by

$$\nu_k(B) = \nu(B \cap D_k),\ k \in \mathbb{N}.$$

Given $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, each $x \mapsto \nu_k(E_x)$ is $\mathcal{F}_1$-measurable, hence

$$\nu(E_x) = \sum_{i=1}^{\infty} \nu(E_x \cap D_i) = \sum_{i=1}^{\infty} \nu_i(E_x),$$

is also $\mathcal{F}_1$-measurable. Arguing similarly for the function $y \mapsto \mu(E^y)$ finishes the proof.[26] $\square$

Proof of Theorem 7.7. We start with uniqueness of the measure $\mu \otimes \nu$. Since $\mathcal{F}_1 \times \mathcal{F}_2$ is a π-system that generates $\mathcal{F}_1 \otimes \mathcal{F}_2$, we can apply the π-λ Theorem to get uniqueness. Moreover, to get existence as well as the second formula, it is enough to show that each of those two integrals yield a measure which satisfies the first one.

We know from Lemma 7.6 that, given $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are $\mathcal{F}_1$-measurable and $\mathcal{F}_2$-measurable, respectively.

So we can define the set functions $(\mu \otimes \nu)_1 : \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0, \infty]$ and $(\mu \otimes \nu)_2 : \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0, \infty]$ by

$$(\mu \otimes \nu)_1(E) = \int_{\Omega_1} \nu(E_x)\mu(\mathrm{d}x) \quad \text{and} \quad (\mu \otimes \nu)_2(E) = \int_{\Omega_2} \mu(E^y)\nu(\mathrm{d}y).$$

Let us check that $(\mu \otimes \nu)_1$ is a measure:

(i) $(\mu \otimes \nu)_1(\emptyset) = \int_{\Omega_1} \nu(\emptyset)\mu(\mathrm{d}x) = 0.$

[26] Expanded from [Coh13, 5.1.3].

(ii) Let $E_1, E_2, \dots \in \mathcal{F}_1 \otimes \mathcal{F}_2$, where $E_i \cap E_j = \emptyset$ for $i \neq j$. By countable additivity of ν, we have

$$(\mu \otimes \nu)_1(\bigcup_{i=1}^{\infty} E_i) = \int_{\Omega_1} \nu((\bigcup_{i=1}^{\infty} E_i)_x)\mu(\mathrm{d}x) = \int_{\Omega_1} \nu(\bigcup_{i=1}^{\infty}(E_i)_x)\mu(\mathrm{d}x) =$$
$$= \sum_{i=1}^{\infty} \int_{\Omega_1} \nu((E_i)_x)\mu(\mathrm{d}x) = \sum_{i=1}^{\infty}(\mu \otimes \nu)_1(E_i).$$

By analogous argument, $(\mu \otimes \nu)_2$ is also a measure.

Moreover, for every $A \times B \in \mathcal{F}_1 \times \mathcal{F}_2$, by (B.5) we have

$$(\mu \otimes \nu)_1(A \times B) = \mu(A)\nu(B),$$

and analogously we can prove $(\mu \otimes \nu)_2(A \times B) = \mu(A)\nu(B)$. As noted in the first paragraph, this concludes the proof.[27] $\square$

[27] Expanded from [Coh13, 5.1.4].

References

[Bar95] R. G. Bartle. *The elements of integration and Lebesgue measure*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1995. doi.

[Coh13] D. L. Cohn. *Measure theory*. Birkhäuser Advanced Texts: Basel Textbooks. Birkhäuser/Springer, New York, second edn., 2013. doi.

[Dem20] A. Dembo. *Probability theory: STAT310/MATH230*, 2020. Lecture Notes, Stanford. Available online as of June 2020.

[Kle14] A. Klenke. *Probability theory*. Universitext. Springer, London, second edn., 2014. doi.

[Kub15] C. S. Kubrusly. *Essentials of measure theory*. Springer, Cham, 2015. doi.

[Shi16] A. N. Shiryaev. *Probability. 1*, vol. ‘ of *Graduate Texts in Mathematics*. Springer, New York, third edn., 2016. doi.

[Tao11] T. Tao. *An introduction to measure theory*, vol. 126 of *Graduate Studies in Mathematics*. American Mathematical Society, 2011. doi. Available online as of June 2020.