An introduction to probability theory II

Stefan Geiss
Department of Mathematics and Statistics
University of Jyväskylä

June 1, 2009

Contents

Introduction

This script is a continuation of [5] so that we shall assume that the reader is familiar with the basics from this previous course. Nevertheless, some basic facts will be recalled in Chapter 0.

Let us motivate this course by some examples.

Example 1. Assume that we perform some experiment several times under identical conditions and get, each time, a measurement denoted by $f_1, f_2, f_3, ...$ To get the *true* quantity (whatever this means) we naturally consider

$$S_n = \frac{1}{n}(f_1 + \cdots + f_n)$$

for large n and hope that S_n converges to this true value as $n \to \infty$. To make this precise we have to clarify at least three things:

- What does it mean that we recall an experiment under identical conditions? This leads us to the mathematical notion of *(stochastic) independence.*
- In what sense does the convergence take place? This will take us to the *almost sure convergence.*
- And finally we have to identify the limit.

Altogether we end up with famous and fundamental *Strong Law of Large Numbers.*

Example 2. We consider a random walk

$$S_n := \varepsilon_1 + \cdots + \varepsilon_n$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3, ...$ are independent and

$$\text{probability}(\varepsilon_k = 1) = \text{ probability}(\varepsilon_k = -1) = \frac{1}{2}.$$

Now we think about a particle which moves up and down in an infinitesimal way with probability $1/2$ and let us consider the time interval $[0, 1]$. To get an approximation of the movement we divide the time interval into n equal sub-intervals and let the particle move down or up with probability $1/2$ at each time point k/n with $k = 0, ..., n-1$ by the same jump-size. What is the right jump-size we have to choose in our model that the model converges in some sense as $n \to \infty$? In other words, what is the *right* rescaling factor $c_n > 0$ in the state variable such that

$$\frac{\varepsilon_1 + \cdots + \varepsilon_n}{c_n}$$

converges to a non-degenerate random variable as $n \to \infty$? This will lead us to the *weak convergence* and to the *Central Limit Theorem.*

Example 3. We assume a source which sends out particles to one side with a uniformly distributed (random) angle which ranges from zero to 180 degree. There is a wall in distance of 1 meter. The function $f(\theta)$ gives the position of the particle which hits the wall where $\theta = \pi/2$ gives position 0. The function f is random because it depends on the random angle θ. Consider a second wall parallel to the first one, but with a distance of 2 meters. Now we think in two different ways: firstly, the particle sent out hits the first wall and will be resent out to the second wall. Secondly, we do the experiment in one step and wait until the particle hits the second wall. Is it possible that both experiments give the same distribution? The answer is *yes.* Knowing this we can analyze the distribution in an abstract manner: Namely, we obtain that the *distributions* of

$$f_1 + f_2 \quad \text{and} \quad 2f$$

are the same, where f_1 and f_2 are *independent copies* of f. The property of this distribution is to be 1-stable, the distribution is called CAUCHY [1] distribution. Stable distributions are used (for example) in stochastic modeling and in probabilistic methods in functional analysis. We shall prove the existence of stable distributions by the help of *characteristic functions.*

[1] Augustin Louis Cauchy, 21 Aug 1789 (Paris) - 23 May 1857 (Sceaux), real and complex analysis.

Chapter 0

Preliminaries

In this section we recall some basic concepts needed in the sequel.

1. Probability spaces. A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is a non-empty set of elementary events,
- $\mathcal{F} \subseteq 2^{\Omega}$ is a system of observable events,
- $\mathbb{P}$ is a function that gives to each event $A \in \mathcal{F}$ a probability $\mathbb{P}(A) \in [0,1]$ to occur.

Example 1. Standard examples for the set of elementary events Ω are the following:

(a) $\Omega = \{1, ..., , 6\}$ are the possible outcomes if one rolls a die.

(b) $\Omega = (0, \infty)$ are possible share prices (for example of NOKIA).

Usually we cannot decide whether a particular $\omega \in \Omega$ in our system occurs. But for certain subsets $A \subseteq \Omega$ we can say whether $\omega \in A$ or $\omega \notin A$. A system of subsets $A \subseteq \Omega$ with this property is called a system of observable events. This system is usually assumed to be a σ-algebra. Let us formulate this in the

Definition 2. Let Ω be a non-empty set.

- A system $\mathcal{F}$ of subsets $A \subseteq \Omega$ is called *algebra,* provided that
  (i) $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$,
  (ii) if $A, B \in \mathcal{F}$, then $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\} \in \mathcal{F}$ and
  (iii) if $A \in \mathcal{F}$, then $A^{\mathrm{C}} = \{\omega \in \Omega : \omega \notin A\} \in \mathcal{F}$.

- $\mathcal{F}$ is called *σ-algebra,* provided that (ii) is replaced by

  (ii') if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

- The pair $(\Omega, \mathcal{F})$ is called *measurable space,* if $\mathcal{F}$ is a σ-algebra.

- The elements of a σ-algebra $\mathcal{F}$ are called *events*. We say that an event $A \in \mathcal{F}$ occurs if $\omega \in A$ and that A does not occur if $\omega \notin A$.

We give some basic examples of σ-algebras.

Example 3. (a) The system of all subsets of Ω, denoted by 2^{Ω}, is the largest σ-algebra on Ω.

(b) The system $\{\emptyset, \Omega\}$ is the smallest σ-algebra on Ω. Any σ-algebra $\mathcal{F}$ on Ω satisfies $\{\emptyset, \Omega\} \subseteq \mathcal{F} \subseteq 2^{\Omega}$.

(c) Assume that you have two dice, but only the sum of the dice is known. The corresponding measurable space is given by

$$\begin{aligned} \Omega &:= \{(k,l) : k,l = 1, ..., 6\}, \\ \mathcal{F} &:= \{A \subseteq \Omega : A \text{ is the empty set or a union of } A_2, ..., A_{12}\}, \end{aligned}$$

where $A_m := \{(k,l) : k + l = m\}$.

In important cases there is not a complete description of the σ-algebra. Surprisingly, there are abstract approaches which can be used practically. The basic idea is

Lemma 4. *Let $\Omega \neq \emptyset$ and $\mathcal{G}$ be a non-empty system of subsets of Ω and*

$$\sigma(\mathcal{G}) := \bigcap_{\mathcal{F}\ \sigma\text{-algebra and } \mathcal{G} \subseteq \mathcal{F}} \mathcal{F}.$$

Then one has the following:

(i) *The system $\sigma(\mathcal{G})$ is a σ-algebra containing all sets from $\mathcal{G}$.*

(ii) *If $\mathcal{F}$ is any σ-algebra on Ω containing all $G \in \mathcal{G}$, then $\sigma(\mathcal{G}) \subseteq \mathcal{F}$.*

The proof can be found in [5]. Using this lemma one can introduce Borel [1] σ-algebras on metric spaces or even on more general spaces. To do this on $\mathbb{R}^d$, with the final goal to define the Lebesgue measure on it, we recall the notion of an open set.

[1] Félix Edouard Justin Émile Borel, 07/01/1871-03/02/1956, French mathematician.

Definition 5. (i) For $x, y \in \mathbb{R}^d$ we let $|x-y| = \left(\sum_{i=1}^d |x_i - y_i|^2\right)^{\frac{1}{2}}$

(ii) A subset $G \in \mathbb{R}^d$ is *open,* if and only if for all $x \in G$ there exists $\varepsilon > 0$ such that $y \in G$ for all $y \in \mathbb{R}^d$ such that $|x-y| < \varepsilon$.

Now we are ready to introduce the Borel-σ-algebra:

Definition 6. The *Borel-σ-algebra* $\mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d$ is the smallest σ-algebra that contains all open sets of $\mathbb{R}^d$.

Finally we recall the notion of a measure:

Definition 7. Let $(\Omega, \mathcal{F})$ be a measurable space.

(i) A map $\mu : \mathcal{F} \to [0,\infty]$ is called *measure* on $\mathcal{F}$ if

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

for all pair-wise disjoint $A_1, A_2, \ldots \in \mathcal{F}$.

(ii) A measure μ on $\mathcal{F}$ is called *σ-finite,* provided that there exists $\Omega_1, \Omega_2, \ldots \in \mathcal{F}$, $\Omega_n \cap \Omega_m = \emptyset$ for $n \neq m$, $\bigcup_{n=1}^{\infty} \Omega_n = \Omega$ such that $\mu(\Omega_n) < \infty$ for all $n \geq 1$.

(iii) A measure μ on $\mathcal{F}$ is called *probability measure,* if $\mu(\Omega) = 1$.

If μ is a measure, then $(\Omega, \mathcal{F}, \mu)$ is called *measure space.* If μ is a probability measure, then $(\Omega, \mathcal{F}, \mu)$ is a *probability space.*

Example 8. Let Ω be an arbitrary non-empty set and $\mathcal{F} := 2^{\Omega}$.

(a) A trivial (but sometimes important) example is the *counting measure*

$$\mu(B) := \mathrm{card}(B).$$

This measure is σ-finite if and only if Ω is countable.

(b) Another measure we obtain as follows: we fix $\omega_0 \in \Omega$ and let $\mu(B) = 1$ if $\omega_0 \in B$, otherwise $\mu(B) = 0$. This measure is called DIRAC-*measure* [2] at $\omega_0 \in \Omega$.

(c) Let $B_0 \subseteq \Omega$ with $\mathrm{card}(B_0) > 1$, and let $\mu(B) = 1$ if $B_0 \cap B \neq \emptyset$, otherwise $\mu(B) = 0$. One notes, that μ is not a measure.

[2] Paul Adrien Maurice Dirac, 08/08/1902 (Bristol, England) - 20/10/1984 (Tallahassee, Florida, USA), Nobel price in Physics 1933.

2. Construction of measures. Now we consider the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and look for a measure λ such that

$$\lambda((a,b]) = \lambda((a,b)) = \lambda([a,b]) = b - a$$

for all $-\infty < a < b < \infty$. For its construction we need

Proposition 9 (Carathéodory [3]). *Let $\Omega \neq \emptyset$, $\mathcal{G}$ be an algebra such that $\mathcal{F} = \sigma(\mathcal{G})$. Assume a map $\mu_0 : \mathcal{G} \to [0,\infty]$ such that*

(i) $\mu_0(\Omega_n) < \infty$, $n = 1, 2, ...,$ *for some partition* $\Omega = \bigcup_{n=1}^\infty \Omega_n$ *with* $\Omega_n \in \mathcal{G}$,

(ii) $\mu_0\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu_0(A_n)$ *for pair-wise disjoint* $A_n \in \mathcal{G}$ *such that*

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

Then there exists a unique measure $\mu : \mathcal{F} \to [0,\infty]$ such that

$$\mu(A) = \mu_0(A) \quad \text{for all} \quad A \in \mathcal{G}.$$

It is clear that the measure μ is automatically σ-finite. As an application one can prove that the LEBESGUE [4] measure exists on $\mathbb{R}^d$.

Proposition 10. *There is a unique σ-finite measure λ_d on $\mathcal{B}(\mathbb{R}^d)$ such that*

$$\lambda_d((a_1,b_1] \times \ldots \times (a_d,b_d]) = \prod_{i=1}^d (b_i - a_i).$$

This measure is called Lebesgue measure.

3. Product spaces. Product spaces are obtained by the following

Proposition 11 (Product measure). *Assume that $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2, ..., d$ are σ-finite measure spaces and let*

$$\mathcal{F} = \otimes_{i=1}^d \mathcal{F}_i := \sigma(B_1 \times \cdots \times B_d : B_i \in \mathcal{F}_i).$$

Then there exists a unique measure $\mu = \otimes_{i=1}^d \mu_i$ on $\mathcal{F}$ (which is automatically σ-finite) such that

$$\mu(B_1 \times \cdots \times B_d) = \mu_1(B_1) \cdots \mu_n(B_d).$$

Proposition 12. *We have that*

$$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda_d) = \otimes_{i=1}^d (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda_1).$$

---

[3] Constantin Carathéodory, 13/09/1873 (Berlin, Germany) - 02/02/1950 (Munich, Germany).

[4] Henri Léon Lebesgue, 28/06/1875-26/07/1941, French mathematician generalized the Riemann integral by the Lebesgue integral; continuation of work of Emile Borel and Camille Jordan.

4. Random variables

Definition 13. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. A map $f : \Omega_1 \to \Omega_2$ is called *measurable,* provided that for all $B \in \mathcal{F}_2$ one has

$$f^{-1}(B) = \{\omega_1 \in \Omega_1 : f(\omega_1) \in B\} \in \mathcal{F}_1.$$

The reason for the definition above is that a measurable map can transport a measure from one measurable space to another.

Proposition 14 (Image measures)**.** *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, $f : \Omega_1 \to \Omega_2$ be a measurable map, and μ_1 be a measure on $\mathcal{F}_1$. Define*

$$\mu_2(B_2) := \mu_1(f^{-1}(B_2)) \text{ for } B_2 \in \mathcal{F}_2.$$

Then one has the following:

(i) $(\Omega_2, \mathcal{F}_2, \mu_2)$ *is a measure space with* $\mu_1(\Omega_1) = \mu_2(\Omega_2)$.

(ii) *There are examples that μ_1 is σ-finite, but μ_2 is not.*

We shall use the following notation:

(i) One writes $\mu_2 = (\mu_1)_f$ and says that μ_2 is the *image(measure)* of μ_1 with respect to f.

(ii) If $(\Omega_2, \mathcal{F}_2) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then one also says that μ_2 is the *law* of f (w.r.t. μ_1).

The property that a map is measurable can be checked sometimes easily by generating systems.

Proposition 15. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, $\mathcal{G}_2$ be a system of subsets of Ω_2 such that $\sigma(\mathcal{G}_2) = \mathcal{F}_2$ and $f : \Omega_1 \to \Omega_2$ be a map such that $f^{-1}(B) \in \mathcal{F}_1$ for all $B \in \mathcal{G}_2$. Then f is measurable.*

Proof. Define $\mathcal{B}_2 := \{B \subseteq \Omega_2 : f^{-1}(B) \in \mathcal{F}_1\}$. Then $\mathcal{B}_2$ is a σ-algebra and $\mathcal{G}_2 \subseteq \mathcal{B}_2$ by assumption. Hence $\mathcal{F}_2 \subseteq \mathcal{B}_2$. □

What are the typical measurable maps $h : \Omega \to \mathbb{R}$? Let $(\Omega, \mathcal{F})$ be a measurable space. Then $h : \Omega \to \mathbb{R}$ is called (measurable) *step function*, if there are $B_1, \ldots, B_n \in \mathcal{F}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$h(\omega) = \sum_{k=1}^{n} \chi_{B_k}(\omega)\alpha_k,$$

where

$$\chi_{B_k}(\omega) := \begin{cases} 1, & \text{if } \omega \in \mathrm{B_k} \\ 0, & \text{if } \omega \notin \mathrm{B_k}. \end{cases}$$

Proposition 16. *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \mathbb{R}$. Then the following assertions are equivalent.*

(i) *f is measurable as a map from $(\Omega, \mathcal{F})$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

(ii) *$f^{-1}((a,b)) \in \mathcal{F}$ for all $-\infty < a < b < \infty$.*

(iii) *There exists (measurable) step functions $h_n : \Omega \to \mathbb{R}$ such that*

$$\lim_{n\to\infty} h_n(\omega) = f(\omega) \ for\ all\ \omega \in \Omega.$$

Proof. (i)$\Rightarrow$(ii) is clear since $(a,b) \in \mathcal{B}(\mathbb{R})$.
(ii)$\Rightarrow$(iii): For $-\infty < a < b < \infty$ one has

$$f^{-1}([a,b)) = \bigcap_{n=1}^{\infty} f^{-1}((a - \frac{1}{n}, b)) \in \mathcal{F},$$

since $f^{-1}((a - \frac{1}{n}, b)) \in \mathcal{F}$ for all $n = 1, 2, \ldots$. Letting for $n \in \{1, 2, \ldots\}$

$$h_n(\omega) = \sum_{k=-2^{n+1}}^{2^{n+1}} \chi_{f^{-1}([\frac{k-1}{2^n}, \frac{k}{2^n}))}(\omega) \frac{k-1}{2^n}$$

we get a sequence $(h_n)_{n=1}^{\infty}$ of step functions and $\lim_{n\to\infty} h_n(\omega) = f(\omega)$ for all $\omega \in \Omega$.
(iii)$\Rightarrow$(ii): We write

$$\begin{aligned}
f^{-1}((a,b)) &= \{\omega \in \Omega : f(\omega) \in (a,b)\} \\
&= \{\omega \in \Omega : \lim_n h_n(\omega) \in (a,b)\} \\
&= \{\omega \in \Omega : \exists N \geq 1 \forall n \geq N h_n(\omega) \in (a,b)\} \\
&= \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega \in \Omega : h_n(\omega) \in (a,b)\} \\
&= \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} h_n^{-1}((a,b)) \in \mathcal{F}.
\end{aligned}$$

(ii)$\Rightarrow$(i) follows from Proposition 15. $\square$

By the above proof we also get the following

Proposition 17. *If $h_n : \Omega \to \mathbb{R}$ are measurable, where $(\Omega, \mathcal{F})$ is a measurable space and if*

$$f(\omega) = \lim_{n\to\infty} h_n(\omega) \ for\ all\ \omega \in \Omega,$$

then f is measurable.

Definition 18. Let $(\Omega, \mathcal{F})$ be a measurable space. A map $f : \Omega \to \mathbb{R}$ which satisfies one of the equivalent properties of Proposition 16 is called *random variable* or *Borel measurable.*

5. Integration We recall briefly the construction of the expected value.

Definition 19. Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space.

(i) If $h = \sum_{i=1}^n \alpha_i \chi_{A_i}$ with $A_i \in \mathcal{F}$ and $\alpha_i \geq 0$, then

$$\int_\Omega h d\mu = \mathbb{E}h := \sum_{i=1}^n \alpha_i \mu(A_i) \in [0, \infty].$$

(ii) If $f : \Omega \to \mathbb{R}$ is measurable and non-negative, then

$$\int_\Omega f d\mu := \sup \left\{ \int_\Omega h d\mu : 0 \leq h \leq f; h \text{ measurable step-function} \right\}.$$

(iii) If $f : \Omega \to \mathbb{R}$ is measurable,

$$f^+ := \max\{f, 0\} \quad \text{and} \quad f^- := \max\{-f, 0\},$$

then $f = f^+ - f^-$ and we define

$$\int_\Omega f d\mu := \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu$$

provided that $\int_\Omega f^+ d\mu < \infty$ or $\int_\Omega f^- d\mu < \infty$.

(iv) A measurable function $f : \Omega \to \mathbb{R}$ is *integrable* if

$$\int_\Omega |f| d\mu < \infty.$$

We will often use the theorem about monotone convergence.

Proposition 20 (Monotone Convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and $f_n, f : \Omega \to \mathbb{R}$ be measurable such that*

$$0 \leq f_n(\omega) \uparrow f(\omega)$$

for all $\omega \in \Omega$. Then one has

$$\lim_n \int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

Let us consider some examples.

Example 21. (a) Lorentz sequence spaces: Let $\Omega := \{1, 2, ...\}$, $\mathcal{F} := 2^\Omega$, and let μ be the counting measure, i.e. $\mu(A) := \mathrm{card}(A)$. Then, for $f \geq 0$,

$$\int_\Omega f d\mu = \sum_{k=1}^\infty f(k).$$

To check this rigorously, we use the step-functions $h_n(k) := f(k)$ if $0 \leq k \leq n$ and $h_n(k) := 0$ for $k > n$. Then

$$\int_\Omega f d\mu = \lim_n \int_\Omega h_n d\mu = \lim_n \sum_{k=1}^n f(k)\mu(\{k\}) = \lim_n \sum_{k=1}^n f(k).$$

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be continuous except in finitely many points, where in these points the left and right-hand side limits exist and are finite. Assume that

$$\lim_{n\to\infty} R - \int_{-n}^{n} |f(x)| dx < \infty$$

where $R - \int$ is the Riemann integral. Then

$$\int_{\mathbb{R}} f(x) d\lambda(x) = \lim_{n} R - \int_{-n}^{n} f(x) dx.$$

(c) Assume an increasing and right-continuous function $F : \mathbb{R} \to \mathbb{R}$. Consider the unique σ-finite measure μ on $\mathcal{B}(\mathbb{R})$ with

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

The measure μ is called Lebesgue-Stieltjes [5] measure. The corresponding integral

$$\int f(t) dF(t) := \int_{\mathbb{R}} f(t) d\mu(t)$$

is an extension of the Riemann-Stieltjes integral.

6. Fubini's theorem For the following it is convenient to allow that the random variables may take infinite values. Moreover, we assume that $(\Omega, \mathcal{F}, \mu)$ and $(\Omega_i, \mathcal{F}_i, \mu_i)$ are σ-finite measure spaces.

Definition 22 (Extended random variable)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ is called *extended random variable* if and only if

$$f^{-1}(B) := \{\omega : f(\omega) \in B\} \in \mathcal{F} \quad \text{for all} \quad B \in \mathcal{B}(\mathbb{R}) \text{ or } B = \{-\infty\}.$$

If we have a non-negative extended random variable, we let (for example)

$$\int_{\Omega} f d\mu = \lim_{N\to\infty} \int_{\Omega} [f \wedge N] d\mu.$$

Proposition 23 (Fubini's Theorem)**.** [6] *Let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a non-negative $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable function such that*

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mu_1 \times \mu_2)(\omega_1, \omega_2) < \infty. \tag{1}$$

Then one has the following:

[5] Thomas Jan Stieltjes 29/12/1856 (Zwolle, Overijssel, The Netherlands) - 31/12/1894 (Toulouse, France); analysis, number theory.

[6] Guido Fubini, 19/01/1879 (Venice, Italy) - 06/06/1943 (New York, USA).

(i) *The functions* $\omega_1 \to f(\omega_1, \omega_2^0)$ *and* $\omega_2 \to f(\omega_1^0, \omega_2)$ *are* $\mathcal{F}_1$*-measurable and* $\mathcal{F}_2$*-measurable, respectively, for all* $\omega_i^0 \in \Omega_i$.

(ii) *The functions*

$$\omega_1 \to \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \quad \textit{and} \quad \omega_2 \to \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1)$$

are extended $\mathcal{F}_1$*-measurable and* $\mathcal{F}_2$*-measurable, respectively, random variables.*

(iii) *One has that*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(\mu_1 \times \mu_2) &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right] d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right] d\mu_2(\omega_2). \end{aligned}$$

Remark 24. (i) It should be noted, that item (iii) together with Formula (1) automatically implies that

$$\mu_2 \left(\left\{ \omega_2 : \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) = \infty \right\} \right) = 0$$

and

$$\mu_1 \left(\left\{ \omega_1 : \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) = \infty \right\} \right) = 0.$$

(ii) Fubini's theorem for general $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ one gets by decomposing $f = f^+ - f^-$.

Let us finish with an

Example 25. Assume that $\Omega_1 = \Omega_2 = [0,1]$, $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{B}([0,1])$ where $\mathcal{B}([0,1])$ is generated by the open sets in $[0,1]$. Assume that both measurable spaces are equipped with the Lebesgue measure λ. Let $\varphi : [0,1] \to [0,1]$ be a (say) continuous function and let

$$B := \{(s,t) \in [0,1]^2 : t < f(s)\}.$$

(i) One has $B \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

(ii) Applying Fubini's theorem we get

$$\begin{aligned} \lambda \otimes \lambda(B) &= \int_0^1 \int_0^1 \chi_{\{t \le f(s)\}} d\lambda(t) d\lambda(s) \\ &= \int_0^1 f(s) d\lambda(s). \end{aligned}$$

(iii) Applying Fubini's theorem in the other way we get

$$\begin{aligned}\lambda \otimes \lambda(B) &= \int_0^1 \int_0^1 \chi_{\{t<f(s)\}} d\lambda(s) d\lambda(t) \\ &= \int_0^1 \lambda(\{s \in [0,1] : f(s) > t\}) d\lambda(t).\end{aligned}$$

(iv) Combining both estimates we have

$$\int_0^1 f(s) d\lambda(s) = \int_0^1 \lambda(\{s \in [0,1] : f(s) > t\}) d\lambda(t).$$

(v) Exercise: Prove in the same way that, for $0 < p < \infty$,

$$\int_0^1 f(s)^p d\lambda(s) = p \int_0^1 \lambda(\{s \in [0,1] : f(s) > t\}) t^{p-1} d\lambda(t).$$

Chapter 1

Modes of convergence

1.1 Almost sure convergence

Definition 1.1.1. Let f_n, $f : \Omega \to \mathbb{R}$ be random variables where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. We say that f_n *converges almost surely to* f if

$$\mathbb{P}(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \underset{n}{\longrightarrow} 0\}) = 1.$$

We write $f_n \underset{a.s.}{\longrightarrow} f$.

Remark 1.1.2. (i) To formulate the above definition we need $\{\omega : |f_n(\omega) - f(\omega)| \underset{n}{\longrightarrow} 0\}) \in \mathcal{F}$. This follows from

$$\begin{aligned}
& \{\omega \in \Omega : |f_n(\omega) - f(\omega)| \underset{n}{\longrightarrow} 0\} \\
= & \left\{\omega : \forall\, m \geq 1\; \exists\, k \geq 1\; s.t.\; \forall\, n \geq k \;\; |f_n(\omega) - f(\omega)| < \frac{1}{m}\right\} \\
= & \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{\omega : |f_n(\omega) - f(\omega)| < \frac{1}{m}\right\} \in \mathcal{F}.
\end{aligned}$$

(ii) The above definition depends on the measure $\mathbb{P}$. In general one does not have that

$$\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| \underset{n}{\longrightarrow} 0\}) = 1$$

if and only if

$$Q(\{\omega : |f_n(\omega) - f(\omega)| \underset{n}{\longrightarrow} 0\}) = 1$$

if Q is another measure on $\mathcal{F}$.

(iii) Only few properties of f_n are transferred to f by almost sure convergence. Take, for example, $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, and λ to be the Lebesgue measure ($\lambda([a, b]) = b - a$). Let f_n be

$$f_n(\omega) := \begin{cases} n^2 2^{n+1}\omega, & \omega \in \left[0, \frac{1}{2n}\right] \\ n2^{n+1} - n^2 2^{n+1}\omega, & \omega \in \left(\frac{1}{2n}, \frac{1}{n}\right] \\ 0, & \omega \in \left(\frac{1}{n}, 1\right]. \end{cases}$$

The function $f_n : [0,1] \to \mathbb{R}$ is continuous so that f_n is a random variable. Moreover $\lim_n f_n(\omega) = 0$ for all $\omega \in [0,1]$. On the other side

$$\int_0^1 f_n(\omega)\, d\lambda(\omega) = \int_0^1 f_n(t)\, dt = 2^n \xrightarrow[n]{} \infty.$$

A useful characterization of the almost sure convergence is given by

Proposition 1.1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_n, f : \Omega \to \mathbb{R}$ be random variables. Then the following assertions are equivalent.*

(i) $f_n \xrightarrow[a.s.]{} f$.

(ii) $\lim_n \mathbb{P}(\{\omega : \sup_{k \geq n} |f_k(\omega) - f(\omega)| > \varepsilon\}) = 0$ *for all* $\varepsilon > 0$.

Proof. For $\varepsilon > 0$ and $n \geq 1$ define

$$A_n^\varepsilon = \{\omega : \sup_{k \geq n} |f_k(\omega) - f(\omega)| > \varepsilon\} = \bigcup_{k=n}^{\infty} \{\omega : |f_k(\omega) - f(\omega)| > \varepsilon\} \in \mathcal{F}$$

so that, since $A_1^\varepsilon \supseteq A_2^\varepsilon \supseteq \dots$,

$$\lim_{n \to \infty} \mathbb{P}(A_n^\varepsilon) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} A_n^\varepsilon\right)$$

and

$$\bigcap_{n=1}^{\infty} A_n^\varepsilon = \{\omega : \forall\ n = 1, 2, \dots \ \sup_{k \geq n} |f_k(\omega) - f(\omega)| > \varepsilon\}.$$

(i)$\Longrightarrow$(ii) Let $\Omega_0 := \{\omega : \lim_n f_n(\omega) = f(\omega)\} \in \mathcal{F}$. Hence for all $\omega \in \Omega_0$ there exists $n(\omega) \geq 1$ with $\sup_{k \geq n(\omega)} |f_k(\omega) - f(\omega)| \leq \varepsilon$ so that

$$\Omega_0 \subseteq \left(\bigcap_{n=1}^{\infty} A_n^\varepsilon\right)^{\mathrm{c}} \quad \text{and} \quad \bigcap_{n=1}^{\infty} A_n^\varepsilon \subseteq \Omega_0^{\mathrm{c}}.$$

Hence

$$0 = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n^\varepsilon\right) = \lim_n \mathbb{P}\left(A_n^\varepsilon\right).$$

(ii)$\Longrightarrow$(i) We have

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} (A_n^\varepsilon)^{\mathrm{c}}\right) = 1 \quad \text{and} \quad \mathbb{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{n=1}^{\infty} \left(A_n^{\frac{1}{N}}\right)^{\mathrm{c}}\right) = 1\,.$$

Finally,

$$\omega \in \bigcap_{N=1}^{\infty} \bigcup_{n=1}^{\infty} \left(A_n^{\frac{1}{N}}\right)^{\mathrm{c}}$$

if and only if for all $N = 1, 2, \dots$ there exists $n \geq 1, 2, \dots$ such that

$$\sup_{k \geq n} |f_k(\omega) - f(\omega)| \leq \frac{1}{N}.$$

$\square$

1.2 Convergence in probability

Although we saw in Remark 1.1.2(iii) that a.s. convergence may be a weak notation, this notation is still sometimes too strong.

Example 1.2.1. $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, λ is the Lebesgue measure on $\mathcal{F}$. Define

$$\begin{aligned} f_1(\omega) &= \chi_{[0,\frac{1}{2})}(\omega)\,,\ f_2(\omega) = \chi_{[\frac{1}{2},1)}(\omega)\,, \\ f_3(\omega) &= \chi_{[0,\frac{1}{4})}(\omega)\,,\ f_4(\omega) = \chi_{[\frac{1}{4},\frac{1}{2})}(\omega)\,,\ldots,\ f_6(\omega) = \chi_{[\frac{3}{4},1)}(\omega)\,, \\ f_7(\omega) &= \chi_{[0,\frac{1}{8})}(\omega)\,,\ \ldots \end{aligned}$$

We have the feeling that $\lim_n f_n(\omega) = 0$, but in what sense? We do not have a.s. convergence since $\#\{n : f_n(\omega) = 1\} = \infty$ for all $\omega \in [0,1)$.

The way out is

Definition 1.2.2. Let $f_n\,, f : \Omega \to \mathbb{R}$ be random variables, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Then f_n *converges to* f *in probability* if

$$\lim_n \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) = 0$$

for all $\varepsilon > 0$. We will write $f_n \underset{\mathbb{P}}{\longrightarrow} f$.

Example (Example 1.2.1 continued)**.** We have that $f_n \underset{\lambda}{\longrightarrow} 0$. In fact,

$$\lim_n \lambda(\{\omega \in [0,1] : |f_n(\omega)| > \varepsilon\}) \le \lim_n \lambda(\{\omega \in [0,1] : f_n(\omega) \neq 0\}) = 0$$

since

$$\lambda(\{\omega \in [0,1] : f_n(\omega) \neq 0\}) = \begin{cases} \frac{1}{2} & \text{for } n = 1,2 \\ \frac{1}{4} & \text{for } n = 3,4,5,6 \\ \frac{1}{8} & \text{for } n = 7,\ldots \\ \vdots & \end{cases}.$$

Proposition 1.2.3. *Let $f_n, f : \Omega \to \mathbb{R}$ be random variables and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then one has the following:*

(i) *If $f_n \underset{a.s.}{\longrightarrow} f$, then $f_n \underset{\mathbb{P}}{\longrightarrow} f$.*

(ii) *If $f_n \underset{\mathbb{P}}{\longrightarrow} f$, then there exists $n_1 < n_2 < n_3 < \ldots$ such that $f_{n_k} \underset{a.s.}{\longrightarrow} f$, as $k \to \infty$.*

Proof. (i) Follows from Proposition 1.1.3 and

$$\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \le \mathbb{P}(\{\omega : \sup_{k\ge n} |f_k(\omega) - f(\omega)| > \varepsilon\}).$$

(ii) Let n_1 be the smallest $n \ge 1$ such that for all $k \ge n$ one has

$$\mathbb{P}\left(\left\{\omega : |f_k(\omega) - f(\omega)| > \frac{1}{2}\right\}\right) < \frac{1}{2}.$$

Let n_2 be the smallest $n > n_1$ such that for all $k \ge n$ one has

$$\mathbb{P}\left(\left\{\omega : |f_k(\omega) - f(\omega)| > \frac{1}{2^2}\right\}\right) < \frac{1}{2^2}.$$

Continuing the same way we get

$$\mathbb{P}\left(\left\{\omega : |f_k(\omega) - f(\omega)| > \frac{1}{2^\ell}\right\}\right) < \frac{1}{2^\ell}.$$

for $k \ge n_\ell$ and $1 \le n_1 < n_2 < n_3 < \dots$. Letting

$$A_\ell := \left\{\omega : |f_{n_\ell}(\omega) - f(\omega)| > \frac{1}{2^\ell}\right\}$$

we get

$$\mathbb{P}(A_\ell) < \frac{1}{2^\ell} \quad \text{and} \quad \sum_{\ell=1}^{\infty} \mathbb{P}(A_\ell) < \infty.$$

The Borel-Cantelli-lemma implies that

$$\mathbb{P}(\{\omega : \#\{n : \omega \in A_n\} = \infty\}) = 0.$$

Hence $\mathbb{P}(\{\omega : \#\{n : \omega \in A_n\} < \infty\}) = 1$. For those ω we have that

$$|f_{n_\ell}(\omega) - f(\omega)| \le \frac{1}{2^\ell}$$

for $\ell \ge \ell(\omega)$ which gives $f_{n_\ell} \xrightarrow[a.s.]{} f$. □

Example (Example 1.2.1 continued)**.** What is a possible sub-sequence? One can take

$$f_1 = \chi_{[0\frac{1}{2})},\ f_3 = \chi_{[0\frac{1}{4})},\ f_7 = \chi_{[0\frac{1}{8})}, \dots.$$

Proposition 1.2.4. *For random variables $f, g : \Omega \to \mathbb{R}$ we let*

$$d(f,g) := \int_\Omega \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} d\mathbb{P}(\omega),$$

Then the following assertions are equivalent:

(i) $d(f_n, f) \underset{n}{\longrightarrow} 0$.

(ii) $f_n \underset{\mathbb{P}}{\longrightarrow} f$.

Proof. (i) $\Longrightarrow$ (ii) $d(f_n, f) \underset{n}{\longrightarrow} 0$ means

$$\int_\Omega \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} d\mathbb{P}(\omega) \underset{n}{\longrightarrow} 0,$$

so that by Ĉebyšev's inequality, for $\lambda > 0$,

$$\lambda \mathbb{P}\left(\left\{\omega : \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} > \lambda\right\}\right) \leq \int_\Omega \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \underset{n}{\longrightarrow} 0.$$

Given $\varepsilon > 0$ we find $\lambda(\varepsilon) > 0$ such that if $|x| > \varepsilon$, then $\frac{|x|}{1+|x|} > \lambda(\varepsilon)$. Hence

$$\begin{aligned} \mathbb{P}\left(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}\right) &\leq \mathbb{P}\left(\left\{\omega : \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} > \lambda(\varepsilon)\right\}\right) \\ &\leq \frac{1}{\lambda(\varepsilon)} \int_\Omega \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \end{aligned}$$

and $\mathbb{P}\left(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}\right) \underset{n}{\longrightarrow} 0$.
(ii) $\Longrightarrow$ (i) For all $\varepsilon > 0$ we have that

$$\begin{aligned} \int_\Omega \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) &= \int_{\{|f_n - f| > \varepsilon\}} \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \\ &\quad + \int_{\{|f_n - f| \leq \varepsilon\}} \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \\ &\leq \mathbb{P}\left(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \varepsilon\}\right) + \frac{\varepsilon}{1 + \varepsilon}, \end{aligned}$$

since the function $\frac{x}{1+x} = 1 - \frac{1}{1+x}$ is monotone for $x \geq 0$. Given $\theta > 0$, we take $\varepsilon > 0$ such that $\frac{\varepsilon}{1+\varepsilon} \leq \frac{\theta}{2}$ and then $n_0 \geq 1$ such that $\mathbb{P}\left(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}\right) \leq \frac{\theta}{2}$ for $n \geq n_0$. Hence $d(f_n, f) \leq \theta$ for $n \geq n_0$. □

Proposition 1.2.5. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be fixed and* $f, g, h : \Omega \to \mathbb{R}$ *be random variables. Then one has that*

(i) $d(f, g) = 0$ *if and only if* $\mathbb{P}(f = g) = 1$.

(ii) $d(f, g) = d(g, f)$.

(iii) $d(f, h) \leq d(f, g) + d(g, h)$.

(iv) *If* $f_1, f_2, ... : \Omega \to \mathbb{R}$ *are random variables such that for all* $\varepsilon > 0$ *there exists an* $n(\varepsilon) \geq 1$ *such that* $d(f_n, f_m) \leq \varepsilon$ *for all* $m, n \geq n(\varepsilon)$, *i.e. if* $(f_n)_{n\geq 1}$ *is a Cauchy sequence with respect to* d, *then there exists a random variable* $f : \Omega \to \mathbb{R}$ *such that*

$$\lim_n d(f_n, f) = 0.$$

Proof. (i) We have that

$$\begin{aligned} d(f,g) = 0 \quad &\iff \quad \int_\Omega \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} d\mathbb{P}(\omega) = 0 \\ &\iff \quad \mathbb{P}\left(\left\{\omega : \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} = 0\right\}\right) = 1 \\ &\iff \quad \mathbb{P}\left(\{\omega : f(\omega) = g(\omega)\}\right) = 1 \\ &\iff \quad f = g \text{ a.s.} \end{aligned}$$

(ii) follows by definition.
(iii) Here we get

$$\begin{aligned} d(f,h) &= \int_\Omega \frac{|f(\omega) - h(\omega)|}{1 + |f(\omega) - h(\omega)|} d\mathbb{P}(\omega) \\ &\leq \int_\Omega \frac{|f(\omega) - g(\omega)| + |g(\omega) - h(\omega)|}{1 + |f(\omega) - g(\omega)| + |g(\omega) - h(\omega)|} d\mathbb{P}(\omega) \\ &\leq \int_\Omega \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} d\mathbb{P}(\omega) + \int_\Omega \frac{|g(\omega) - h(\omega)|}{1 + |g(\omega) - h(\omega)|} d\mathbb{P}(\omega) \\ &= d(f,g) + d(g,h). \end{aligned}$$

(iv) For the proof of this statement we need

Definition 1.2.6. A sequence $f_1, f_2, ... : \Omega \to \mathbb{R}$ of random variables is a *Cauchy sequence in probability* (or *fundamental in probability*) provided that for all $\varepsilon > 0$ there exists $n(\varepsilon) \geq 1$ such that for all $k, l \geq n(\varepsilon)$ one has that

$$\mathbb{P}(|f_k - f_l| > \varepsilon) \leq \varepsilon.$$

Proposition 1.2.7. *Let* $(f_n)_{n=1}^\infty$ *be a Cauchy sequence in probability. Then there exists a random variable* $f : \Omega \to \mathbb{R}$ *such that* $f_n \to_{\mathbb{P}} f$.

Proof. We proceed as in the proof of Proposition 1.2.3 and find $1 \leq n_1 < n_2 < \ldots$ such that

$$\mathbb{P}\left(\left\{\omega : |f_k(\omega) - f_l(\omega)| > \frac{1}{2^j}\right\}\right) < \frac{1}{2^j}$$

for $k, l \geq n_j$. Taking the sequence $(f_{n_j})_{j=1}^\infty$ we get that

$$\mathbb{P}\left(\left\{\omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \frac{1}{2^j}\right\}\right) < \frac{1}{2^j}$$

and

$$\sum_{j=1}^\infty \mathbb{P}\left(\left\{\omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \frac{1}{2^j}\right\}\right) < \infty.$$

Applying the Borel-Cantelli Lemma implies

$$\mathbb{P}\left(\left\{\omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \frac{1}{2^j} \text{ infinitely often}\right\}\right) = 0.$$

Hence

$$\mathbb{P}\left(\left\{\omega : \sum_{j=1}^\infty |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| < \infty\right\}\right) = 1.$$

We set

$$f(\omega) := \begin{cases} f_{n_1}(\omega) + \sum_{j=1}^\infty (f_{n_{j+1}}(\omega) - f_{n_j}(\omega)) & : \quad \sum_{j=1}^\infty |f_{n_{j+1}} - f_{n_j}| < \infty \\ 0 & : \quad \text{else} \end{cases}$$

and get that $f_{n_j} \xrightarrow[a.s.]{} f$. Finally, we have to check that $f_n \xrightarrow[\mathbb{P}]{} f$. For $\varepsilon > 0$ one gets

$$\begin{aligned} & \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \\ \leq\ & \mathbb{P}(\{\omega : |f_n(\omega) - f_{n_j}(\omega)| + |f_{n_j}(\omega) - f(\omega)| > \varepsilon\}) \\ \leq\ & \mathbb{P}\left(\left\{\omega : |f_n(\omega) - f_{n_j}(\omega)| > \frac{\varepsilon}{2}\right\}\right) + \mathbb{P}\left(\left\{\omega : |f_{n_j}(\omega) - f(\omega)| > \frac{\varepsilon}{2}\right\}\right) \end{aligned}$$

and

$$\lim_{n,j\to\infty} \mathbb{P}\left(\left\{\omega : |f_n(\omega) - f_{n_j}(\omega)| > \frac{\varepsilon}{2}\right\}\right) = 0,$$

$$\lim_{j\to\infty} \mathbb{P}\left(\left\{\omega : |f_{n_j}(\omega) - f(\omega)| > \frac{\varepsilon}{2}\right\}\right) = 0.$$

□

Now we can finish our proof by showing that $(f_n)_{n=1}^\infty$ is a Cauchy sequence with respect to d if and only if it is a Cauchy sequence in probability.

Assume that $(f_n)_{n=1}^\infty$ is a Cauchy sequence with respect to d. For $\lambda > 0$ we have that

$$\lambda \mathbb{P}\left(\left\{\omega : \frac{|f_k(\omega) - f_l(\omega)|}{1 + |f_k(\omega) - f_l(\omega)|} > \lambda\right\}\right) \leq d(f_k, f_l) \leq \eta$$

for $k, l \geq n(\eta) \geq 1$. For $\lambda := \frac{\varepsilon}{1+\varepsilon}$ with $\varepsilon > 0$ this gives that

$$\frac{\varepsilon}{1+\varepsilon} \mathbb{P}\left(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}\right) \leq \eta$$

for $k, l \geq n(\eta) \geq 1$. Choosing $\eta = \varepsilon^2$ we end up with

$$\mathbb{P}\left(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}\right) \leq \varepsilon(1+\varepsilon)$$

for $k, l \geq n(\varepsilon^2) \geq 1$. Hence $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability.

Now assume that $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability. Then for all $\varepsilon > 0$ there exists $n(\varepsilon) \geq 1$ such that for all $k, l \geq n(\varepsilon)$ $\mathbb{P}\left(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}\right) \leq \varepsilon$. Consequently,

$$\begin{aligned}
\int_\Omega \frac{|f_k(\omega) - f_l(\omega)|}{1 + |f_k(\omega) - f_l(\omega)} d\mathbb{P}(\omega) &\leq \mathbb{P}(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}) + \frac{\varepsilon}{1+\varepsilon} \\
&\leq \varepsilon + \frac{\varepsilon}{1+\varepsilon}
\end{aligned}$$

for $k, l \geq n(\varepsilon) \geq 1$. □

To formalize the above we recall the notion of a metric space:

Definition 1.2.8. Let $M \neq \emptyset$. The pair (M, d) is called *metric space,* if $d : M \times M \to [0, \infty)$ satisfies

(i) $d(x, y) = 0$ if and only if $x = y$ (reflexivity),

(ii) $d(x, y) = d(y, x)$ (symmetry),

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

Convergence in probability can be described by a suitable metric space that consists of equivalence classes.

Definition 1.2.9. Let M be an arbitrary set. We say that a relation $x \sim y$ is *an equivalence class relation* on M, provided that

(i) $x \sim x$ for all $x \in M$ (reflexivity),

(ii) if $x \sim y$, then $y \sim x$ (symmetry),

(iii) if $x \sim y$ and $y \sim z$, then $x \sim z$ (transitivity).

For two elements $x, y \in M$ we have either $x \sim y$ or x and y are not in the relation $x \sim y$. Based on this one introduces *equivalence classes:* We have

$$M = \bigcup_{i \in I} M_i,$$

with

(i) $M_i \neq \emptyset$,

(ii) $M_i \cap M_j = \emptyset$, if $i \neq j$,

(iii) two elements $x, y \in M$ belong to the same set M_i if and only if $x \sim y$.

An element $x_i \in M_i$ is called *representative.* In probability theory we use

Definition 1.2.10. (i) For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $f : \Omega \to \mathbb{R}$ we let

$$\hat{f} := \{g : \Omega \to \mathbb{R} \text{ random variable}, \mathbb{P}(\{\omega \in \Omega : f(\omega) = g(\omega)\}) = 1\}.$$

The set $\hat{f}$ is called *equivalence class* (with respect to $\mathbb{P}$) and $f \in \hat{f}$ is called representative.

(ii) $\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all random variables $f : \Omega \to \mathbb{R}$.

(iii) $L_0(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all equivalence classes from $\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$.

(iv) For $f \in \hat{f}$ and $g \in \hat{g}$ we let

$$\hat{d}(\hat{f}, \hat{g}) := d(f, g).$$

In other words, for random variables $f, g : \Omega \to \mathbb{R}$ one has $f \sim g$ if and only if

$$\mathbb{P}(\{\omega : f(\omega) = g(\omega)\}) = 1.$$

For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ and $\lambda \in \mathbb{R}$ we introduce the linear operations

$$\begin{aligned} \lambda \hat{f} &:= \widehat{\lambda f}, \\ \hat{f} + \hat{g} &:= \widehat{f + g}. \end{aligned}$$

Finally, we have

Proposition 1.2.11. *The space $[L_0(\Omega, \mathcal{F}, \mathbb{P}), \hat{d}]$ is a linear complete metric space.*

1.3 Convergence in mean

Definition 1.3.1. Let $p \in (0, \infty)$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Given random variables $f_n, f : \Omega \to \mathbb{R}$ we say that f_n converges to f *with respect to the p-th mean* ($f_n \xrightarrow[L_p]{} f$) provided that

$$\lim_{n \to \infty} \int_\Omega |f_n(\omega) - f(\omega)|^p d\mathbb{P}(\omega) = 0.$$

Since f_n and f are random variables and $x \mapsto |x|^p$ is a continuous function, $\omega \mapsto |f_n(\omega) - f(\omega)|^p$ is a non-negative random variable and we may integrate.

Proposition 1.3.2. *If $f_n \xrightarrow[L_p]{} f$, then $f_n \xrightarrow[\mathbb{P}]{} f$.*

Proof. Let $\varepsilon > 0$. Then

$$\begin{aligned}\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) &= \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)|^p > \varepsilon^p\}) \\ &\leq \frac{1}{\varepsilon^p} \int_\Omega |f_n(\omega) - f(\omega)|^p d\mathbb{P}(\omega) \xrightarrow[n]{} 0.\end{aligned}$$

□

Proposition 1.3.3 (Minkowski inequality)**.** *Let $0 < p < \infty$ and $f, g : \Omega \to \mathbb{R}$ be random variables. Then one has*

$$\begin{aligned}&\left(\int_\Omega |f(\omega) + g(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}} \\ &\leq c_p \left[\left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}} + \left(\int_\Omega |g(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}}\right],\end{aligned}$$

where $c_p = 1$ for $1 \leq p < \infty$ and $c_p = 2^{\frac{1}{p}-1}$ for $0 < p < 1$.

Proof. We only prove the case $0 < p \leq 1$. Here we get $|a + b|^p \leq |a|^p + |b|^p$ for all $a, b \in \mathbb{R}$, so that

$$\begin{aligned}&\left(\int_\Omega |f(\omega) + g(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}} \\ &\leq \left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega) + \int_\Omega |g(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}} \\ &\leq 2^{\frac{1}{p}-1} \left[\left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}} + \left(\int_\Omega |g(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}}\right],\end{aligned}$$

where, for $1 \leq q = \frac{1}{p} < \infty$, we have used $|a + b|^q \leq 2^{q-1}(|a|^q + |b|^q)$. □

From the Minkowski inequality, we get immediately

Proposition 1.3.4. *If $f_n, g_n, f, g : \Omega \to \mathbb{R}$, $f_n \xrightarrow[L_p]{} f$ and $g_n \xrightarrow[L_p]{} g$, then $f_n + g_n \xrightarrow[L_p]{} f + g$.*

Now, we recall

Proposition 1.3.5 (Hölder's inequality)**.** *Let $f, g : \Omega \to \mathbb{R}$ be random variables, $1 < p < \infty$, $1 < q < \infty$ and $1 = \frac{1}{p} + \frac{1}{q}$. Then*

$$\int_\Omega |f(\omega)g(\omega)| d\mathbb{P}(\omega) \leq \left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega)\right)^{\frac{1}{p}} \left(\int_\Omega |g(\omega)|^q d\mathbb{P}(\omega)\right)^{\frac{1}{q}}.$$

As an application we get

Proposition 1.3.6. *Let $0 < p < q < \infty$. Then from $f_n \xrightarrow[L_q]{} f$ it follows $f_n \xrightarrow[L_p]{} f$.*

Proof. Let $r := \frac{q}{p} \in (1, \infty)$ and let $1 = \frac{1}{r} + \frac{1}{r*}$. Then

$$
\begin{aligned}
& \int_\Omega |f_n(\omega) - f(\omega)|^p d\mathbb{P}(\omega) \\
= & \int_\Omega |f_n(\omega) - f(\omega)|^p \cdot 1 d\mathbb{P}(\omega) \\
\leq & \left(\int_\Omega (|f_n(\omega) - f(\omega)|^p)^r \, d\mathbb{P}(\omega) \right)^{\frac{1}{r}} \left(\int_\Omega 1^{r*} d\mathbb{P}(\omega) \right)^{\frac{1}{r*}} \\
= & \left(\int_\Omega |f_n(\omega) - f(\omega)|^q d\mathbb{P}(\omega) \right)^{\frac{p}{q}}.
\end{aligned}
$$

□

We conclude with the Lorentz-spaces L_p. Before we do this we recall the notion of a Banach space.

Definition 1.3.7. Let X be a linear space and $\|\cdot\| : X \to [0, \infty)$. Then $[X, \|\cdot\|]$ is called Banach space, provided that

(i) $\|x\| = 0$ if and only if $x = 0$,

(ii) $\|\lambda x\| = |\lambda| \|x\|$ for all $\lambda \in \mathbb{R}$ ($\lambda \in \mathbb{C}$) and $x \in X$,

(iii) $\|x + y\| \leq \|x\| + \|y\|$ and

(iv) if $(x_n)_{n=1}^\infty \subseteq X$ is a Cauchy sequence, i.e. for all $\varepsilon > 0$ there exists $n(\varepsilon) \geq 1$ with $\|x_m - x_n\| \leq \varepsilon$ whenever $m, n \geq n(\varepsilon)$, then there is an $x \in X$ such that
$$\lim_n \|x_n - x\| = 0.$$

Remark 1.3.8. Properties (i), (ii), and (iii) say that $\|\cdot\|$ is a norm.

Example 1.3.9. For $X = \mathbb{R}^n$ and $\|x\| = \|(x_1, ..., x_n)\| := (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}}$ we obtain a norm on $\mathbb{R}^n$.

In our setting there are some problems with property (i). For this reason we have to work with the equivalence classes introduced in Definition 1.2.10.

Definition 1.3.10. For $p \in [1, \infty)$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we let

(i) $\mathcal{L}_p := \left\{ f : \Omega \to \mathbb{R} \text{ random variable}, \|f\|_{L_p} := \left(\int_\Omega |f|^p d\mathbb{P} \right)^{\frac{1}{p}} < \infty \right\}$,

(ii) $L_p(\Omega, \mathcal{F}, \mathbb{P}) := \left\{ \hat{f} : f \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P}) \right\}$ and $\|\hat{f}\|_{L_p} := \|f\|_{L_p}$.

Proposition 1.3.11. $\left[L_p(\Omega, \mathcal{F}, \mathbb{P}), \| \cdot \|_{L_p} \right]$ *is a Banach space.*

Proof. (i) $\|\hat{f}\|_{L_p} = 0$ if and only if $f = 0$ a.s. if and only if $\hat{f} = 0$.
(ii) is clear and (iii) is a consequence of the Minkowski inequality.
(iv) Assume a Cauchy sequence $(\hat{f}_n)_{n=1}^\infty \subseteq L_p$. Then $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability since

$$\mathbb{P}\left(\{ \omega : |f_n(\omega) - f_m(\omega)| > \lambda \} \right) \le \frac{1}{\lambda^p} \|\hat{f}_n(\omega) - \hat{f}_m(\omega)\|_{L_p}^p \le \varepsilon$$

for $n, m \ge n(\lambda, \varepsilon) \ge 1$. Hence there is a limit $f : \Omega \to \mathbb{R}$ such that $f_n \xrightarrow[\mathbb{P}]{} f$. It remains to show that $f_n \xrightarrow[L_p]{} f$ as well. We pick a sub-sequence $(f_{n_k})_{k=1}^\infty$ such that $\lim_k f_{n_k} = f$ $\mathbb{P} - a.s.$. Applying the lemma of Fatou gives

$$\int_\Omega |f_m(\omega) - f(\omega)|^p d\mathbb{P}(\omega) \le \liminf_k \int_\Omega |f_m(\omega) - f_{n_k}(\omega)|^p d\mathbb{P}(\omega) \le \varepsilon$$

for $m \ge m(\varepsilon) \ge 1$. $\square$

Remark 1.3.12. (i) One can define $L_p(\Omega, \mathcal{F}, \mathbb{P})$ for $0 < p < 1$ in the same way. We get

(a) $\|x + y\|_{L_p} \le c_p \left(\|x\|_{L_p} + \|y\|_{L_p} \right)$ (quasi-triangle inequality) and

(b) (L_p, d) is a complete metric space with $d(\hat{f}, \hat{g}) := \|\hat{f} - \hat{g}\|_{L_p}^p$.

(ii) One can define $L_p(\Omega, \mathcal{F}, \mathbb{P})$ for $p = \infty$ as well, where

$$\|\hat{f}\|_{L_\infty} := \operatorname{ess\,sup}_{\omega \in \Omega} |f(\omega)| = \inf \left\{ \sup_{\omega \in \Omega \setminus N} |f(\omega)| : N \in \mathcal{F}, \mathbb{P}(N) = 0 \right\}.$$

Again, $[L_\infty(\Omega, \mathcal{F}, \mathbb{P}), \| \cdot \|_{L_\infty}]$ is a Banach space.

Proposition 1.3.13. *Assume that* $\lim_n f_n = f$ *a.s.,* $p \in (0, \infty)$ *and* $\mathbb{E} \sup_n |f_n|^p < \infty$. *Then* $\hat{f} \in L_p$ *and* $\lim_n \mathbb{E}|f_n - f|^p = 0$.

Proof. By Lebesgue's theorem of dominated convergence we get that

$$\mathbb{E}|f|^p = \lim_n \mathbb{E}|f_n|^p < \infty.$$

Hence

$$\mathbb{E} \sup_n |f_n - f|^p \le c_p \mathbb{E} \left(\sup_n |f_n|^p + |f|^p \right) < \infty$$

and, again by dominated convergence,

$$0 = \mathbb{E} \lim_n |f_n - f|^p = \lim_n \mathbb{E}|f_n - f|^p.$$

□

In the last proposition one can replace the condition $\lim_n f_n = f$ a.s. by $\lim_n f_n = f$ in probability. Indeed, assuming that $\|f_{n_k} - f\|^p \geq \delta > 0$ for a sub-sequence $n_1 < n_2 < \cdots$, we find one more sub-sequence $(n_{k_l})_{l=1}^{\infty}$ such that $\lim_l f_{n_{k_l}} = f$ a.s. and get $\lim_l \mathbb{E}|f_{n_{k_l}} - f|^p = 0$ which is a contradiction.

Summary:

a.s. convergence	$\Longrightarrow$	L_p-convergence if $\mathbb{E}\sup_n |f_n|^p < \infty$
L_p-convergence	$\Longrightarrow$	a.s. convergence for a sub-sequence
a.s. convergence	$\Longrightarrow$	convergence in probability
convergence in probability	$\Longrightarrow$	a.s. convergence for a sub-sequence
L_p-convergence	$\Longrightarrow$	convergence in probability
convergence in probability	$\Longrightarrow$	L_p-convergence if $\mathbb{E}\sup_n |f_n|^p < \infty$

1.4 Uniform integrability

Now we want to understand the condition $\mathbb{E}\sup_n |f_n|^p < \infty$ for $p = 1$ from Proposition 1.3.13 better. We recall that $\int_\Omega |f| d\mathbb{P} < \infty$ implies that

$$\lim_{c \to \infty} \int_{\{|f| \geq c\}} |f| d\mathbb{P} = 0.$$

In fact, the latter condition is (of course) equivalent to $\int_\Omega |f| d\mathbb{P} < \infty$. This leads us to the following definition of uniform integrability.

Definition 1.4.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_i : \Omega \to \mathbb{R}$, $i \in I$, be a family of random variables. Then the family $(f_i)_{i \in I}$ is called *uniformly integrable* provided that for all $\varepsilon > 0$ there is a constant $c > 0$ such that

$$\sup_{i \in I} \int_{|f_i| \geq c} |f_i| d\mathbb{P} \leq \varepsilon,$$

or equivalently,

$$\lim_{c \uparrow \infty} \sup_{i \in I} \int_{|f_i| \geq c} |f_i| d\mathbb{P} = 0.$$

Example 1.4.2. Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and $f_n(t) := n\chi_{[0,\frac{1}{n}]}(t)$, $n \in I = \{1, 2, 3, ...\}$. This family is not uniformly integrable because for any $c > 0$ we have that

$$\sup_n \int_{|f_n| \geq c} |f_n(t)| dt = 1.$$

An important sufficient criteria is

Lemma 1.4.3. *Let $G : [0,\infty) \to [0,\infty)$ non-negative and increasing such that*

$$\lim_{t\to\infty} \frac{G(t)}{t} = \infty$$

and $(f_i)_{i\in I}$ be a family of random variables $f_i : \Omega \to \mathbb{R}$ such that

$$\sup_{i\in I} \mathbb{E}G(|f_i|) < \infty.$$

Then $(f_i)_{i\in I}$ is uniformly integrable.

Proof. We let $\varepsilon > 0$ and $M := \sup_{i\in I} \mathbb{E}G(|f_i|)$ and find a $c > 0$ such that

$$\frac{M}{\varepsilon} \leq \frac{G(t)}{t}$$

for $t \geq c$. Then

$$\int_{|f_i|\geq c} |f_i| d\mathbb{P} \leq \frac{\varepsilon}{M} \int_{|f_i|\geq c} G(|f_i|) d\mathbb{P} \leq \varepsilon.$$

$\square$

Example 1.4.4.
(i) Let $G(t) = t^p$ with $1 < p < \infty$. Hence $\sup_{i\in I} \mathbb{E}|f_i|^p < \infty$ implies that $(f_i)_{i\in I}$ is uniformly integrable.

(ii) One cannot take the function $G(t) = t$ as we have in Example 1.4.2 that $\mathbb{E}|f_n| = 1$.

The main statement of this section is

Proposition 1.4.5. *Assume random variables $f_n, f : \Omega \to \mathbb{R}$ such that $f_n \to_{\mathbb{P}} f$, $\mathbb{E}|f_n| < \infty$ and $\mathbb{E}|f| < \infty$. Then the following assertions are equivalent:*

(i) *$f_n \to f$ in L_1.*

(ii) *$(f_n)_{n=1}^\infty$ is u.i.*

(iii) *$\|f_n\|_{L_1} \to_n \|f\|_{L_1}$.*

Remark 1.4.6. The proposition is applied in the following way: Given a sequence $f_n \to_{\mathbb{P}} f$ such that $(f_n)_{n=1}^\infty$ is uniformly integrable, we get $f_n \to f$ in L_1. Moreover, the statement says that the uniform integrability is a necessary condition in order to have this implication. From the mathematics side, the implication (iii) $\Longrightarrow$ (i) is rather surprising.

For the proof of Proposition 1.4.5 we need some lemmata:

Lemma 1.4.7. *The conditions $f_n \to_{\mathbb{P}} f$, $|f_n| \le g$ and $|f| \le g$ for some g with $\mathbb{E}g < \infty$ imply that*

$$\lim_n \int_\Omega f_n d\mathbb{P} = \int_\Omega f d\mathbb{P}.$$

Proof. Assume that the conclusion is not true. Then there is a $\varepsilon > 0$ and a sub-sequence $n_1 < n_2 < n_3 < \cdots$ such that

$$\left| \int_\Omega f_{n_k} d\mathbb{P} - \int_\Omega f d\mathbb{P} \right| \ge \varepsilon.$$

But we can find one more sub-sequence n_{k_l} such that

$$f_{n_{k_l}} \to f \quad a.s. \text{ as } l \to \infty.$$

Applying dominated convergence yields in this case that

$$\lim_l \int_\Omega f_{n_{k_l}} d\mathbb{P} = \int_\Omega f d\mathbb{P}$$

which is a contradiction. $\square$

Lemma 1.4.8. *For $f \in L_1$ and $\mathbb{P}(A_n) \to_n 0$ one has*

$$\lim_n \int_{A_n} f d\mathbb{P} = 0.$$

Proof. We apply Lemma 1.4.7 to $\widetilde{f_n} := f\chi_{A_n}$ and get

$$|\widetilde{f_n}| \le |f| =: g \quad \text{and} \quad \widetilde{f_n} \to_{\mathbb{P}} 0$$

because

$$\mathbb{P}(|\widetilde{f_n}| > c) \le \mathbb{P}(A_n) \to_n 0.$$

$\square$

Proof of Proposition 1.4.5. $(i) \Longrightarrow (ii)$. We have that

$$\begin{aligned}
\int_{\{|f_n| \ge c\}} |f_n| d\mathbb{P} &\le \int_{\{|f_n| \ge c\}} |f_n - f| d\mathbb{P} + \int_{\{|f_n| \ge c\}} |f| d\mathbb{P} \\
&\le \|f_n - f\|_{L_1} + \int_{\{|f_n - f| \ge c/2\}} |f| d\mathbb{P} + \int_{\{|f| \ge c/2\}} |f| d\mathbb{P}
\end{aligned}$$

and

$$\lim_n \|f_n - f\|_{L_1} = \lim_{c \uparrow \infty} \int_{\{|f| \ge c/2\}} |f| d\mathbb{P} = 0.$$

For the middle term we use Lemma 1.4.8 to deduce that

$$\int_{\{|f_n| \ge c\}} |f_n| d\mathbb{P} \le \varepsilon$$

for $c \geq c(\varepsilon)$ and $n \geq n(c(\varepsilon), \varepsilon)$.

$(ii) \Longrightarrow (iii)$ By Lemma 1.4.7 we get that

$$\int_\Omega (|f_n| \wedge c) d\mathbb{P} \to \int_\Omega (|f| \wedge c) d\mathbb{P}$$

because $(|f_n| \wedge c) \to_{\mathbb{P}} (|f| \wedge c)$ which is left as an exercise. But then

$$\begin{aligned} & |\mathbb{E}|f_n| - \mathbb{E}|f|| \\ \leq\ & |\mathbb{E}(|f_n| \wedge c) - \mathbb{E}(|f| \wedge c)| + |\mathbb{E}|f_n| - \mathbb{E}(|f_n| \wedge c)| \\ & \qquad + |\mathbb{E}|f| - \mathbb{E}(|f| \wedge c)| \\ =\ & |\mathbb{E}(|f_n| \wedge c) - \mathbb{E}(|f| \wedge c)| + \int_{\{|f_n| \geq c\}} |f_n| d\mathbb{P} + \int_{\{|f| \geq c\}} |f| d\mathbb{P}. \end{aligned}$$

$(iii) \Longrightarrow (i)$ We have that

$$||f_n - f| - |f_n| + |f|| \leq 2|f|$$

and

$$|f_n - f| - |f_n| + |f| \to_{\mathbb{P}} 0.$$

Hence, by Lemma 1.4.7,

$$\mathbb{E}[|f_n - f| - |f_n| + |f|] \to_n 0$$

and

$$\mathbb{E}|f_n - f| \to_n 0$$

because $\mathbb{E}|f_n| \to_n \mathbb{E}|f|$. $\square$

1.5 Independence

In this section we will recall the important notion of independence.

Definition 1.5.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(f_i)_{i \in I}$ with $I \neq \emptyset$ be a family of random variables $f_i : \Omega \to \mathbb{R}$. The family $(f_i)_{i \in I}$ is called *independent* provided that for all distinct $i_1, ..., i_n \in I$ and $B_1, ..., B_n \in \mathcal{B}(\mathbb{R})$ one has that

$$\mathbb{P}(f_{i_1} \in B_1, ..., f_{i_n} \in B_n) = \mathbb{P}(f_{i_1} \in B_1) \cdots \mathbb{P}(f_{i_n} \in B_n).$$

The main mathematical question is in the beginning to what extent independent random variables do exist. This question one might easily overlook. The positive answer is

Proposition 1.5.2. *Let* $I = \{1, 2, ..., n\}$ *or* $I = \{1, 2, 3, ...\}$. *Given probability measures* μ_k *on* $\mathcal{B}(\mathbb{R})$, $k \in I$, *there exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and independent random variables* $(f_k)_{k \in I}$, $f_k : \Omega \to \mathbb{R}$ *such that* $\mathrm{law}(f_k) = \mu_k$, *i.e.*

$$\mathbb{P}(f_k \in B) = \mu_k(B) \quad \textit{for all} \quad B \in \mathcal{B}(\mathbb{R}).$$

Proof. (i) Assume that $I = \{1, 2, ..., n\}$ and take $\Omega := \mathbb{R}^n$, $\mathcal{F} := \mathcal{B}(\mathbb{R}^n)$, and $\mathbb{P} := \otimes_{k=1}^n \mu_k$. As random variables we choose the coordinate functionals

$$f_k : \Omega \to \mathbb{R} \quad \text{with} \quad f_k(x_1, ..., x_n) := x_k.$$

The maps f_k are measurable as

$$f_k^{-1}(B_k) = \mathbb{R} \times \cdots \times \mathbb{R} \times B_k \times \mathbb{R} \times \cdots \times \mathbb{R} \in \mathcal{B}(\mathbb{R}^n)$$

for all $B_k \in \mathcal{B}(\mathbb{R})$. Finally,

$$\begin{aligned}
\mathbb{P}(f_1 \in B_1, ..., f_n \in B_n) &= \mathbb{P}(B_1 \times \cdots \times B_n) \\
&= \otimes_{k=1}^n \mu_k(B_1 \times \cdots \times B_n) \\
&= \prod_{k=1}^n \mu_k(B_k)
\end{aligned}$$

and, by fixing k and letting $B_l = \mathbb{R}$ for $l \neq k$,

$$\begin{aligned}
\mathbb{P}(f_k \in B_k) &= \mathbb{P}(f_1 \in \mathbb{R}, ..., f_{k-1} \in \mathbb{R}, f_k \in B_k, f_{k+1} \in \mathbb{R}, ..., f_n \in \mathbb{R}) \\
&= \cdots \\
&= \mu_1(\mathbb{R}) \cdots \mu_{k-1}(\mathbb{R}) \mu_k(B_k) \mu_{k+1}(\mathbb{R}) \cdots \mu_n(\mathbb{R}) \\
&= \mu_k(B_k)
\end{aligned}$$

which proves our statement for a finite index set I.

(ii) Now let us turn to the case $I = \{1, 2, 3, ...\}$. As measurable space we choose $(\Omega, \mathcal{F}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ with

$$\mathbb{R}^{\mathbb{N}} := \{(x_k)_{k=1}^\infty : x_k \in \mathbb{R}\}$$

and $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ being the smallest σ-algebra containing all cylinder sets

$$\mathbb{R} \times \cdots \times \mathbb{R} \times B_k \times \mathbb{R} \times \mathbb{R} \times \cdots \quad \text{with} \quad B_k \in \mathcal{B}(\mathbb{R}).$$

As family of random variables $(f_k)_{k=1}^\infty$ we again use the coordinate functionals $f_k(x_1, x_2, ...) := x_k$. Using exactly the same argument as in (i) (the reader should check this) the proof is complete if we find a probability measure $\mathbb{P}$ on $\mathcal{F}$ such that

$$\mathbb{P}(B_1 \times B_2 \times \cdots \times B_n \times \mathbb{R} \times \mathbb{R} \cdots) = \mu_1(B_1) \cdots \mu_n(B_n).$$

Obviously, this measure will be denoted by

$$\otimes_{k=1}^{\infty}\mu_k.$$

We use Carathéodory's theorem to construct this infinite product measure. As algebra we take $\mathcal{A}$ to be

$$\{\text{ finite unions of } (a_1,b_1]\times\cdots\times(a_n,b_n]\times\mathbb{R}\times\mathbb{R}\times\cdots, -\infty\le a_k\le b_k\le\infty\}$$

with the convention that $(a,\infty]=(a,\infty)$. It is clear that $\mathcal{A}$ is an algebra and that $\sigma(\mathcal{A})=\mathcal{B}(\mathbb{R}^{\mathbb{N}})$. Define the 'pre-measure' $\mathbb{P}_0$ on $\mathcal{A}$ by

$$\mathbb{P}_0((a_1,b_1]\times\cdots\times(a_n,b_n]\times\mathbb{R}\times\mathbb{R}\times\cdots)=\prod_{k=1}^{n}\mu_k((a_k,b_k]).$$

We skip the formal check that $\mathbb{P}_0$ is correctly defined. But we indicate that $\mathbb{P}_0$ is σ-additive **on** $\mathcal{A}$. Here we have to show that

$$\mathbb{P}_0\left(\bigcup_{l=1}^{\infty}A_l\right)=\sum_{l=1}^{\infty}\mathbb{P}_0(A_l)\quad\text{for}\quad A_1,A_2,...\in\mathcal{A}\quad\textbf{and}\quad\bigcup_{l=1}^{\infty}A_l\in\mathcal{A}$$

whenever $A_1,A_2,...$ are pair-wise disjoint. It follows from the properties of finite product measures and the monotonicity of $\mathbb{P}_0$ that

$$\sum_{l=1}^{L}\mathbb{P}_0(A_l)\le\mathbb{P}_0\left(\bigcup_{l=1}^{L}A_l\right)\le\mathbb{P}_0\left(\bigcup_{l=1}^{\infty}A_l\right).$$

By $L\uparrow\infty$ this implies

$$\sum_{l=1}^{\infty}\mathbb{P}_0(A_l)\le\mathbb{P}_0\left(\bigcup_{l=1}^{\infty}A_l\right).$$

Assume that we have a strict inequality, i.e. there is some $\delta>0$ such that

$$\mathbb{P}_0\left(\bigcup_{l=1}^{\infty}A_l\right)\ge\sum_{l=1}^{\infty}\mathbb{P}_0(A_l)+\delta.$$

Defining

$$B_L:=\left(\bigcup_{l=1}^{\infty}A_l\right)\setminus(A_1\cup\cdots\cup A_L)$$

one gets $B_1\supseteq B_2\supseteq\cdots$ and $\bigcap_{l=1}^{\infty}B_l=\emptyset$. With this notation the above assumption translates into

$$\mathbb{P}_0(B_L)\ge\delta\quad\text{for all}\quad L=1,2,...$$

By induction one can define a sequence of sets C_n such that

- $C_n \subseteq B_n$,
- $C_1 \supseteq C_2 \supseteq \cdots$,
- C_n is a finite union of sets of type $[a_1, b_1] \times \cdots \times [a_K, b_K] \times \mathbb{R} \times \mathbb{R} \times \cdots$.

Take for example $C_1 \subseteq B_1$ with $\mathbb{P}_0(B_1 \setminus C_1) \leq \delta/8$ and continue with $C_2 \subseteq B_2 \cap C_1$ with $\mathbb{P}_0((B_2 \cap C_1) \setminus C_2) \leq \delta/16$ etc. Hence

$$\emptyset = \bigcap_{l=1}^{\infty} B_l \supseteq \bigcap_{l=1}^{\infty} C_l \neq \emptyset$$

which is a contradiction. □

1.6 Two examples for almost sure convergence

1.6.1 Strong law of large numbers (SLLN)

There are different versions of the SLLN. We prove a version which goes back to Kolmogorov.

Proposition 1.6.1 (Kolmogorov [1]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ random variables such that*

(i) *$f_1, f_2, \ldots$ are independent,*

(ii) *$f_1, f_2, \ldots$ have the same distribution and*

(iii) *$\mathbb{E}|f_1| < \infty$.*

Then one has

$$\lim_{n\to\infty} \frac{1}{n} \left(f_1(\omega) + \cdots + f_n(\omega) \right) = \mathbb{E} f_1 \ \ a.s. \ .$$

For the proof we need a couple of preparations.

Proposition 1.6.2 (Inequality of Kolmogorov). *Let $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be independent and such that $\mathbb{E} f_n = 0$ and $\mathbb{E} f_n^2 < \infty$. Then, for $S_k := f_1 + \cdots + f_k$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \varepsilon \right) \leq \frac{\mathbb{E} S_n^2}{\varepsilon^2}$$

for all $\varepsilon > 0$.

[1] Andrey Nikolaevich Kolmogorov, 25/04/1903 (Tambov, Russia) - 20/10/1987 (Moscow, Russia).

Proof. Let $A_k := \{|S_1| < \varepsilon, \ldots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\}$. Then

$$\begin{aligned}
\varepsilon^2 \mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq \varepsilon\right) &= \varepsilon^2 \sum_{k=1}^n \mathbb{P}(A_k) \leq \sum_{k=1}^n \int_{A_k} S_k^2 d\mathbb{P} \\
\leq \sum_{k=1}^n \Bigg[\int_{A_k} S_k^2 d\mathbb{P} &+ 2\int_{A_k} S_k \left(\sum_{i=k+1}^n f_i\right) d\mathbb{P} + \int_{A_k}\left(\sum_{i=k+1}^n f_i\right)^2 d\mathbb{P}\Bigg] \\
= \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} &\leq \int_\Omega S_n^2 d\mathbb{P}.
\end{aligned}$$

□

Proposition 1.6.3 (Kolmogorov-Chinčin)**.** *Let $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be independent and such that $\mathbb{E} f_n = 0$ and $\sum_{n=1}^\infty \mathbb{E} f_n^2 < \infty$. Then*

$$\mathbb{P}\left(\left\{\omega : \sum_{n=1}^\infty f_n(\omega) \text{ is convergent}\right\}\right) = 1.$$

Proof. Given $\varepsilon > 0$, we get

$$\begin{aligned}
\mathbb{P}\left(\sup_{k\geq 1} |S_{n+k} - S_n| \geq \varepsilon\right) &= \lim_{N\to\infty} \mathbb{P}\left(\sup_{1\leq k\leq N} |S_{n+k} - S_n| \geq \varepsilon\right) \\
&\leq \lim_{N\to\infty} \frac{1}{\varepsilon^2} \sum_{k=n+1}^N \mathbb{E}|f_k|^2 \\
&= \frac{1}{\varepsilon^2} \sum_{k=n+1}^\infty \mathbb{E}|f_k|^2 \underset{n}{\longrightarrow} 0,
\end{aligned}$$

where we have used Kolmogorov's inequality. Hence

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geq 1} |S_{n+k} - S_n| \geq \varepsilon\right) = 0.$$

As an exercise we show that this is equivalent to

$$\mathbb{P}\left((S_n)_{n=1}^\infty \text{ is a Cauchy sequence}\right) = 1.$$

□

Next we need two Lemmata from analysis:

Lemma 1.6.4 (Töplitz [2]). *Let $a_n \geq 0$, $b_n := a_1 + a_2 + \cdots + a_n > 0$, $\lim_n b_n = \infty$. Assume $(x_n)_{n=1}^\infty \subseteq \mathbb{R}$ such that $\lim_n x_n = x$. Then*

$$\frac{1}{b_n}\sum_{i=1}^n a_i x_i \underset{n}{\longrightarrow} x.$$

Proof. The proof is an exercise. □

Lemma 1.6.5 (Kronecker [3]). *Let $0 < b_1 \leq b_2 \leq b_3 \leq \ldots$ with $b_n \to \infty$ and $(x_n)_{n=1}^\infty \subseteq \mathbb{R}$ so that $\sum_{n=1}^\infty x_n$ is convergent. Then*

$$\frac{1}{b_n}\sum_{j=1}^n b_j x_j \underset{n}{\longrightarrow} 0.$$

Proof. We let $b_0 = 0$, $S_0 = 0$ and $S_n := x_1 + \cdots + x_n$. Then

$$\begin{aligned}
\frac{1}{b_n}\sum_{j=1}^n b_j x_j &= \frac{1}{b_n}\sum_{j=1}^n b_j (S_j - S_{j-1}) \\
&= \frac{1}{b_n}\left(b_n S_n - b_0 S_0 - \sum_{j=1}^n S_{j-1}(b_j - b_{j-1}) \right) \\
&= S_n - \frac{1}{b_n}\sum_{j=1}^n S_{j-1}(b_j - b_{j-1}) \underset{n}{\longrightarrow} x - x = 0.
\end{aligned}$$

□

Lemma 1.6.6. *Let $f : \Omega \to \mathbb{R}$ be a random variable. Then*

$$\sum_{n=1}^\infty \mathbb{P}(|f| \geq n) \leq \int_\Omega |f| d\mathbb{P} \leq 1 + \sum_{n=1}^\infty \mathbb{P}(|f| \geq n).$$

Proof.

$$\begin{aligned}
\sum_{n=1}^\infty \mathbb{P}(|f| \geq n) &= \sum_{n=1}^\infty \sum_{k \geq n} \mathbb{P}(k \leq |f| < k+1) \\
&= \sum_{k=1}^\infty k \mathbb{P}(k \leq |f| < k+1) \\
&\leq \mathbb{E}|f|.
\end{aligned}$$

The other inequality is proved in the same way. □

[2] Otto Töplitz, 01/08/1881 (Wroclaw, Poland) - 15/02/1940 (Jerusalem, Israel).

[3] Leopold Kronecker, 07/12/1823 (Legnica, Poland) - 29/12/1891 (Berlin, Germany).

Proof of Proposition 1.6.1. We can assume that $\mathbb{E}f_1 = 0$. The idea is to truncate the random variables and to apply the theorem of Kolmogorov-Chinčin 1.6.3. We let

$$\tilde{f}_n(\omega) := \begin{cases} f_n(\omega) & : \ |f_n(\omega)| < n \\ 0 & : \ |f_n(\omega)| \geq n \end{cases}$$

and $g_n(\omega) := \tilde{f}_n(\omega) - \mathbb{E}\tilde{f}_n$.
(i) Now we compute

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^2}\mathbb{E}g_n^2 &\leq \sum_{n=1}^{\infty} \frac{1}{n^2}\mathbb{E}\tilde{f}_n^2 \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2}\mathbb{E}\chi_{\{|f_n|<n\}} f_n^2 \\
&= \sum_{n=1}^{\infty}\sum_{k=1}^{n} \frac{1}{n^2}\mathbb{E}\chi_{\{k-1\leq|f_1|<k\}} f_1^2 \\
&= \sum_{k=1}^{\infty}\left(\sum_{n=k}^{\infty} \frac{1}{n^2}\right)\mathbb{E}\chi_{\{k-1\leq|f_1|<k\}} f_1^2 \\
&\leq 2\sum_{k=1}^{\infty} \frac{1}{k}\mathbb{E}\chi_{\{k-1\leq|f_1|<k\}} f_1^2 \\
&\leq 2\sum_{k=1}^{\infty} \mathbb{E}\chi_{\{k-1\leq|f_1|<k\}} |f_1| \\
&= 2\mathbb{E}|f_1| < \infty,
\end{aligned}$$

where we have used that

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{1}{k} + \int_k^{\infty} \frac{1}{x^2}dx = \frac{1}{k} + \left(-x^{-1}|_k^{\infty}\right) = \frac{2}{k}.$$

(ii) Applying Proposition 1.6.3 gives that $\sum_{n=1}^{\infty} \frac{1}{n} g_n$ is convergent almost surely. Applying Kronecker's Lemma gives that

$$\frac{1}{n}\left(1 \cdot \frac{1}{1}g_1 + 2 \cdot \frac{1}{2}g_2 + \ldots + n \cdot \frac{1}{n}g_n\right) = \frac{1}{n}\left(g_1 + \ldots + g_n\right) \xrightarrow[a.s.]{} 0.$$

(iii) To replace g by $\tilde{f}$ we need to show that $\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\tilde{f}_k \xrightarrow[n]{} 0$. According to the Töplitz-Lemma it is sufficient to prove that $\mathbb{E}\tilde{f}_n \xrightarrow[n]{} 0$. But this follows from

$$\mathbb{E}\tilde{f}_n = \int_{\{|f_n|<n\}} f_n d\mathbb{P} = \int_{\{|f_1|<n\}} f_1 d\mathbb{P} \xrightarrow[n]{} \int_{\Omega} f_1 d\mathbb{P} = 0$$

because of dominated convergence. So we get $\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\tilde{f}_k \xrightarrow[a.s.]{} 0$.

(iv) To replace $\tilde{f}$ by f we use Borel-Cantelli. We observe

$$\begin{aligned}
\mathbb{E}|f_1| < \infty &\iff \sum_{n=1}^{\infty} \mathbb{P}\left(|f_1| \geq n\right) < \infty \\
&\iff \sum_{n=1}^{\infty} \mathbb{P}\left(|f_n| \geq n\right) < \infty \\
&\iff \mathbb{P}\left(\{\omega : \#\{n : |f_n(\omega)| \geq n\} < \infty\}\right) = 1
\end{aligned}$$

and let $\Omega_0 := \{\omega : \#\{n : |f_n(\omega)| \geq n\} < \infty\}$. Hence, for all $\omega \in \Omega_0$ we get some $n(\omega)$ such that for $n \geq n(\omega)$ one has $f_n(\omega) = \tilde{f}_n(\omega)$ and

$$\lim_{n\to\infty} \frac{1}{n}\left(f_1(\omega) + \ldots + f_n(\omega)\right) = \lim_{n\to\infty} \frac{1}{n}\left(\tilde{f}_1(\omega) + \ldots + \tilde{f}_n(\omega)\right).$$

This completes the proof. □

Now we consider a converse to the Strong Law of Large Numbers.

Proposition 1.6.7. *Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and that $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ are independent random variables having the same distribution. If there is a constant $c \in \mathbb{R}$ such that*

$$\frac{1}{n}\left(f_1 + \ldots + f_n\right) \xrightarrow[n]{} c \ \mathbb{P}\text{-a.s.},$$

then $\mathbb{E}|f_1| < \infty$ and $\mathbb{E} f_1 = c$.

The proof will be subject to an exercise.

As an application we deduce a result of Borel. Let $t \in [0,1)$ and write

$$t = \sum_{n=1}^{\infty} \frac{1}{2^n} t_n, \ t_n \in \{0,1\},$$

where $\#\{n : t_n = 0\} = \infty$. Hence

$$\{t : t_1 = x_1, \ldots, t_n = x_n\} = \left\{t : \frac{x_1}{2} + \ldots + \frac{x_n}{2^n} \leq t < \frac{x_1}{2} + \ldots + \frac{x_n}{2^n} + \frac{1}{2^n}\right\}.$$

Let $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}([0,1))$, λ be the Lebesgue measure.

Proposition 1.6.8 (Borel)**.** *Given $t \in [0,1)$ we let*

$$Z_n(t) := \#\left\{1 \leq k \leq n : t_k = 1\right\}.$$

Then

$$\lambda\left(\left\{t \in [0,1) : \frac{1}{n} Z_n(t) \to \frac{1}{2}\right\}\right) = 1.$$

Proof. Letting $f_n(t) := t_n$ we simply have $\frac{1}{n}Z_n(t) = \frac{1}{n}(f_1(t) + \ldots + f_n(t))$. The random variables $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ satisfy

$$\lambda\left(f_1 = \theta_1, \ldots, f_n = \theta_N\right) = \frac{1}{2^N} = \lambda\left(f_1 = \theta_1\right) \cdots \lambda\left(f_n = \theta_N\right)$$

for all $\theta_1, \ldots, \theta_N \in \{0, 1\}$ and are therefore independent as one can replace the condition $\{f_k = \theta_k\}$ by $\{f \in B_k\}$ for $B_k \in \mathcal{B}(\mathbb{R})$. Hence we can apply the SLLN and are done. □

1.6.2 The law of iterated logarithm (LIL)

Assume that we have independent random variables $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ having the same distribution and mean zero. Define

$$S_n(\omega) := f_1(\omega) + \ldots + f_n(\omega), \; S_0 := 0.$$

We get a random walk. To get information for the random walk one is interested in functions $\varphi(n)$ and $\psi(n)$ such that

$$\limsup_n \frac{S_n}{\varphi(n)} = 1 \quad \text{a.s.}$$

and

$$\liminf_n \frac{S_n}{\psi(n)} = -1 \quad \text{a.s.},$$

so that for all $\epsilon > 0$ the area between $(1+\varepsilon)\varphi(n)$ and $-(1+\varepsilon)\psi(n)$ is left only finitely many times with probability one.

Proposition 1.6.9 (LIL, Chinčin, Kolmogorov, Hartman, Wintner)**.** *Assume that $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ are independent random variables, having the same distribution, such that $\mathbb{E}f_1 = 0$ and $\mathbb{E}f_1^2 = \sigma^2 > 0$. Let*

$$\psi(n) := \sqrt{2\sigma^2 n \log\log n}.$$

Then

$$\mathbb{P}\left(\limsup_n \frac{S_n}{\psi(n)} = 1\right) = \mathbb{P}\left(\liminf_n \frac{S_n}{\psi(n)} = -1\right) = 1.$$

For the proof of a special case we need two lemmata.

Lemma 1.6.10. *Assume that $f_1, \ldots, f_N : \Omega \to \mathbb{R}$ are independent and symmetric, that means*

$$\mathbb{P}\left(f_k \in B\right) = \mathbb{P}\left(-f_k \in B\right)$$

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

$$\mathbb{P}\left(\max_{1 \le n \le N} S_n > a\right) \le 2\mathbb{P}\left(S_N > a\right)$$

for $S_n := f_1 + \ldots + f_n$ and $a \in \mathbb{R}$.

Proof. We let

$$A_n := \{\omega : S_n(\omega) > a, S_1(\omega) \le a, \ldots, S_{n-1}(\omega) \le a\}$$

so that

$$\mathbb{P}\left(\max_{1\le n\le N} S_n > a\right) = \sum_{n=1}^{N} \mathbb{P}(A_n).$$

Moreover,

$$\begin{aligned}
\mathbb{P}(\{S_N > a\} \cap A_n) &\ge \mathbb{P}(\{S_N \ge S_n\} \cap A_n) \\
&= \mathbb{P}(\{S_N \ge S_n\})\,\mathbb{P}(A_n) \\
&= \frac{1}{2}\mathbb{P}(A_n),
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{P}\left(\max_{1\le n\le N} S_n > a\right) &\le 2\sum_{n=1}^{N} \mathbb{P}(\{S_N > a\} \cap A_n) \\
&\le 2\mathbb{P}(S_N > a).
\end{aligned}$$

□

Lemma 1.6.11. *Let $S_n \sim N(0,\sigma^2(n))$, $\sigma^2(n) \nearrow \infty$ and $\frac{a(n)}{\sigma(n)} \underset{n}{\longrightarrow} \infty$, where $a(n), \sigma(n) > 0$ and $n \ge 1$. Then*

$$\lim_{n\to\infty} \mathbb{P}(S_n > a(n)) \left(\frac{\sigma(n)}{\sqrt{2\pi}a(n)} e^{-\frac{a^2(n)}{2\sigma^2(n)}} \right)^{-1} = 1.$$

Proof. The assertion follows immediately from

$$\lim_{x\to\infty} \frac{\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{y^2}{2}}dy}{\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}} = \lim_{x\to\infty} \frac{-e^{-\frac{x^2}{2}}}{-x^{-2}e^{-\frac{x^2}{2}} - e^{-\frac{x^2}{2}}} = \lim_{x\to\infty} \frac{1}{\frac{1}{x^2}+1} = 1.$$

□

Proof of Proposition 1.6.9. We only indicate the proof for $f_n \sim N(0,1)$, $n = 1, 2, \ldots$ and show that

$$\mathbb{P}\left(\limsup_n \frac{S_n}{\psi(n)} \le 1\right) = 1.$$

Assume $\varepsilon > 0$ and a sequence $n_{k_0}^{(\varepsilon)} \leq n_{k_0+1}^{(\varepsilon)} \leq \ldots$ Define

$$A_k^\varepsilon := \left\{ \omega \in \Omega : S_n(\omega) > (1+\varepsilon)\psi(n) \quad \text{for some} \quad n \in (n_k^{(\varepsilon)}, n_{k+1}^{(\varepsilon)}] \right\}$$

and

$$A_\varepsilon := \limsup_{k \geq k_0} A_k^\varepsilon = \{ \omega \in \Omega : \omega \in A_k^\varepsilon \quad \text{for infinitely many} \quad k \geq k_0 \}.$$

Assume that we have $\mathbb{P}(A_\varepsilon) = 0$. Then

$$\# \{ n : S_n(\omega) > (1+\varepsilon)\psi(n) \} < \infty$$

for all $\omega \in A_\varepsilon$. Letting

$$\Omega_0 := \bigcap_{N=1}^{\infty} A_{\frac{1}{N}}^{\mathrm{c}}$$

gives a set of measure 1. Moreover,

$$\# \left\{ n : \frac{S_n(\omega)}{\psi(n)} > 1 + \frac{1}{N} \right\} < \infty \quad \text{for all} \quad \omega \in \Omega_0 \text{ and } N = 1, 2, \ldots .$$

Hence

$$\limsup_n \frac{S_n(\omega)}{\psi(n)} \leq 1 + \frac{1}{N} \quad \text{for all} \quad \omega \in \Omega_0 \text{ and } N = 1, 2, \ldots$$

and

$$\limsup_n \frac{S_n(\omega)}{\psi(n)} \leq 1 \quad \text{for all} \quad \omega \in \Omega_0.$$

So it remains to show that $\mathbb{P}(A_\varepsilon) = 0$. We take $n_k^{(\varepsilon)} = \lambda^k = (1+\varepsilon)^k$ and some (large) $k_0 \geq 1$. According to Borel-Cantelli it is sufficient to prove that $\sum_{k=k_0}^{\infty} \mathbb{P}(A_k^\varepsilon) < \infty$. We get by Lemma 1.6.10 and 1.6.11 that

$$\begin{aligned}
\mathbb{P}(A_k^\varepsilon) &= \mathbb{P}\left(S_n > \lambda\psi(n) \quad \text{for some} \quad n \in (n_k^{(\varepsilon)}, n_{k+1}^{(\varepsilon)}] \right) \\
&\leq \mathbb{P}\left(S_n > \lambda\psi\left(n_k^{(\varepsilon)}\right) \quad \text{for some} \quad n \in (n_k^{(\varepsilon)}, n_{k+1}^{(\varepsilon)}] \right) \\
&\leq \mathbb{P}\left(S_n > \lambda\psi\left(n_k^{(\varepsilon)}\right) \quad \text{for some} \quad n \leq n_{k+1}^{(\varepsilon)} \right) \\
&\leq 2\mathbb{P}\left(S_{\left[n_{k+1}^{(\varepsilon)}\right]} > \lambda\psi\left(n_k^{(\varepsilon)}\right) \right) \\
&\leq 2c_0 \frac{\sqrt{\left[n_{k+1}^{(\varepsilon)}\right]}}{\sqrt{2\pi}\lambda\psi(n_k^{(\varepsilon)})} e^{-\frac{\lambda^2 \psi(n_k^{(\varepsilon)})^2}{2\left[n_{k+1}^{(\varepsilon)}\right]}} \\
&= 2c_0 \frac{\lambda^{\frac{k+1}{2}}}{\sqrt{2\pi}\lambda\sqrt{2\lambda^k \log\log \lambda^k}} e^{-\frac{\lambda^2}{\lambda^{k+1}}\lambda^k \log\log(\lambda^k)}
\end{aligned}$$

$$
\begin{aligned}
&\leq c_1 e^{-\lambda \log\log(\lambda^k)} \\
&= c_1 e^{-\lambda \log(k \log \lambda)} \\
&= c_1 e^{-\lambda \log k}\ e^{-\lambda \log\log \lambda} \\
&= c_2 k^{-\lambda}
\end{aligned}
$$

so that

$$\sum_{k \geq k_0} \mathbb{P}\left(A_k^{\varepsilon}\right) \leq c \sum_{k \geq k_0} k^{-\lambda} < \infty.$$

$\square$

Chapter 2

Characteristic functions

2.1 Complex numbers

We identify

$$\mathbb{C} \cong \mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

and write

$$\mathbb{C} \ni z = x + iy \cong (x, y) \in \mathbb{R}^2,$$

where $x = Re(z)$ is the real part of z and $y = Im(z)$ is the imaginary part of z. The complex numbers are an extension of the real numbers by using $\mathbb{R}$ as $\mathbb{C}$ with $x \mapsto (x, 0) = x + i \cdot 0$. We recall some definitions:

Addition. For $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ we let

$$z_1 + z_2 := (x_1 + x_2, y_1 + y_2) = (x_1 + x_2) + i(y_1 + y_2).$$

Multiplication. For $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ we let

$$z_1 z_2 := (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1).$$

Remark 2.1.1. (i) If we interpret $i^2 = -1$, we get this formally by

$$(x_1 + iy_1)(x_2 + iy_2) = x_1 x_2 + i^2 y_1 y_2 + i(x_1 y_2 + x_2 y_1).$$

(ii) If $z_1 = (x_1, 0)$, then $z_1 z_2 = (x_1 x_2, x_1 y_2) = x_1(x_2, y_2)$ and, in the same way, if $z_2 = (x_2, 0)$, then $z_1 z_2 = x_2(x_1, y_1)$.

Length of a complex number: If $z = (x, y)$, then $|z| = \sqrt{x^2 + y^2}$.

Conjugate complex number: If $z = (x, y)$, then $\bar{z} := (x, -y) = x - iy$. We have that $z\bar{z} = x^2 + y^2 = |x|^2$.

Polar coordinates: These coordinates are given as (r, φ) where $r \geq 0$ and $\varphi \in [0, 2\pi)$ are determined by

$$\begin{aligned} x &= r\cos\varphi, \\ y &= r\sin\varphi. \end{aligned}$$

Note that $r = |z|$ and that φ is not unique whenever $r = 0$.

Now we recall the notion of the complex **exponential function**.

Definition 2.1.2. *For $z \in \mathbb{C}$ we let*

$$e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!},$$

where the convergence is considered with respect to the euclidean metric in $\mathbb{R}^2$.

Proposition 2.1.3. (i) *For all $z_1, z_2 \in \mathbb{C}$ one has $e^{z_1+z_2} = e^{z_1}e^{z_2}$.*

(ii) *One has $e^{ix} = \cos x + i \sin x$ for $x \in \mathbb{R}$ (Euler's [1] formula).*

Proof. (i) is an exercise. (ii) follows from

$$e^{ix} = \frac{x^0}{0!} - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + i\left(\frac{x^1}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\right) = \cos x + i \sin x.$$

□

Complex valued random variables.

Definition 2.1.4. Let $(\Omega, \mathcal{F})$ be a measurable space.

(i) A map $f : \Omega \to \mathbb{C}$ is called *measurable*, or *random variable* in case $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, provided that $f : \Omega \to \mathbb{R}^2$ is a random variable, i.e. for $f = (f_1, f_2)$ the maps $f_1, f_2 : \Omega \to \mathbb{R}$ are measurable.

(ii) A random variable $f : \Omega \to \mathbb{C}$ is called *integrable* provided that

$$\int_\Omega |f(\omega)| d\mathbb{P}(\omega) < \infty.$$

In this case we let

$$\int_\Omega f(\omega) d\mathbb{P}(\omega) = \int_\Omega Re\,(f(\omega))\, d\mathbb{P}(\omega) + i \int_\Omega Im\,(f(\omega))\, d\mathbb{P}(\omega).$$

[1] Leonhard Euler 15/04/1707 (Basel, Switzerland) - 18/09/1783 (St Petersburg, Russia), Swiss mathematician.

2.2 Definition and basic properties of characteristic functions

The concept of characteristic functions is one of the most important tools in probability theory, and it also gives a link to harmonic analysis. The idea is to describe properties of random variables $f : \Omega \to \mathbb{R}$ or of measures on $\mathcal{B}(\mathbb{R}^d)$ by a Fourier transform.

Definition 2.2.1. Given $d \in \{1, 2, \ldots\}$, we let $\mathcal{M}_1^+(\mathbb{R}^d)$ be the set of all probability measures μ on $\mathcal{B}(\mathbb{R}^d)$.

The number '1' stands for $\mu(\mathbb{R}^d) = 1$ and '+' for $\mu(A) \geq 0$.

Definition 2.2.2. (i) Let $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then

$$\widehat{\mu}(x) := \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y), \ x \in \mathbb{R}^d,$$

is called *Fourier transform* of μ.

(ii) Let $f : \Omega \to \mathbb{R}^d$ be a random variable. Then

$$\widehat{f}(x) := \mathbb{E}e^{i\langle x,f\rangle} = \int_\Omega e^{i\langle x,f(\omega)\rangle} d\mathbb{P}(\omega)$$

is called *characteristic function* of f.

Remark 2.2.3. (i) $\widehat{\mu}$ and $\hat{f}$ exist, since $\left|e^{i\langle x,y\rangle}\right| = \left|e^{i\langle x,f\rangle}\right| = 1$ and $y \mapsto e^{i\langle x,y\rangle}$ is continuous, so that $y \mapsto e^{i\langle x,y\rangle}$ and $\omega \mapsto e^{i\langle x,f(\omega)\rangle}$ are measurable.

(ii) If μ_f is the law of $f : \Omega \to \mathbb{R}^d$, then $\widehat{f}(x) = \widehat{\mu}_f(x)$ for $x \in \mathbb{R}^d$. In fact, this follows from the change of variable formula where we get, for $\psi(y) = e^{i\langle x,y\rangle}$,

$$\int_\Omega e^{i\langle x,f(\omega)\rangle} d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} \psi(y) d\mu_f(y) = \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu_f(y).$$

Example 2.2.4. (a) Let $a \in \mathbb{R}^d$ and δ_a be the Dirac-measure with

$$\delta_a(B) = \begin{cases} 1, & a \in B \\ 0, & a \notin B. \end{cases}$$

Then

$$\widehat{\delta_a}(x) = \int_{\mathbb{R}} e^{i\langle x,y\rangle} d\delta_a(y) = e^{i\langle x,a\rangle}.$$

(b) Let $a_1, \ldots, a_n \in \mathbb{R}^d$, $0 \le \theta_i \le 1$, $\sum_{i=1}^n \theta_i = 1$ and $\mu = \sum_{i=1}^n \theta_i \delta_{a_i}$. Then

$$\widehat{\mu}(x) = \sum_{i=1}^n \theta_i e^{i\langle x, a_i \rangle} = \sum_{i=1}^n \theta_i \left(\cos\langle x, a_i\rangle + i \sin\langle x, a_i \rangle\right),$$

which is a trigonometric polynomial.

(c) Binomial distribution: Let $0 < p < 1$, $d = 1$, $n \in \{1, 2, \ldots\}$ and

$$\mu(\{k\}) := \binom{n}{k} p^{n-k}(1-p)^k, \quad \text{for} \quad k = 0, \ldots, n.$$

Then

$$\begin{aligned}
\widehat{\mu}(x) &= \int_{\mathbb{R}} e^{ixy} d\mu(y) \\
&= \sum_{k=0}^n \binom{n}{k} p^{n-k}(1-p)^k e^{ixk} \\
&= \sum_{k=0}^n \binom{n}{k} p^{n-k}\left((1-p)e^{ix}\right)^k \\
&= \left(p + (1-p)e^{ix}\right)^n.
\end{aligned}$$

Proposition 2.2.5. *Let $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then the following is true:*

(i) *The function $\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is uniformly continuous, that means that for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|\widehat{\mu}(x) - \widehat{\mu}(y)| \le \varepsilon$ whenever $|x - y| = \left(\sum_{i=1}^d |x_i - y_i|^2\right)^{\frac{1}{2}} \le \delta$.*

(ii) *For all $x \in \mathbb{R}^d$ one has $|\widehat{\mu}(x)| \le \widehat{\mu}(0) = 1$.*

(iii) *The function $\widehat{\mu}$ is positive semi-definite, that means that for all $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ it follows that*

$$\sum_{k,l=1}^n \lambda_k \bar{\lambda}_l \widehat{\mu}(x_k - x_l) \ge 0.$$

Proof. (ii) This part follows from

$$\begin{aligned}
\left| \int_{\mathbb{R}^d} e^{i\langle x, y\rangle} d\mu(y) \right| &= \left| \int_{\mathbb{R}^d} \cos\langle x, y\rangle\, d\mu(y) + i \int_{\mathbb{R}^d} \sin\langle x, y \rangle\, d\mu(y) \right| \\
&= \left(\left| \int_{\mathbb{R}^d} \cos\langle x, y\rangle\, d\mu(y) \right|^2 + \left| \int_{\mathbb{R}^d} \sin\langle x, y\rangle\, d\mu(y) \right|^2 \right)^{\frac{1}{2}}
\end{aligned}$$

$$
\begin{aligned}
&\le \left(\int_{\mathbb{R}^d} |\cos\langle x, y\rangle|^2 \, d\mu(y) + \int_{\mathbb{R}^d} |\sin\langle x, y\rangle|^2 \, d\mu(y) \right)^{\frac{1}{2}} \\
&= 1,
\end{aligned}
$$

where we have used Hölder's inequality.
(iii) Here we have that

$$
\begin{aligned}
\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \widehat{\mu}(x_k - x_l) &= \sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \int_{\mathbb{R}^d} e^{i\langle x_k - x_l, y\rangle} d\mu(y) \\
&= \sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \int_{\mathbb{R}^d} e^{i\langle x_k, y\rangle} e^{-i\langle x_l, y\rangle} d\mu(y).
\end{aligned}
$$

Since $\overline{e^{i\alpha}} = \overline{\cos\alpha + i\sin\alpha} = \cos\alpha - i\sin\alpha = \cos(-\alpha) + i\sin(-\alpha) = e^{-i\alpha}$, we can continue to

$$
\begin{aligned}
\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \widehat{\mu}(x_k - x_l) &= \int_{\mathbb{R}^d} \left[\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l e^{i\langle x_k, y\rangle} \overline{e^{-i\langle x_l, y\rangle}} d\mu(y) \right] \\
&= \int_{\mathbb{R}^d} \left[\sum_{k=1}^{n} \lambda_k e^{i\langle x_k, y\rangle} \right] \left[\sum_{l=1}^{n} \bar{\lambda}_l \overline{e^{-i\langle x_l, y\rangle}} \right] d\mu(y) \\
&= \int_{\mathbb{R}^d} \left| \sum_{k=1}^{n} \lambda_k e^{i\langle x_k, y\rangle} \right|^2 d\mu(y) \ge 0.
\end{aligned}
$$

(i) Let $\varepsilon > 0$. Choose a ball of radius $R > 0$ such that $\mu(\mathbb{R}^d \setminus B_R(0)) < \frac{\varepsilon}{3}$, where

$$
B_R(0) := \left\{ x \in \mathbb{R}^d : \left(\sum_{i=1}^{d} |x_i|^2 \right)^{\frac{1}{2}} \le R \right\}.
$$

Take $\delta := \frac{\varepsilon}{3R}$. Then, since $\left| e^{i\alpha} - e^{i\beta} \right| \le |\alpha - \beta|$ and by Hölder's inequality, if $\|x_1 - x_2\|_2 \le \delta$

$$
\begin{aligned}
&|\widehat{\mu}(x_1) - \widehat{\mu}(x_2)| \\
\le\ & \int_{B_R(0)} \left| e^{i\langle x_1, y\rangle} - e^{i\langle x_2, y\rangle} \right| d\mu(y) + \int_{\mathbb{R}^d \setminus B_R(0)} \left| e^{i\langle x_1, y\rangle} - e^{i\langle x_2, y\rangle} \right| d\mu(y) \\
\le\ & \int_{B_R(0)} |\langle x_1 - x_2, y\rangle| \, d\mu(y) + \int_{\mathbb{R}^d \setminus B_R(0)} 2 d\mu(y) \\
\le\ & \|x_1 - x_2\|_2 \int_{B_R(0)} \|y\|_2 d\mu(y) + 2\frac{\varepsilon}{3} \\
\le\ & \frac{\varepsilon}{3R} R + 2\frac{\varepsilon}{3} = \varepsilon.
\end{aligned}
$$

$\square$

Now we connect our Fourier transform to the Fourier transform for functions $\psi : \mathbb{R}^d \to \mathbb{C}$.

Definition 2.2.6. For a Borel-measurable $\psi : \mathbb{R}^d \to \mathbb{C}$ such that

$$\int_{\mathbb{R}^d} |\psi(y)| d\lambda(y) < \infty$$

we let

$$\hat{\psi}(x) := \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \psi(y) d\lambda(y)$$

be the Fourier transform of φ, where $x \in \mathbb{R}^d$ and λ is the Lebesgue measure on $\mathbb{R}^d$.

Before we give the connection to our previous definition of a Fourier transform we extend Definition 2.2.2 to more general measures.

Definition 2.2.7. Let $(\Omega, \mathcal{F})$ be a measurable space. A map $\mu : \mathcal{F} \to \mathbb{R}$ is called *finite signed measure*, provided that $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are finite measures $\mu^+, \mu^- : \mathcal{F} \to [0, \infty)$. The collection of finite signed measures is denoted by $\mathcal{M}(\Omega, \mathcal{F})$. In case of $\mathcal{M}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ we simply write $\mathcal{M}(\mathbb{R}^d)$.

Definition 2.2.8. For $\mu \in \mathcal{M}(\mathbb{R}^d)$ we let

$$\widehat{\mu} := \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^+(y) - \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^-(y).$$

The expression does not depend on the decomposition $\mu = \mu^+ - \mu^-$. Let us now connect the different definitions.

Proposition 2.2.9. *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be Borel-measurable such that $\int_{\mathbb{R}^d} |\varphi(x)| d\lambda(x) < \infty$. Define*

$$\mu(B) := \int_{\mathbb{R}^d} \chi_B(x) \varphi(x) d\lambda(x).$$

Then one has the following:

(i) $\mu \in \mathcal{M}(\mathbb{R}^d)$ *with* $\mu = \mu^+ - \mu^-$, *where*

$$\mu^\pm(B) := \int_{\mathbb{R}^d} \chi_B(x) \varphi^\pm(x) d\lambda(x),$$

$\varphi^+ := \max\{\varphi, 0\}$ *and* $\varphi^- := \max\{-\varphi, 0\}$.

(ii) $\widehat{\varphi}(x) = \widehat{\mu}(x)$ *for all* $x \in \mathbb{R}^d$.

Proof. Assertion (i) is obvious, (ii) follows from

$$\begin{aligned} \hat{\mu}(x) &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^+(y) - \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu^-(y) \\ &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \varphi^+(y) d\lambda(y) - \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \varphi^-(y) d\lambda(y) \\ &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \varphi(y) d\lambda(y). \end{aligned}$$

□

2.3 Convolutions

Often one has the problem to compute the distribution of $f_1 + \cdots + f_n$ where $f_1, ..., f_n : \Omega \to \mathbb{R}^d$ are random variables. We already considered some problems in this direction in the SLLN and LIL. Now we introduce a technique especially designed for this purpose: the convolution.

Definition 2.3.1 (Convolution of measures)**.** *Let* $\mu_1, \ldots, \mu_n \in \mathcal{M}_1^+(\mathbb{R}^d)$. *Then* $\mu_1 * \cdots * \mu_n \in \mathcal{M}_1^+(\mathbb{R}^d)$ *is the law of* $f : \Omega \to \mathbb{R}^d$ *defined by*

(i) $(\Omega, \mathcal{F}, \mathbb{P}) := \big(\mathbb{R}^d \times \cdots \times \mathbb{R}^d,\ \mathcal{B}(\mathbb{R}^d) \otimes \cdots \otimes \mathcal{B}(\mathbb{R}^d),\ \mu_1 \times \cdots \times \mu_n\big)$,

(ii) $f(x_1, \ldots, x_n) := x_1 + \cdots + x_n$,

that is

$$(\mu_1 * \cdots * \mu_n)(B) = (\mu_1 \times \cdots \times \mu_n)\left(\{(x_1, \ldots, x_n) : x_1 + \cdots + x_n \in B\}\right).$$

The measure $\mu_1 * \cdots * \mu_n$ *is called* convolution of $\mu_1, \ldots, \mu_n$.

Remark 2.3.2. One has $\mathbb{R}^d \times \cdots \times \mathbb{R}^d = \mathbb{R}^{nd}$ and $\mathcal{B}(\mathbb{R}^d) \otimes \cdots \otimes \mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}^{nd})$.

Example 2.3.3. Let $a \in \mathbb{R}^d$ and

$$\delta_a(B) := \begin{cases} 1, & a \in B \\ 0, & a \notin B \end{cases}$$

be the Dirac-measure $\delta_a \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then $\delta_{a_1} * \cdots * \delta_{a_n} = \delta_{a_1 + \cdots + a_n}$.

Proof. By definition,

$$\begin{aligned}
& (\delta_{a_1} * \cdots * \delta_{a_n})(B) \\
= \; & (\delta_{a_1} \times \cdots \times \delta_{a_n}) \left(\{(x_1, \ldots, x_n) : x_1 + \cdots + x_n \in B\}\right) \\
= \; & \delta_{(a_1, \ldots, a_n)} \left(\{(x_1, \ldots, x_n) : x_1 + \cdots + x_n \in B\}\right) \\
= \; & \begin{cases} 1, & a_1 + \cdots + a_n \in B \\ 0, & a_1 + \cdots + a_n \notin B. \end{cases}
\end{aligned}$$

□

Example 2.3.4. $\delta_0 * \mu = \mu * \delta_0 = \mu$ for all $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$, that means that δ_0 is a unit with respect to the convolution.

Proof. By Fubini's theorem

$$\begin{aligned}
(\delta_0 * \mu)(B) &= (\delta_0 \times \mu) \left(\{(x_1, x_2) : x_1 + x_2 \in B\}\right) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \chi_B(x_1 + x_2) d\delta_0(x_1) \right] d\mu(x_2) \\
&= \int_{\mathbb{R}^d} \chi_B(x_2) d\mu(x_2) = \mu(B).
\end{aligned}$$

For the other direction one can use Proposition 2.3.5 below. □

Proposition 2.3.5. *For $\mu_1, \mu_2, \mu_3 \in \mathcal{M}_1^+(\mathbb{R}^d)$ one has*

(i) $\mu_1 * \mu_2 = \mu_2 * \mu_1$,

(ii) $\mu_1 * (\mu_2 * \mu_3) = (\mu_1 * \mu_2) * \mu_3 = \mu_1 * \mu_2 * \mu_3$.

Proof. Let $f_1, f_2, f_3 : \Omega \to \mathbb{R}$ be independent random variables such that $law(f_j) = \mu_j$. Now (i) follows from the fact that $f_1 + f_2$ and $f_2 + f_1$ have the same distribution, and (ii) follows from $f_1 + (f_2 + f_3) = (f_1 + f_2) + f_3 = f_1 + f_2 + f_3$. □

As before, we consider the convolution for functions as well. What is a good candidate for this? Assume $\varphi_1, \varphi_2 : \mathbb{R} \to [0, \infty)$ to be continuous and such that $\int_{\mathbb{R}} \varphi_1(x)dx = \int_{\mathbb{R}} \varphi_2(x)dx = 1$ and define $\mu_j(B) := \int_B \varphi_j(x)dx$, $B \in \mathcal{B}(\mathbb{R})$, for $j = 1, 2$ so that $\mu_1, \mu_2 \in \mathcal{M}_1^+(\mathbb{R})$. Now let us formally compute

$$(\mu_1 * \mu_2)(B) = (\mu_1 \times \mu_2)(\{(x_1, x_2) : x_1 + x_2 \in B\})$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi_B(x_1+x_2)\varphi_1(x_1)d\lambda(x_1) \right] \varphi_2(x_2)d\lambda(x_2) \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi_B(x)\varphi_1(x-x_2)d\lambda(x) \right] \varphi_2(x_2)d\lambda(x_2) \\
&= \int_{\mathbb{R}} \chi_B(x) \left[\int_{\mathbb{R}} \varphi_1(x-x_2)\varphi_2(x_2)d\lambda(x_2) \right] d\lambda(x)
\end{aligned}
$$

by Fubini's theorem. Hence $(\varphi_1 * \varphi_2)(x) := \int_{\mathbb{R}} \varphi_1(x-y)\varphi_2(y)d\lambda(y)$ seems to be a good candidate.

Definition 2.3.6 (Convolutions of functions)**.** (i) Let $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$ be Borel-functions such that $f_1(x) \geq 0$ and $f_2(x) \geq 0$ for all $x \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} f_1(x)d\lambda_d(x) < \infty$ and $\int_{\mathbb{R}^d} f_2(x)d\lambda_d(x) < \infty$. Then

$$
(f_1 * f_2)(x) := \int_{\mathbb{R}^d} f_1(x-y)f_2(y)d\lambda_d(y).
$$

The function $f_1 * f_2 : \mathbb{R}^d \to \mathbb{R}$ is called *convolution* of f_1 and f_2.

(ii) For Borel-functions $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}} |f_1(x)|d\lambda_d(x) < \infty$ and $\int_{\mathbb{R}} |f_2(x)|d\lambda_d(x) < \infty$ we let

$$
(f_1 * f_2)(x) = (f_1^+ * f_2^+)(x) - (f_1^+ * f_2^-)(x) - (f_1^- * f_2^+)(x) + (f_1^- * f_2^-)(x)
$$

if all terms on the right-hand side are finite.

We need to justify the definition above:

Proposition 2.3.7. *Let* $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$ *be as in* Definition 2.3.6(i)*. Then one has the following:*

(i) $f_1 * f_2 : \mathbb{R}^d \to [0, \infty]$ *is an extended measurable function.*

(ii) $\lambda_d(\{x \in \mathbb{R}^d : (f_1 * f_2)(x) = \infty\}) = 0$.

(iii) $\int_{\mathbb{R}^d} (f_1 * f_2)(x)d\lambda_d(x) = \int_{\mathbb{R}^d} f_1(x)d\lambda_d(x) \int_{\mathbb{R}^d} f_2(x)d\lambda_d(x)$.

Proof. Consider the map $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$
g(x,y) := f_1(x-y)f_2(y).
$$

We get a non-negative measurable function $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and can apply Fubini's theorem. We observe that

$$
\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x,y)d\lambda_d(x)d\lambda_d(y) = \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} f_1(x-y)f_2(y)d\lambda_d(x) \right] d\lambda_d(y)
$$

$$\begin{aligned} &= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} f_1(x-y) d\lambda_d(x) \right] f_2(y) d\lambda_d(y) \\ &= \int_{\mathbb{R}^d} f_2(y) d\lambda_d(y) \int_{\mathbb{R}^d} f_1(x-y) d\lambda_d(x) \\ &< \infty. \end{aligned}$$

implies (iii). Moreover, (i) and (ii) follow as a byproduct from Fubini's theorem. □

Remark 2.3.8. The Lebesgue measure of those $x \in \mathbb{R}^d$ for which we cannot define $(f_1 * f_2)(x)$ in Definition 2.3.6(ii) is zero so that we (can) agree about $(f_1 * f_2)(x) = 0$ for those x.

Now we connect the two convolutions to each other.

Proposition 2.3.9. *Let $\mu_1, \mu_2 \in \mathcal{M}_1^+(\mathbb{R}^d)$ with*

$$\mu_i(B) := \int_{\mathbb{R}^d} \chi_B(x) p_i(x) d\lambda_d(x), \ i = 1, 2,$$

for Borel-measurable and non-negative $p_1, p_2 : \mathbb{R}^d \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^d} p_1(x) d\lambda_d(x) = \int_{\mathbb{R}^d} p_2(x) d\lambda_d(x) = 1.$$

Then

$$(\mu_1 * \mu_2)(B) = \int_{\mathbb{R}^d} (p_1 * p_2)(x) \chi_B(x) d\lambda_d(x).$$

Exercise 2.3.10. *Compute $\mu_1 * \mu_2$, where μ_i is the uniform distribution on $[a_i, a_i + 1] \subseteq \mathbb{R}$ for $i = 1, 2$.*

2.4 Some important properties

Proposition 2.4.1. *For $\mu, \nu \in \mathcal{M}_1^+(\mathbb{R}^d)$ one has the following:*

(i) $\widehat{\mu} + \widehat{\nu} = \widehat{\mu + \nu}$ *where* $(\mu + \nu)(B) := \mu(B) + \nu(B)$.

(ii) $\widehat{\mu * \nu} = \widehat{\mu}\widehat{\nu}$.

(iii) *If* $A = (a_{ij})_{i,j=1}^d : \mathbb{R}^d \to \mathbb{R}^d$ *is a linear transformation, then*

$$\widehat{A(\mu)}(x) = \widehat{\mu}(A^{\mathrm{T}} x),$$

where $A(\mu)(B) := \mu(x \in \mathbb{R}^d : Ax \in B)$.

(iv) *If* $S_a : \mathbb{R}^d \to \mathbb{R}^d$ *is the shift operator* $S_a x := x + a$, $a \in \mathbb{R}^d$, *then*

$$\widehat{S_a(\mu)} = \widehat{\delta_a}\widehat{\mu}.$$

Proof. (i) is clear, (ii) follows from

$$\begin{aligned} \widehat{\mu * \nu}(x) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{i\langle x, y+z\rangle} d\mu(y) d\nu(z) \\ &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y) \int_{\mathbb{R}^d} e^{i\langle x,z\rangle} d\nu(z) \\ &= \widehat{\mu}(x)\widehat{\nu}(x). \end{aligned}$$

(iii) is an exercise and

$$\widehat{S_a(\mu)}(x) = \int_{\mathbb{R}^d} e^{i\langle x,y+a\rangle} d\mu(y) = e^{i\langle x,a\rangle}\widehat{\mu}(x)$$

implies (iv). □

Definition 2.4.2. A system Π of subsets $A \in \Omega$ is called π-system, provided that $A \cap B \in \Pi$ for all $A, B \in \Pi$.

Proposition 2.4.3. *Let* $(\Omega, \mathcal{F}, \mathbb{P}_1)$ *and* $(\Omega, \mathcal{F}, \mathbb{P}_2)$ *be probability spaces such that* $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ *for all* $A \in \Pi$ *where* Π *is a* π*-system which generates* $\mathcal{F}$. *Then* $\mathbb{P}_1 = \mathbb{P}_2$.

Examples 2.4.4. (a) If $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are measurable spaces, then

$$\Pi := \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$$

is a π-system.

(b) Assume a metric space M with the Borel-σ-algebra generated by the open sets. Then the system of open sets of a metric space is a π-system that generates the Borel σ-algebra. Another π-system that generates the Borel σ-algebra is the system of closed sets.

Proposition 2.4.5. *Assume that* $f_1, \ldots, f_n : \Omega \to \mathbb{R}^d$ *are independent random variables and that* $\mu_1, \ldots, \mu_n \in \mathcal{M}_1^+(\mathbb{R}^d)$ *are the laws of* $f_1, \ldots, f_n$, *that means that*

$$\mathbb{P}(f_k \in B) = \mu_k(B) \quad \text{for all} \quad B \in \mathcal{B}(\mathbb{R}^d).$$

Then $\mu_1 * \cdots * \mu_n$ *is the law of* $S := f_1 + \cdots + f_n$.

Proof. Define the product space $\overset{n}{\underset{1}{\times}}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_k) =: (M, \Sigma, Q)$ and consider the random vector $F : \Omega \to \mathbb{R}^{nd}$ given by $F(\omega) := (f_1(\omega), \ldots, f_n(\omega))$. Let μ be the law of F. Then, by independence,

$$\begin{aligned}
\mu(B_1 \times \cdots \times B_n) &= \mathbb{P}(f_1 \in B_1, \ldots, f_n \in B_n) \\
&= \prod_{k=1}^{n} \mathbb{P}(f_k \in B_k) \\
&= \mu_1(B_1) \cdots \mu_n(B_n) \\
&= Q(B_1 \times \cdots \times B_n).
\end{aligned}$$

Since the system $\Pi := \{B_1 \times \cdots \times B_n : B_k \in \mathcal{B}(\mathbb{R}^d)\}$ is a generating π-system, we get that $\mu = Q$. Consequently,

$$\begin{aligned}
\mathbb{P}(S \in B) &= \mu(\{(x_1, \ldots x_n) : x_1 + \cdots + x_n \in B\}) \\
&= Q(\{(x_1, \ldots x_n) : x_1 + \cdots + x_n \in B\}) \\
&= (\mu_1 * \cdots * \mu_n)(B).
\end{aligned}$$

□

A hopefully motivating example.

Example 2.4.6. Let $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be independent random variables having the same distribution. Let

$$S_n := \frac{1}{\sqrt{n}}(f_1 + \cdots + f_n).$$

We are interested in the convergence of S_n and compute their characteristic functions. Here we get that

$$\begin{aligned}
\widehat{S_n}(t) &= \left(\widehat{\frac{f_1}{\sqrt{n}} + \cdots + \frac{f_n}{\sqrt{n}}}\right)(t) = \left(\widehat{\mu_{\frac{f_1}{\sqrt{n}}} * \cdots * \mu_{\frac{f_n}{\sqrt{n}}}}\right)(t) \\
&= \left(\widehat{\mu_{\frac{f_1}{\sqrt{n}}}}(t)\right)^n = \left(\mathbb{E}e^{i\frac{f_1}{\sqrt{n}}t}\right)^n = \left(\widehat{f_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n,
\end{aligned}$$

where $\mu_{\frac{f_k}{\sqrt{n}}}$ is the law of $\frac{f_k}{\sqrt{n}}$.

Now we ask:

(Q1) Under what conditions does $(\widehat{f_1}(\frac{t}{\sqrt{n}}))^n$ converge to a function φ? If yes, is the limit φ a characteristic function, i.e. does there exist a probability measure μ such that $\widehat{\mu} = \varphi$.

(Q2) And finally, if there is a measure μ, what is its connection to the distributions of S_n?

The second question will be answered now.

Proposition 2.4.7 (Uniqueness). *Let $\mu, \nu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then $\mu = \nu$, if and only if $\widehat{\mu} = \widehat{\nu}$.*

Proof. It is clear that if $\mu = \nu$, then $\widehat{\mu} = \widehat{\nu}$. Let us show the other direction and assume $f \in L_1(\mathbb{R}^d; \mathbb{C})$. By Fubini's theorem we get

$$
\begin{aligned}
\int_{\mathbb{R}^d} \widehat{\mu}(y) f(y) d\lambda_d(y) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} f(y) d\mu(x) d\lambda_d(y) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} e^{i\langle x,y\rangle} f(y) d\lambda_d(y) \right] d\mu(x) \\
&= \int_{\mathbb{R}^d} \widehat{f}(x) d\mu(x).
\end{aligned}
$$

Hence $\widehat{\mu} = \widehat{\nu}$ implies that $\int_{\mathbb{R}^d} \widehat{f}(x) d\mu(x) = \int_{\mathbb{R}^d} \widehat{f}(x) d\nu(x)$ for all $f \in L_1(\mathbb{R}^d; \mathbb{C})$. Now we need a little more background and interrupt the proof.

Definition 2.4.8. Let

$$
C_0(\mathbb{R}^d; \mathbb{C}) := \left\{ g : \mathbb{R}^d \to \mathbb{C} \text{ continuous and } \lim_{|x| \to \infty} |g(x)| = 0 \right\}
$$

and

$$
\|g\|_{C_0} := \sup_{x \in \mathbb{R}^d} |g(x)|.
$$

Proposition 2.4.9 (Riemann & Lebesgue). *For all $f \in L_1(\mathbb{R}^d; \mathbb{C})$ one has $\widehat{f} \in C_0(\mathbb{R}^d; \mathbb{C})$.*

Proof. First of all we remark that by decomposing f into the positive and negative parts of the real and imaginary part (so that we have four parts) and applying Proposition 2.2.5 we get that $\widehat{f}$ is continuous. Next we recall that

$$
\begin{aligned}
\left| \widehat{f}(x) - \widehat{g}(x) \right| &= \left| \int_{\mathbb{R}^d} [f(y) - g(y)] \, e^{ixy} d\lambda_d(y) \right| \\
&\leq \int_{\mathbb{R}^d} |f(y) - g(y)| \, d\lambda_d(y) \\
&= \|f - g\|_{L_1}.
\end{aligned}
$$

Assume that $E \subseteq L_1(\mathbb{R}^d; \mathbb{C})$ is a dense subset such that $\widehat{f_0} \in C_0(\mathbb{R}^d; \mathbb{C})$ for $f_0 \in E$. Letting $f \in L_1(\mathbb{R}^d; \mathbb{C})$ we find $f_n \in E$ such that $\lim_n \|f_n - f\|_{L_1} = 0$

so that $\lim_n \sup_{x\in\mathbb{R}^d} |\widehat{f}_n(x) - \widehat{f}(x)| = 0$. Given $\varepsilon > 0$ we find an $n \geq 1$ such that $\|\widehat{f}_n - \widehat{f}\|_{C_0} < \varepsilon$ so that

$$\overline{\lim_{x\to\infty}} \left|\widehat{f}(x)\right| = \overline{\lim_{x\to\infty}} \left|\widehat{f}(x) - \widehat{f}_n(x)\right| \leq \varepsilon.$$

Since this holds for all $\varepsilon > 0$, we are done. What is the set E? We can take all linear combinations of indicator functions $g(x_1, \ldots, x_d) = \chi_{(a_1,b_1)}(x_1) \cdots \chi_{(a_d,b_d)}(x_d)$ for $-\infty < a_k < b_k < \infty$. In this case we obtain that $\widehat{g}(x_1, \ldots, x_d) = \widehat{\chi}_{(a_1,b_1)}(x_1) \cdots \widehat{\chi}_{(a_d,b_d)}(x_d)$ and

$$\widehat{\chi}_{(a_k,b_k)}(x_k) = \int_{a_k}^{b_k} e^{ix_k y_k} dy_k = \frac{1}{ix_k}\left(e^{ix_k b_k} - e^{ix_k a_k}\right) \longrightarrow 0, \quad \text{as } |x_k| \longrightarrow \infty.$$

□

The second result we need is

Proposition 2.4.10 (Stone & Weierstrass)**.** *Assume that* $A \subseteq C_0(\mathbb{R}^d;\mathbb{C})$ *satisfies the following properties:*

(i) *A is a linear space.*

(ii) *$g_1, g_2 \in A$ implies $g_1 g_2 \in A$.*

(iii) *$g \in A$ implies $\bar{g} \in A$.*

(iv) *For all $x_0 \in \mathbb{R}^d$ there is a $g \in A$ such that $g(x_0) \neq 0$.*

(v) *For all $x_0 \neq x_1$ there is a $g \in A$ such that $g(x_0) \neq g(x_1)$.*

Then A is dense, that means that for all $g \in C_0(\mathbb{R}^d;\mathbb{C})$ there exists $g_n \in A$ such that $\lim_n \sup_x |g_n(x) - g(x)| = 0$.

We leave this without proof.

Corollary 2.4.11. *One has that*

$$A := \left\{\widehat{f} : \mathbb{R}^d \to \mathbb{C} : f \in L_1(\mathbb{R}^d;\mathbb{C})\right\} \subseteq C_0(\mathbb{R}^d;\mathbb{C})$$

is dense.

Proof. Proposition 2.4.9 implies $A \subseteq C_0(\mathbb{R}^d;\mathbb{C})$. Now we have to check that the conditions of Proposition 2.4.10 are satisfied.
(i) is clear.
(ii) $\widehat{f_1}\widehat{f_2} = \widehat{f_1 * f_2}$ and $\|f_1 * f_2\|_{L_1} \le \|f_1\|_{L_1}\|f_2\|_{L_1}$
(iii) Here we have that

$$\overline{\widehat{f}}(x) = \overline{\int_{\mathbb{R}^d} e^{i\langle x,y\rangle} f(y)dy} = \int_{\mathbb{R}^d} e^{-i\langle x,y\rangle}\overline{f}(y)dy = \widehat{\overline{f}}(-x).$$

(iv) Let $x_0 \in \mathbb{R}^d$, $V := \{x_0 + y : |y| \le 1\}$, $f(x) := \chi_V(x)e^{-i\langle x_0,x\rangle} \in L_1(\mathbb{R}^d;\mathbb{C})$. Then

$$\widehat{f}(x_0) = \int_{\mathbb{R}^d} e^{i\langle x_0,x\rangle}\chi_V(x)e^{-i\langle x_0,x\rangle}d\lambda_d(x) = \lambda_d(V) > 0.$$

(v) Let $x_0 \ne x_1$ and $z_0 := (x_0 - x_1)/|x_0 - x_1|^2$. Then $e^{i\langle x_0,z_0\rangle} \ne e^{i\langle x_1,z_0\rangle}$ because of $e^{i\langle x_0-x_1,z_0\rangle} = e^i \ne 1$ and $e^{i\langle x_0,z\rangle} \ne e^{i\langle x_1,z\rangle}$ for all $z \in W$, where W is a small neighborhood of z_0. Let $f(x) := \chi_W(x)(e^{-i\langle x_0,x\rangle} - e^{-i\langle x_1,x\rangle})$. Then

$$\begin{aligned}
\widehat{f}(x_0) - \widehat{f}(x_1) &= \int_W \left[e^{-i\langle x_0,y\rangle} - e^{-i\langle x_1,y\rangle}\right]\left[e^{i\langle x_0,y\rangle} - e^{i\langle x_1,y\rangle}\right] d\lambda_d(y) \\
&= \int_W \left|e^{i\langle x_0,y\rangle} - e^{i\langle x_1,y\rangle}\right|^2 d\lambda_d(y) > 0.
\end{aligned}$$

$\square$

Now we finish the proof of the uniqueness theorem Proposition 2.4.7. We got that

$$\int_{\mathbb{R}^d} \widehat{f}(x)d\mu(x) = \int_{\mathbb{R}^d} \widehat{f}(x)d\nu(x)$$

for $f \in L_1(\mathbb{R}^d;\mathbb{C})$ and want to show that $\mu = \nu$. Assuming $g \in C_0(\mathbb{R}^d;\mathbb{C})$ and $f_n \in L_1(\mathbb{R}^d;\mathbb{C})$ such that $\|g - \widehat{f_n}\|_{C_0} \xrightarrow[n]{} 0$, we get that

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} g(x)d\mu(x) - \int_{\mathbb{R}^d} g(x)d\nu(x)\right| \\
= &\left|\int_{\mathbb{R}^d}\left[g(x) - \widehat{f_n}(x)\right]d\mu(x) - \int_{\mathbb{R}^d}\left[g(x) - \widehat{f_n}(x)\right]d\nu(x)\right| \\
\le &\ 2\left\|g - \widehat{f_n}\right\|_{C_0} \xrightarrow[n]{} 0.
\end{aligned}$$

Let us define now the system

$$\Pi := \{(a_1, b_1] \times \cdots \times (a_d, b_d] : -\infty < a_k \le b_k < \infty,\ k = 1, \ldots, d\}$$

For large n we find functions

$$g_k^{(n)}(x) := \begin{cases} 1 : & x \in [a_k + \frac{1}{n}, b_k] \\ 0 : & x \le a_k \text{ or } x \ge b_k + \frac{1}{n} \\ \text{linear} : & \text{otherwise} \end{cases}$$

so that

$$\lim_{n\to\infty} \prod_{k=1}^{d} g_k^{(n)}(x_k) = \chi_{\prod_{k=1}^d (a_k,b_k]}(x).$$

Since $g^n(x) := \prod_{k=1}^d g_k^{(n)}(x_k) \in C_0(\mathbb{R}^d;\mathbb{C})$ we get by majorized convergence that

$$\begin{aligned} \mu\left(\prod_{k=1}^{d}(a_k,b_k]\right) &= \int_{\mathbb{R}^d} \lim_{n\to\infty} g^n(x)d\mu(x) \\ &= \lim_{n\to\infty}\int_{\mathbb{R}^d} g^n(x)d\mu(x) \\ &= \lim_{n\to\infty}\int_{\mathbb{R}^d} g^n(x)d\nu(x) \\ &= \nu\left(\prod_{k=1}^{d}(a_k,b_k]\right). \end{aligned}$$

Since Π is a π-system which generates $\mathcal{B}(\mathbb{R}^d)$, we are done. □

Next, we state the important

Proposition 2.4.12 (Bochner & Chinčin). *Assume that $\varphi : \mathbb{R}^d \to \mathbb{C}$ is continuous with $\varphi(0) = 1$. Then the following assertions are equivalent*

(i) *φ is the Fourier transform of some $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$.*

(ii) *φ is positive semi-definite, i.e. for all $n = 1, 2, \ldots$ for all $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$*

$$\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \varphi(x_k - x_l) \geq 0.$$

Next we have an explicit inversion formula.

Proposition 2.4.13. (i) *Let $\mu \in \mathcal{M}_1^+(\mathbb{R})$ and let $F(b) = \mu((-\infty, b])$ be its distribution function. Then*

$$F(b) - F(a) = \lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-iya} - e^{-iyb}}{iy}\widehat{\mu}(y)d\lambda(y),$$

if $a < b$ and a and b are points of continuity of F.

(ii) *If* $\mu \in \mathcal{M}_1^+(\mathbb{R})$ *and* $\int_{\mathbb{R}} |\widehat{\mu}(x)| d\lambda(x) < \infty$*, then* μ *has a continuous density* $f : \mathbb{R} \to [0, \infty)$*, i.e.*

$$\mu(B) = \int_B f(x) d\lambda(x).$$

Moreover,

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixy} \widehat{\mu}(y) d\lambda(y).$$

Proposition 2.4.14. *Let* $\mu \in \mathcal{M}_1^+(\mathbb{R})$*. Then* $\mu(B) = \mu(-B)$ *for all* $B \in \mathcal{B}(\mathbb{R})$*, if and only if* $\widehat{\mu}(x) \in \mathbb{R}$ *for all* $x \in \mathbb{R}$*.*

The proof is an exercise.

The Bochner & Chinčin assumption *positive semi-definite* is sometimes difficult to check. There is an easier sufficient condition:

Proposition 2.4.15 (Polya)**.** *Let* $\varphi : \mathbb{R} \to [0, \infty)$ *be*

(i) *continuous,*

(ii) *even (i.e.* $\varphi(x) = \varphi(-x)$*),*

(iii) *convex on* $[0, \infty)$*,*

(iv) *and assume that* $\varphi(0) = 1$ *and* $\lim_{x \to \infty} \varphi(x) = 0$*.*

Then there exists some $\mu \in \mathcal{M}_1^+(\mathbb{R})$ *such that* $\widehat{\mu}(x) = \varphi(x)$*.*

2.5 Examples

Normal distribution on $\mathbb{R}$. Recall that

$$\gamma(B) := \int_B e^{-\frac{x^2}{2}} \frac{d\lambda_1(x)}{\sqrt{2\pi}}$$

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

Lemma 2.5.1. *One has that*

$$\widehat{\gamma}(x) = e^{-\frac{x^2}{2}}.$$

Proof. We will not give all details. By definition

$$\widehat{\gamma}(x) = \int_{\mathbb{R}} e^{ixy} e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}} = \int_{\mathbb{R}} \cos(xy) e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}}.$$

But then

$$\widehat{\gamma}'(x) = -\int_{\mathbb{R}} \sin(xy) y e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}} = -x \int_{\mathbb{R}} \cos(xy) e^{-\frac{y^2}{2}} \frac{d\lambda_1(y)}{\sqrt{2\pi}} = -x\widehat{\gamma}(x)$$

by partial integration. Letting $y \geq 0$ and $E(y) := \widehat{\gamma}(\sqrt{2y})$ gives

$$E'(y) = \sqrt{2}\frac{1}{2}\frac{1}{\sqrt{y}}\widehat{\gamma}'(\sqrt{2y}) = \frac{1}{\sqrt{2y}}\widehat{\gamma}'(\sqrt{2y}) = -\frac{\sqrt{2y}}{\sqrt{2y}}\widehat{\gamma}(\sqrt{2y}) = -E(y)$$

for $y > 0$. Since $E(0) = 1$, we get $E(y) = e^{-y}$ and $\widehat{\gamma}(x) = E(\frac{x^2}{2}) = e^{-\frac{x^2}{2}}$. □

Now we shift and stretch the normal distribution. Let $\sigma > 0$ and $m \in \mathbb{R}$. Then

$$\gamma_{m,\sigma^2}(B) := \int_B e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}}.$$

Proposition 2.5.2. (i) $\gamma_{m,\sigma^2} \in \mathcal{M}_1^+(\mathbb{R})$.

(ii) $\int_{\mathbb{R}} x d\gamma_{m,\sigma^2}(x) = m$ *is the mean.*

(iii) $\int_{\mathbb{R}} (x-m)^2 d\gamma_{m,\sigma^2}(x) = \sigma^2$ *is the variance.*

(iv) $\widehat{\gamma}_{m,\sigma^2}(x) = e^{imx} e^{-\frac{1}{2}\sigma^2 x^2}$.

Proof. (i) follows from

$$\int_{\mathbb{R}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} = \int_{\mathbb{R}} e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} = 1,$$

where we have used $y := \frac{x-m}{\sigma}$ and $\sigma dy = dx$ and the last equality was shown in the basic course.
(ii) is a consequence of

$$\int_{\mathbb{R}} x d\gamma_{m,\sigma^2}(x) = \int_{\mathbb{R}} x e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} \frac{dx}{\sqrt{2\pi\sigma^2}} = \int_{\mathbb{R}} (\sigma y + m) e^{-\frac{1}{2}y^2} \frac{dy}{\sqrt{2\pi}} = m$$

for $y := \frac{x-m}{\sigma}$.
(iii) is an exercise.
(iv) Here we get

$$\widehat{\gamma}_{m,\sigma^2}(x) = \int_{\mathbb{R}} e^{ixy} e^{-\frac{1}{2}\frac{(y-m)^2}{\sigma^2}} \frac{dy}{\sqrt{2\pi\sigma^2}}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} e^{ix(\sigma z+m)} e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}} \\
&= e^{ixm}\widehat{\gamma}(x\sigma) \\
&= e^{imx} e^{-\frac{1}{2}\sigma^2 x^2}.
\end{aligned}
$$

□

Definition 2.5.3. A measure $\mu \in \mathcal{M}_1^+(\mathbb{R})$ is a *Gaussian measure,* provided that $\mu = \delta_{\{m\}}$ for some $m \in \mathbb{R}$ or $\mu = \gamma_{m,\sigma^2}$ for $\sigma^2 > 0$ and $m \in \mathbb{R}$.

Normal distribution on $\mathbb{R}^d$. There are various ways to introduce this distribution.

Definition 2.5.4. The measure $\gamma = \gamma^{(d)} \in \mathcal{M}_1^+(\mathbb{R}^d)$ given by

$$\gamma(B) := \int_B e^{-\frac{\langle x,x\rangle}{2}} \frac{dx}{\sqrt{2\pi}^d}$$

is called *standard Gaussian measure* on $\mathbb{R}^d$.

Lemma 2.5.5. $\gamma^{(d)} \in \mathcal{M}_1^+(\mathbb{R}^d)$, *that means*

$$\int_{\mathbb{R}^d} e^{-\frac{\langle x,x\rangle}{2}} \frac{dx}{\sqrt{2\pi}^d} = 1.$$

Proof. By Fubini's theorem we get that

$$
\begin{aligned}
\int_{\mathbb{R}^d} e^{-\frac{\langle x,x\rangle}{2}} \frac{dx}{\sqrt{2\pi}^d} &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} e^{-\frac{1}{2}x_1^2} \cdots e^{-\frac{1}{2}x_d^2} \frac{dx_1}{\sqrt{2\pi}} \cdots \frac{dx_d}{\sqrt{2\pi}} \\
&= \left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \right)^d = 1.
\end{aligned}
$$

□

Definition 2.5.6. A matrix $R = (r_{ij})_{i,j=1}^d$ is called *positive semi-definite*, provided that

$$\langle Rx, x\rangle = \sum_{k,l=1}^d r_{kl} x_k x_l \geq 0$$

for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. A matrix $R = (r_{ij})_{i,j=1}^d$ is called *symmetric*, provided that $r_{kl} = r_{lk}$ for all $k, l = 1, \ldots, d$.

Proposition 2.5.7. *Let $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then the following assertions are equivalent*

(i) *There exist a matrix $A = (\alpha_{kl})_{k,l=1}^d$ and a vector $m \in \mathbb{R}^d$ such that μ is the image measure of $\gamma^{(d)}$ with respect to the map $x \mapsto Ax + m$, i.e.*

$$\mu(B) = \gamma^{(d)}\left(\left\{x \in \mathbb{R}^d : Ax + m \in B\right\}\right).$$

(ii) *There exist a positive semi-definite and symmetric matrix $R = (r_{kl})_{k,l=1}^d$ and a vector $m' \in \mathbb{R}^d$ such that*

$$\widehat{\mu}(x) = e^{i\langle x,m'\rangle - \frac{1}{2}\langle Rx,x\rangle}.$$

(iii) *For all $b = (b_1, \ldots, b_d) \in \mathbb{R}^d$ the law of $\varphi_b : \mathbb{R}^d \to \mathbb{R}$, $\varphi_b(x) := \langle x, b\rangle$, with respect to $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu\right)$ is a Gaussian measure on the real line $\mathbb{R}$.*

In particular, we have that m, R and m' are unique and that

(a) $m = m'$ *and* $R = AA^{\mathrm{T}}$,

(b) $\int_{\mathbb{R}^d} x_k d\mu(x) = m_k$ *and*

(c) $\int_{\mathbb{R}^d} (x_k - m_k)(x_l - m_l)\, d\mu(x) = r_{kl}$.

Definition 2.5.8. The above measure μ is called Gaussian measure on $\mathbb{R}^d$ with mean $m = (m_k)_{k=1}^d$ and covariance $R = (r_{kl})_{k,l=1}^d$.

Proof of Proposition 2.5.7. (i) $\Rightarrow$ (ii) follows from

$$\begin{aligned}
\widehat{\mu}(x) &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\mu(y) = \int_{\mathbb{R}^d} e^{i\langle x,m+Ay\rangle} d\gamma^{(d)}(y) \\
&= e^{i\langle x,m\rangle} \int_{\mathbb{R}} e^{i\left\langle A^{\mathrm{T}}x,y\right\rangle} d\gamma^{(d)}(y) = e^{i\langle x,m\rangle}\widehat{\gamma}^{(d)}\left(A^{\mathrm{T}}x\right) \\
&= e^{i\langle x,m\rangle} e^{-\frac{1}{2}\left\langle A^{\mathrm{T}}x,A^{\mathrm{T}}x\right\rangle} = e^{i\langle x,m\rangle} e^{-\frac{1}{2}\left\langle AA^{\mathrm{T}}x,x\right\rangle}
\end{aligned}$$

so that $AA^{\mathrm{T}} = R$ and $m = m'$ where we have used that

$$\begin{aligned}
\widehat{\gamma}^{(d)}(x) &= \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} d\gamma^{(d)}(y) \\
&= \int_{\mathbb{R}} e^{ix_1y_1} d\gamma^{(1)}(y_1) \cdots \int_{\mathbb{R}} e^{ix_dy_d} d\gamma^{(d)}(y_d) \\
&= e^{-\frac{x_1^2}{2}} \cdots e^{-\frac{x_d^2}{2}} = e^{-\frac{1}{2}\langle x,x\rangle}.
\end{aligned}$$

(ii) $\Rightarrow$ (iii): We compute the Fourier transform of law(φ_b):

$$\begin{aligned}
\widehat{\mathrm{law}(\varphi_b)}(t) &= \int_{\mathbb{R}} e^{its} d\mathrm{law}(\varphi_b)(s) = \int_{\mathbb{R}^d} e^{it\varphi_b(y)} d\mu(y) \\
&= \int_{\mathbb{R}^d} e^{it\langle b,y\rangle} d\mu(y) = \widehat{\mu}(tb) \\
&= e^{i\langle tb,m'\rangle - \frac{1}{2}\langle Rtb,tb\rangle} = e^{it\langle b,m'\rangle - \frac{1}{2}t^2\langle Rb,b\rangle}.
\end{aligned}$$

But this is the Fourier transform of a Gaussian measure on $\mathbb{R}$.
(iii) $\Rightarrow$ (i): For all $b \in \mathbb{R}^d$ there are $m_b \in \mathbb{R}^d$ and $\sigma_b \geq 0$ such that

$$\int_{\mathbb{R}^d} e^{it\langle b,y\rangle} d\mu(y) = e^{im_b t - \frac{1}{2}\sigma_b^2 t^2}.$$

Now we compute m_b and σ_b. From Proposition 2.5.2. We know that $\int_{\mathbb{R}^d} \langle b,x\rangle \, d\mu(x) = m_b$ and $\int_{\mathbb{R}^d} (\langle b,x\rangle - m_b)^2 \, d\mu(x) = \sigma_b^2$. The first equation implies that

$$\left\langle b, \left(\int_{\mathbb{R}^d} \langle e_1, x\rangle \, d\mu(x), \ldots, \int_{\mathbb{R}^d} \langle e_d, x\rangle \, d\mu(x) \right) \right\rangle = m_b$$

and $\langle b, m''\rangle = m_b$ with $m'' = \left(\int_{\mathbb{R}^d} \langle e_k, x\rangle \, d\mu(x) \right)_{k=1}^d$. Knowing this, we can rewrite the second equation as

$$\int_{\mathbb{R}^d} \langle b, x - m''\rangle^2 \, d\mu(x) = \sigma_b^2.$$

Defining

$$r'_{kl} := \int_{\mathbb{R}^d} \langle e_k, x - m''\rangle \langle e_l, x - m''\rangle \, d\mu(x)$$

we get that $R' = (r'_{kl})_{k,l=1}^d$ is symmetric and that

$$\begin{aligned}
\langle R'b, b\rangle &= \sum_{k,l=1}^d r'_{kl} b_k b_l \\
&= \sum_{k,l=1}^d \int_{\mathbb{R}^d} \langle e_k, x - m''\rangle \langle e_l, x - m''\rangle \, d\mu(x) b_k b_l \\
&= \int_{\mathbb{R}^d} \left(\sum_{k=1}^d b_k \langle e_k, x - m''\rangle \right)^2 d\mu(x) \\
&= \int_{\mathbb{R}^d} \langle b, x - m''\rangle^2 \, d\mu(x) = \sigma_b^2.
\end{aligned}$$

Consequently,

$$\int_{\mathbb{R}^d} e^{i\langle b,y\rangle} d\mu(y) = e^{i\langle b,m''\rangle - \frac{1}{2}\langle R'b,b\rangle},$$

where R' is positive semi-definite because of $\langle R'b, b\rangle = \sigma_b^2 \geq 0$. To get (i) we use algebra: There exists a matrix A such that $R' = AA^{\mathrm{T}}$. Hence

$$\int_{\mathbb{R}^d} e^{i\langle b,y\rangle} d\mu(y) = e^{i\langle b,m''\rangle - \frac{1}{2}\langle A^{\mathrm{T}}b, A^{\mathrm{T}}b\rangle} = \widehat{\mathrm{law}(\varphi)}(b),$$

where $\varphi(x) = m'' + Ax$ and $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \gamma^{(d)})$ is taken as probability space. Finally, we check (a), (b) and (c). Using (ii) $\Rightarrow$ (iii) gives that $\int_{\mathbb{R}^d} \langle x, b\rangle \, d\mu(x) = \langle b, m'\rangle$ and $\int_{\mathbb{R}^d} \langle x - m', b\rangle^2 \, d\mu(x) = \langle Rb, b\rangle$, so that $\langle Rb, b\rangle = \langle R'b, b\rangle$ for all $b \in \mathbb{R}^d$. Since R and R' are symmetric, we may deduce

$$\begin{aligned} \langle Rb_1, b_2\rangle &= \frac{1}{4}\left[\langle R(b_1+b_2), b_1+b_2\rangle - \langle R(b_1-b_2), b_1-b_2\rangle\right] \\ &= \frac{1}{4}\left[\langle R'(b_1+b_2), b_1+b_2\rangle - \langle R'(b_1-b_2), b_1-b_2\rangle\right] = \langle R'b_1, b_2\rangle \end{aligned}$$

and $R = R'$ which proves (c). Using (i), we get

$$\int_{\mathbb{R}^d} x d\mu(x) = \int_{\mathbb{R}^d} (m + Ax)\, d\gamma^{(d)}(x) = m,$$

so that (b) is proved and that $m = m'$. Finally, $R = AA^{\mathrm{T}}$ follows now from (i) $\Rightarrow$ (ii). $\square$

Definition 2.5.9. A Gaussian measure $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ with mean m and covariance R is *degenerated* if $\mathrm{rank}(R) < d$. The Gaussian measure is called *non-degenerated* if $\mathrm{rank}(R) = d$.

Examples

(a) $\gamma^{(d)}$ is non-degenerated since

$$R = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

has rank d.

(b) Let $d = 2$ and $\mu = \gamma^{(1)} \times \delta_{\{x_{2,0}\}}$, i.e.

$$\mu(B) := \int_{\mathbb{R}} e^{-\frac{1}{2}(x_1^2 + x_{2,0}^2)} \chi_{(x_1, x_{2,0}) \in B} \frac{dx_1}{\sqrt{2\pi}}.$$

Let us compute the mean:

$$\int_{\mathbb{R}^2} x_1 d\mu(x) = 0 \quad \text{and} \quad \int_{\mathbb{R}^2} x_2 d\mu(x) = x_{2,0}.$$

Moreover

$$\int_{\mathbb{R}^2} (x_1 - 0)^2 d\mu(x) = 1,$$

$$\int_{\mathbb{R}^2} (x_2 - x_{2,0})^2 d\mu(x) = \int_{\mathbb{R}} (x_2 - x_{2,0})^2 d\delta_{\{x_{2,0}\}} = 0,$$

and

$$\int_{\mathbb{R}^2} (x_1 - 0)(x_2 - x_{2,0}) d\mu(x) = 0.$$

Consequently, $R = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\operatorname{rank}(R) = 1$.

In the case of non-degenerate measures we have the

Proposition 2.5.10. *Assume that $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ is a non-degenerate Gaussian measure with covariance R and mean m. Then one has that*

$$\mu(B) = \int_B e^{-\frac{1}{2}\left\langle R^{-1}(x-m), x-m \right\rangle} \frac{dx}{(2\pi)^{\frac{d}{2}} |det\ R|^{\frac{1}{2}}}.$$

We will not prove this. The proof is a computation.

Cauchy distribution on $\mathbb{R}$. We start with an experiment. Take $\alpha > 0$ and a point source which sends small particles to a wall. The angle $\varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is distributed uniformly, that means that it has the uniform distribution on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Problem: What is the probability that a particle hits a set B? Here we get, for $0 \leq \varphi < \frac{\pi}{2}$, $\mu_\alpha([0, \alpha \tan \varphi]) = \frac{\varphi}{\pi}$ and, for $0 \leq x < \infty$,

$$\begin{aligned} \mu_\alpha([0,x]) &= \frac{\arctan \frac{x}{\alpha}}{\pi} = \frac{1}{\pi} \int_0^{\frac{x}{\alpha}} \frac{d\xi}{1+\xi^2} \\ &= \frac{1}{\pi\alpha} \int_0^x \frac{d\eta}{1 + \left(\frac{\eta}{\alpha}\right)^2} = \frac{\alpha}{\pi} \int_0^x \frac{d\eta}{\alpha^2 + \eta^2}. \end{aligned}$$

Definition 2.5.11. For $\alpha > 0$ the distribution

$$d\mu_\alpha(x) := \frac{\alpha}{\pi} \frac{1}{\alpha^2 + x^2} dx \in \mathcal{M}_1^+(\mathbb{R}^d)$$

is called *Cauchy distribution with parameter $\alpha > 0$.*

Proposition 2.5.12. *One has that*

$$\widehat{\mu}_\alpha(x) = e^{-\alpha|x|}.$$

Proof. We prove the statement for $\alpha = 1$. (The rest can be done by a change of variables.) Since $\widehat{\mu}_1(-x) = \overline{\widehat{\mu}_1(x)}$ we can restrict our proof to $x \geq 0$. We consider the meromorphic function $f : \mathbb{C} \to \mathbb{C}$, $f(z) := \frac{e^{ixz}}{1+z^2}$, which has its residuals in the $z \in \mathbb{C}$ such that $1 + z^2 = 0$, that means $z_1 = i$ and $z_2 = -i$. From complex analysis it is known that

$$\lim_{z \to i}(z - i)f(z) = \frac{1}{2\pi i}\left[\int_{-R}^{R} f(z)dz + \int_{S_R} f(z)dz\right].$$

Since

$$\int_{S_R} \frac{e^{ixz}}{1+z^2}dz \longrightarrow 0, \text{ as } R \longrightarrow \infty$$

and

$$\lim_{z \to i}(z - i)f(z) = \lim_{z \to i}\frac{e^{ixz}}{z+i} = \frac{1}{2i}e^{-x},$$

we obtain that

$$\frac{1}{2i}e^{-x} = \frac{1}{2\pi i}\lim_{R \to \infty}\int_{-R}^{R}\frac{e^{ixz}}{1+z^2}dz = \frac{1}{2i}\widehat{\mu}_1(x).$$

□

2.6 Independent random variables

We recall two facts from the basic course.

Proposition 2.6.1. *If $f, g : \Omega \to \mathbb{R}$ are independent such that $\mathbb{E}|f| < \infty$ and $\mathbb{E}|g| < \infty$. Then $\mathbb{E}|fg| < \infty$ and $\mathbb{E}fg = \mathbb{E}f\mathbb{E}g$.*

Proposition 2.6.2. *Let $(\Omega, \mathcal{F}, \mathbb{P}) = \times_{i=1}^{n} (\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ and let $g_i : \Omega \to \mathbb{R}$ be random variables. If $f_i(\omega_1, \ldots, \omega_n) := g_i(\omega_i)$, then $f_1, \ldots, f_n$ are independent and $\mathrm{law}(f_i) = \mathrm{law}(g_i)$.*

Proof. Letting $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ we get that

$$\begin{aligned}\mathbb{P}(f_1 \in B_1, \ldots, f_n \in B_n) &= \mathbb{P}\left(\times_{i=1}^{n} \{g_i \in B_i\}\right) = \prod_{i=1}^{n}\mathbb{P}_i(g_i \in B_i) \\ &= \prod_{i=1}^{n}\mathbb{P}(f_i \in B_i).\end{aligned}$$

□

Proposition 2.6.3. *Let $f, g : \Omega \to \mathbb{R}$ be random variables. Then the following assertions are equivalent.*

(i) *f and g are independent.*

(ii) *If $h : \Omega \to \mathbb{R}^2$ is defined by $h(\omega) := (f(\omega), g(\omega))$, then $\widehat{h}(x,y) = \widehat{f}(x)\widehat{g}(y)$.*

Proof. (i)⇒(ii): By Proposition 2.6.1

$$\begin{aligned}\widehat{h}(x,y) &= \int_\Omega e^{i(xf(\omega)+yg(\omega))} d\mathbb{P}(\omega) \\ &= \int_\Omega e^{ixf(\omega)} d\mathbb{P}(\omega) \int_\Omega e^{iyg(\omega)} d\mathbb{P}(\omega) = \widehat{f}(x)\widehat{g}(y).\end{aligned}$$

(ii)⇒(i): Define $H : \Omega \times \Omega :\to \mathbb{R}^2$ by $H(\omega_1, \omega_2) := (f(\omega_1), g(\omega_2))$. Then the coordinates are independent and $\widehat{H}(x,y) = \widehat{f}(x)\widehat{g}(y) = \widehat{h}(x,y)$ so that the law of H and h are the same. But this implies that

$$\begin{aligned}\mathbb{P}(f \in B_1, g \in B_2) &= \mathbb{P}(h \in B_1 \times B_2) \\ &= (\mathbb{P} \times \mathbb{P})(H \in B_1 \times B_2) = \mathbb{P}(f \in B_1)\,\mathbb{P}(g \in B_2).\end{aligned}$$

□

We consider an application of this:

Definition 2.6.4. Two random variables $f, g : \Omega \to \mathbb{R}$ with $\mathbb{E}f^2 + \mathbb{E}g^2 < \infty$ are called *uncorrelated*, provided that $\mathbb{E}(f - \mathbb{E}f)(g - \mathbb{E}g) = 0$.

Remark 2.6.5. If f and g are independent and if $\mathbb{E}f^2 + \mathbb{E}g^2 < \infty$, then they are uncorrelated. In fact, we have (by Proposition 2.6.1) that

$$\mathbb{E}(f - \mathbb{E}f)(g - \mathbb{E}g) = [\mathbb{E}(f - \mathbb{E}f)]\,[\mathbb{E}(g - \mathbb{E}g)] = 0.$$

Proposition 2.6.6. *Let $f, g : \Omega \to \mathbb{R}$ be random variables such that $(f,g) : \Omega \to \mathbb{R}^2$ is a Gaussian random variable. Then the following assertions are equivalent.*

(i) *f and g are uncorrelated.*

(ii) *f and g are independent.*

Proof. We only have to check that (i) implies (ii). We know from Proposition 2.5.7 that for $x = (x_1, x_2)$ one has

$$\widehat{(f,g)}(x_1, x_2) = e^{i(x_1 \mathbb{E}f + x_2 \mathbb{E}g) - \frac{1}{2}\langle Rx, x\rangle},$$

with

$$R = \begin{pmatrix} \mathbb{E}(f - \mathbb{E}f)^2 & \mathbb{E}(f - \mathbb{E}f)(g - \mathbb{E}g) \\ \mathbb{E}(f - \mathbb{E}f)(g - \mathbb{E}g) & \mathbb{E}(g - \mathbb{E}g)^2 \end{pmatrix} = \begin{pmatrix} \mathrm{var}(f) & 0 \\ 0 & \mathrm{var}(g) \end{pmatrix}.$$

Consequently,

$$\widehat{(f,g)}(x_1, x_2) = e^{ix_1 \mathbb{E}f - \frac{1}{2}x_1^2 \mathrm{var}(f)} e^{ix_2 \mathbb{E}g - \frac{1}{2}x_2^2 \mathrm{var}(g)} = \widehat{f}(x_1)\widehat{g}(x_2).$$

Applying Proposition 2.6.3 we get the independence of f and g. □

Warning: We need that the joint distribution of f and g (in other words $\mathrm{law}(f,g) \in \mathcal{M}_1^+(\mathbb{R}^2)$) is Gaussian.

2.7 Moments of measures

There are different types of moments of a measure $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Given integers $l_1, \ldots, l_d \geq 0$ and $0 < p < \infty$ we have for example that

$$\begin{aligned} \int_{\mathbb{R}^d} x_1^{l_1} \cdots x_d^{l_d} d\mu(x_1, \ldots, x_d) &= \text{moment of order } (l_1, \ldots, l_d), \\ \int_{\mathbb{R}^d} \left| x_1^{l_1} \cdots x_d^{l_d} \right| d\mu(x_1, \ldots, x_d) &= \text{absolute moment of order } (l_1, \ldots, l_d), \\ \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} x d\mu(x) \right|^p d\mu(x) &= \text{centered absolute } p\text{-th moment.} \end{aligned}$$

We are interested in the first type and show that one can use the Fourier transform to compute these moments:

Proposition 2.7.1. *Let $\mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ and assume integers $k_1, \ldots, k_d \geq 0$ such that for all integers $0 \leq l_j \leq k_j$ one has that*

$$\int_{\mathbb{R}^d} \left| x_1^{l_1} \cdots x_d^{l_d} \right| d\mu(x_1, \ldots, x_d) < \infty.$$

Then one has the following:

(i)
$$\frac{\partial^{l_1+\cdots+l_d}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \widehat{\mu} \in C_b(\mathbb{R}^d).$$

(ii)
$$\frac{\partial^{l_1+\cdots+l_d}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \widehat{\mu}(x) = i^{l_1+\cdots+l_d} \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} y_1^{l_1} \cdots y_d^{l_d} d\mu(y).$$

(iii)
$$\frac{\partial^{l_1+\cdots+l_d}}{\partial x_1^{l_1} \cdots \partial x_d^{l_d}} \widehat{\mu}(0) = i^{l_1+\cdots+l_d} \int_{\mathbb{R}^d} y_1^{l_1} \cdots y_d^{l_d} d\mu(y).$$

(iv) *The partial derivatives of $\widehat{\mu}$ are uniformly continuous.*

Example 2.7.2. (a) Binomial distribution: $0 < p < 1$, $d = 1$, $n \in \{1, 2, \ldots\}$.

$$\mu(\{k\}) := \binom{n}{k} p^{n-k}(1-p)^k, \; k = 1, \ldots, n.$$

$$\widehat{\mu}(x) = \left[p + (1-p)e^{ix}\right]^n, \quad \widehat{\mu}'(x) = n\left[p + (1-p)e^{ix}\right]^{n-1}(1-p)ie^{ix},$$

$$\widehat{\mu}'(0) = n(1-p)i, \quad \frac{\widehat{\mu}'(0)}{i} = n(1-p).$$

(b) Gaussian measure $\gamma_{0,\sigma^2} \in \mathcal{M}_1^+(\mathbb{R})$: We have that $\widehat{\gamma}_{0,\sigma^2}(x) = e^{-\frac{1}{2}\sigma^2 x^2}$ and get

$$\widehat{\gamma}'_{0,\sigma^2}(x) = -x\sigma^2 e^{-\frac{1}{2}\sigma^2 x^2}, \quad \widehat{\gamma}'_{0,\sigma^2}(0) = 0,$$

$$\widehat{\gamma}''_{0,\sigma^2}(x) = (x^2\sigma^4 - \sigma^2)e^{-\frac{1}{2}\sigma^2 x^2}, \quad \widehat{\gamma}''_{0,\sigma^2}(0) = -\sigma^2.$$

(c) Cauchy distribution $\mu_\alpha \in \mathcal{M}_1^+(\mathbb{R})$ with $\alpha > 0$: Recall that

$$d\mu_\alpha(x) = \frac{\alpha}{\pi}\frac{1}{\alpha^2 + x^2} \quad \text{and} \quad \widehat{\mu}(x) = e^{-\alpha|x|}.$$

Proposition 2.7.3. *For all $\alpha > 0$ one has*

$$\int_{\mathbb{R}} |x|^k d\mu_\alpha(x) = \infty \quad \textit{for} \quad k = 1, 2, \ldots$$

Proof. A first variant of the proof is

$$\lim_{x\downarrow 0}\frac{e^{-\alpha|x|}-1}{x} = -\alpha \neq \alpha = \lim_{x\uparrow 0}\frac{e^{-\alpha|x|}-1}{x}.$$

Or we can use

$$\int_{\mathbb{R}}\frac{|x|^k}{(\alpha^2+x^2)}dx \geq \frac{1}{\alpha^2+1}\int_{|x|\geq 1}\frac{|x|^k}{x^2}dx = 2\int_1^\infty x^{k-2}dx = \infty.$$

□

Proof of Proposition 2.7.1. We only prove the case $l_1 = 1, l_2 = \cdots = l_d = 0$ (the rest follows by induction). Fix $x_2, \ldots, x_d \in \mathbb{R}$ and define $f(x_1, y) := e^{i\langle(x_1,\ldots,x_d),y\rangle}$. Then

$$\frac{\partial}{\partial x_1}\int_{\mathbb{R}^d} e^{i\langle x,y\rangle}d\mu(y) = \int_{\mathbb{R}^d}\frac{\partial}{\partial x_1}e^{i\langle x,y\rangle}d\mu(y) = i\int_{\mathbb{R}^d} e^{i\langle x,y\rangle}y_1 d\mu(y),$$

where the first inequality has to be justified. Now we define $d\upsilon_+(y) = \chi_{\{y_1\geq 0\}}y_1 d\mu(y)$ and $d\upsilon_-(y) = -\chi_{\{y_1<0\}}y_1 d\mu(y)$ and obtain bounded measures, so that

$$x \mapsto \int_{\mathbb{R}^d} e^{i\langle x,y\rangle}d\upsilon_\pm(y)$$

are uniformly continuous and bounded and (iv) follows. □

For the equality we have to justify, we need

Lemma 2.7.4. *Let $f : \mathbb{R}\times\Omega \to \mathbb{C}$ be such that*

(i) $\frac{\partial f}{\partial x}(\cdot,\omega)$ *is continuous for all $\omega\in\Omega$,*

(ii) $\frac{\partial f}{\partial x}(x,\cdot)$ *and $f(x,\cdot)$ are random variables,*

(iii) *There exists a $g : \Omega\to\mathbb{R}$, $g(\omega)\geq 0$, such that $\left|\frac{\partial f}{\partial x}(x,\omega)\right| \leq g(\omega)$ for all $\omega\in\Omega$, $x\in\mathbb{R}$ and $\mathbb{E}g<\infty$,*

(iv) $\int_\Omega |f(x,\omega)|d\mathbb{P}(\omega) < \infty$ *for all $x\in\mathbb{R}$.*

Then

$$\frac{\partial}{\partial x}\int_\Omega f(x,\omega)d\mathbb{P}(\omega) = \int_\Omega\frac{\partial f}{\partial x}(x,\omega)d\mathbb{P}(\omega).$$

2.8 Weak convergence

In the beginning of the lecture we considered the following types of convergence: almost sure convergence, convergence in probability and L_p-convergence. Up to now we needed that the underlying probability spaces are the same. This will be relaxed by the weak convergence, where we only consider the convergence of the laws.

Proposition 2.8.1. *Let $\mu_n, \mu \in \mathcal{M}_1^+(\mathbb{R}^d)$. Then the following assertions are equivalent.*

(i) *For all continuous and bounded functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ one has that $\int \varphi(x) d\mu_n(x) \underset{n}{\longrightarrow} \int \varphi(x) d\mu(x)$.*

(ii) *For all closed sets $A \in \mathcal{B}(\mathbb{R}^d)$ one has $\overline{\lim}_n \mu_n(A) \leq \mu(A)$.*

(iii) *For all open sets $B \in \mathcal{B}(\mathbb{R}^d)$ one has $\underline{\lim}_n \mu_n(B) \geq \mu(B)$.*

(iv) *If $d = 1$ and if $F_n(x) := \mu_n((-\infty, x])$ and $F(x) := \mu((-\infty, x])$, then $F_n(x) \underset{n}{\longrightarrow} F(x)$ for all points $x \in \mathbb{R}$ of continuity of F.*

(v) *$\widehat{\mu}_n(x) \underset{n}{\longrightarrow} \widehat{\mu}(x)$ for $x \in \mathbb{R}^d$.*

Definition 2.8.2. (i) For $\mu_n, \mu \in \mathcal{M}_1^+(\mathbb{R}^d)$ we say that μ_n converges weakly to μ ($\mu_n \Rightarrow \mu$ or $\mu_n \xrightarrow{\text{w}} \mu$) provided that the conditions of Proposition 2.8.1 are satisfied.

(ii) Let $f_n : \Omega_n \to \mathbb{R}^d$ and $f : \Omega \to \mathbb{R}^d$ be random variables over probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$. Then f_n converges to f *weakly* or *in distribution* ($f_n \xrightarrow{\text{d}} f$) provided that the corresponding laws $\mu_n(B) = \mathbb{P}_n(f_n \in B)$ and $\mu(B) = \mathbb{P}(f \in B)$ are converging weakly.

What is the connection to our earlier types of convergence?

Proposition 2.8.3. *For $f_n, f : \Omega \to \mathbb{R}^d$ one has that if f_n converges to f in probability, then f_n converges to f in distribution.*

Proof. Letting $\varphi : \mathbb{R}^d \to \mathbb{R}$ be continuous and bounded, we need to show that

$$\mathbb{E}\varphi(f_n) = \int_{\mathbb{R}^d} \varphi(x) d\mu_n(x) \underset{n}{\longrightarrow} \int_{\mathbb{R}^d} \varphi(x) d\mu(x) = \mathbb{E}\varphi(f).$$

But this follows from (defining $\|\varphi\|_\infty := \sup_x |\varphi(x)|$)

$$|\mathbb{E}\varphi(f_n) - \mathbb{E}\varphi(f)| \leq \int_{\{|f_n - f| > \varepsilon\}} d\mathbb{P} 2\|\varphi\|_\infty + \int_{\{|f_n - f| \leq \varepsilon, |f| > N\}} d\mathbb{P} 2\|\varphi\|_\infty$$

$$
\begin{aligned}
& + \int_{\{|f_n - f| \le \varepsilon, |f| \le N\}} |\varphi(f_n) - \varphi(f)| \, d\mathbb{P} \\
\le \; & 2\|\varphi\|_\infty \left(\mathbb{P}(|f_n - f| > \varepsilon) + \mathbb{P}(|f| > N)\right) \\
& + \sup_{|x-y| \le \varepsilon, \; |y| \le N} |\varphi(x) - \varphi(y)| \\
\le \; & 3\delta,
\end{aligned}
$$

for $N = N(\delta)$, $\varepsilon = \varepsilon(N(\delta), \delta))$ and $n \ge n(\varepsilon)$. □

Proposition 2.8.4 (Central limit theorem)**.** *Let $f_1, f_2, \ldots : \Omega \to \mathbb{R}$ be a sequence of independent random variables which have the same distribution such that $0 < \mathbb{E}(f_k - \mathbb{E}f_k)^2 = \sigma^2 < \infty$ and $\mathbb{E}f_k = m$. Then*

$$
\mathbb{P}\left(\frac{1}{\sqrt{k\sigma^2}}\left((f_1 - m) + \cdots + (f_k - m)\right) \le x\right) \underset{k}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} d\xi.
$$

Proof. Let $f_n^0 := \frac{f_n - m}{\sigma}$. Then we get that $\mathbb{E}f_n^0 = \frac{1}{\sigma}(\mathbb{E}f_n - m) = 0$ and $\mathbb{E}(f_n^0)^2 = \frac{1}{\sigma^2}\mathbb{E}(f_n - m)^2 = 1$. We have to show that

$$
\mathbb{P}\left(\frac{1}{\sqrt{n}}\left(f_1^0 + \cdots + f_n^0\right) \le x\right) \underset{n}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} d\xi
$$

or, for $S_n := f_1^0 + \cdots + f_n^0$, $\frac{1}{\sqrt{n}}S_n \Rightarrow g \sim N(0,1)$. By Proposition 2.8.1 this is equivalent to

$$
\widehat{\frac{1}{\sqrt{n}}S_n}(t) \underset{n}{\longrightarrow} \widehat{g}(t) = e^{-\frac{t^2}{2}}
$$

for all $t \in \mathbb{R}$. Now

$$
\widehat{\frac{1}{\sqrt{n}}S_n}(t) = \left(\frac{1}{\sqrt{n}}\widehat{\left(f_1^0 + \cdots + f_n^0\right)}\right)(t) = \widehat{\frac{f_1^0}{\sqrt{n}}}(t) \cdots \widehat{\frac{f_n^0}{\sqrt{n}}}(t) = \varphi\left(\frac{t}{\sqrt{n}}\right)^n,
$$

if $\varphi(t) = \widehat{f_1^0}(t)$. Since $\mathbb{E}(f_1^0)^2 < \infty$, Proposition 2.7.1 implies that $\varphi'' \in C_b(\mathbb{R})$ and

$$
\begin{aligned}
\varphi(t) &= \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(0) + o(t^2) \\
&= \varphi(0) + ti\mathbb{E}f_1^0 + \frac{t^2}{2}i^2\mathbb{E}(f_1^0)^2 + o(t^2) \\
&= 1 - \frac{t^2}{2} + o(t^2)
\end{aligned}
$$

for $t \in \mathbb{R}$ with

$$
\lim_{t \to 0} \frac{o(t^2)}{t^2} = 0.
$$

So it remains to show that

$$\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \underset{n}{\longrightarrow} e^{-\frac{t^2}{2}}.$$

Letting $\theta = \frac{t^2}{2}$, this is done by

$$\left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \underset{n}{\longrightarrow} e^{-\theta}.$$

Fix $\theta \geq 0$, let $\varepsilon \in (0,1)$, and choose $n(\varepsilon,\theta) \in \mathbb{N}$ such that

$$\left| o\left(\frac{\theta}{n}\right)\right| \leq \varepsilon \left|\frac{\theta}{n}\right|$$

for $n \geq n(\varepsilon,\theta)$. Then

$$\left(1 - \frac{\theta(1+\varepsilon)}{n}\right)^n \leq \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \leq \left(1 - \frac{\theta(1-\varepsilon)}{n}\right)^n$$

for $n \geq n(\varepsilon,\theta)$ such that $\frac{\theta(1+\varepsilon)}{n} < 1$. Because

$$\lim_n \left(1 - \frac{\theta(1+\varepsilon)}{n}\right)^n = e^{-\theta(1+\varepsilon)} \quad \text{and} \quad \lim_n \left(1 - \frac{\theta(1-\varepsilon)}{n}\right)^n = e^{-\theta(1-\varepsilon)}$$

it follows that

$$e^{-\theta(1+\varepsilon)} \leq \liminf_n \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \leq \limsup_n \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \leq e^{-\theta(1-\varepsilon)}.$$

Since this is true for all $\varepsilon > 0$ we end up with

$$e^{-\theta} \leq \lim_n \left(1 - \frac{\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \leq e^{-\theta}$$

which finishes the proof. $\square$

Proposition 2.8.5 (Poisson). *Let $f_{n,1}, \ldots, f_{n,n} : \Omega \to \mathbb{R}$ be independent random variables such that $\mathbb{P}(f_{n,k} = 1) = p_{nk}$ and $\mathbb{P}(f_{n,k} = 0) = q_{nk}$ with $p_{nk} + q_{nk} = 1$. Assume that*

$$\max_{1\leq k\leq n} p_{nk} \to_n 0 \text{ and } \sum_{k=1}^n p_{nk} \to_n \lambda > 0.$$

Then, for $S_n := f_{n,1} + \cdots + f_{n,n}$, the laws $\mu_n := \mathrm{law}(S_n)$ converge weakly to the Poisson distribution

$$\pi_\lambda(B) = \sum_{k=0}^\infty e^{-\lambda} \frac{\lambda^k}{k!} \delta_{\{k\}}(B).$$

Proof. For $\theta = e^{it} - 1$ we get that

$$
\begin{aligned}
\widehat{S}_n(t) &= \prod_{k=1}^{n} \widehat{f}_{n,k}(t) = \prod_{k=1}^{n} \left(p_{nk}e^{it} + q_{nk}\right) \\
&= \prod_{k=1}^{n} \left(1 + p_{nk}\left(e^{it} - 1\right)\right) \\
&= \prod_{k=1}^{n} (1 + p_{nk}\theta) \\
&= 1 + \theta\left(\sum_{k=1}^{n} p_{nk}\right) + \theta^2\left(\sum_{1\le k_1<k_2\le n} p_{nk_1}p_{nk_2}\right) \\
&\quad + \cdots + \theta^l\left(\sum_{1\le k_1<\cdots<k_l\le n} p_{nk_1}\cdots p_{nk_l}\right) + \cdots \\
&\quad + \theta^n p_{n1}\cdots p_{nn} \\
&= 1 + \sum_{l=1}^{n} \frac{(\theta\lambda)^l}{l!} b_{ln}
\end{aligned}
$$

with

$$b_{ln} := \frac{l!}{\lambda^l} \sum_{1\le k_1<\cdots<k_l\le n} p_{nk_1}\cdots p_{nk_l}.$$

We do not give the Poisson's Theorem, we just show that

$$\lim_n b_{ln} = 1$$

for *fixed* l. This can be easily seen by

$$
\begin{aligned}
&\frac{l!}{\lambda^l} \sum_{1\le k_1<\cdots<k_l\le n} p_{nk_1}\cdots p_{nk_l} \\
&= \left(\frac{p_{n1} + \cdots + p_{nn}}{\lambda}\right)^l \frac{\sum_{\substack{1\le k_1,\ldots,k_l\le n \\ \text{all indices are distinct}}} p_{nk_1}\cdots p_{nk_l}}{\sum_{1\le k_1,\ldots,k_l\le n} p_{nk_1}\cdots p_{nk_l}}.
\end{aligned}
$$

The first factor converges to 1 as $n \to \infty$. The second one we write as

$$\frac{\sum_{\substack{1\le k_1,\ldots,k_l\le n \\ \text{all indices are distinct}}} p_{nk_1}\cdots p_{nk_l}}{\sum_{1\le k_1,\ldots,k_l\le n} p_{nk_1}\cdots p_{nk_l}} = 1 - \frac{\sum_{\substack{1\le k_1,\ldots,k_l\le n \\ \text{not all indices are distinct}}} p_{nk_1}\cdots p_{nk_l}}{\sum_{1\le k_1,\ldots,k_l\le n} p_{nk_1}\cdots p_{nk_l}}$$

and can bound the second term by

$$
\begin{aligned}
\sum_{\substack{1\le k_1,\ldots,k_l\le n \\ \text{not all indices are distinct}}} p_{nk_1}\cdots p_{nk_l} &\le (p_{n1} + \cdots + p_{nn})^{l-1} \max_k p_{nk} \\
&\quad + (p_{n1} + \cdots + p_{nn})^{l-2} \max_k p_{nk}^2
\end{aligned}
$$

$$+\cdots$$
$$+(p_{n1}+\cdots+p_{nn})\max_k p_{nk}^{l-1}.$$

This should motivate the convergence

$$\widehat{S}_n(t) \xrightarrow[n]{} e^{\lambda\theta} = e^{\lambda(e^{it}-1)} = \widehat{\pi}_\lambda(t).$$

□

Example 2.8.6. $p_{n,1} = \cdots = p_{n,n} := \frac{\lambda}{n}$ for $n > \lambda$.

Now we consider a limit theorem which is an extension of the central limit theorem.

Proposition 2.8.7. *Assume that $f_{n1}, f_{n2}, ..., f_{nn}$ are independent random variables such that*

$$\mathbb{E}f_{nk} = 0 \quad \textit{and} \quad \sum_{k=1}^{n} \mathbb{E}f_{nk}^2 = 1$$

for all $n = 1, 2, ...$ Let $\mu_{nk} := \text{law}(f_{nk})$ and assume that the Lindeberg condition

$$\sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 d\mu_{nk}(x) \to_n 0$$

is satisfied for all $\varepsilon > 0$. Then one has that

$$S_n \to_d N(0,1)$$

where $S_n := f_{n1} + \cdots + f_{nn}$.

Example 2.8.8. Let us check that the Lindeberg condition is satisfied in the case that we consider the weak limit of

$$\frac{1}{\sqrt{n}}(f_1 + \cdots + f_n)$$

where

(1) $\mathbb{E}f_k = 0$,

(2) $\mathbb{E}f_k^2 = 1$,

(3) $f_1^2, f_2^2, f_3^2, ...$ is a uniformly integrable family of random variables.

In the above notation we can write (in distribution) that

$$f_{nk} =_d \frac{f_k}{\sqrt{n}},$$

so that

$$\begin{aligned}
\sum_{k=1}^{n} \int_{|x|>\varepsilon} x^2 d\mu_{nk}(x) &= \sum_{k=1}^{n} \int_{\left|\frac{f_k}{\sqrt{n}}\right|>\varepsilon} \left(\frac{f_k}{\sqrt{n}}\right)^2 d\mathbb{P} \\
&= \frac{1}{n} \sum_{k=1}^{n} \int_{|f_k|>\sqrt{n}\varepsilon} f_k^2 d\mathbb{P} \\
&\leq \sup_{1\leq k\leq n} \int_{|f_k|>\sqrt{n}\varepsilon} f_k^2 d\mathbb{P} \\
&\leq \sup_{k\geq 1} \int_{|f_k|>\sqrt{n}\varepsilon} f_k^2 d\mathbb{P} \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Examples that the family $(f_k^2)_{k=1}^{\infty}$ is uniformly integrable are

(a) $f_1, f_2, ...$ are identical distributed,

(b) $\sup_k \mathbb{E}|f_k|^p < \infty$ for some $2 < p < \infty$.

2.9 A first ergodic theorem

We conclude the lecture with an extension of the Strong Law of Large Numbers.

Definition 2.9.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

(1) A measurable map $T : \Omega \to \Omega$ is called *measure preserving* provided that

$$\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A) \quad \text{for all} \quad A \in \mathcal{F}.$$

(2) A measure preserving map $T : \Omega \to \Omega$ is called *ergodic* provided that, for $A \in \mathcal{F}$, the condition

$$T^{-1}(A) = A \quad \text{implies} \quad \mathbb{P}(A) \in \{0, 1\}.$$

Now we get the following Ergodic Theorem.

Proposition 2.9.2 (Birkhoff & Chinčin)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $T : \Omega \to \Omega$ be ergodic, and $f : \Omega \to \mathbb{R}$ be a random variable such that $\mathbb{E}|f| < \infty$. Then one has that*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = \mathbb{E}f \quad a.s.$$

Why does the Ergodic Theorem of Birkhoff and Chinčin extend the Strong Law of Large Numbers? Let

$$(M,\Sigma,\mu) := \otimes_{n=1}^{\infty}(\Omega,\mathcal{F},\mathbb{P})$$

and define the shift $T : M \to M$ by

$$T(\omega_1,\omega_2,\omega_3,...) := (\omega_2,\omega_3,...).$$

The shift T is ergodic:

(a) T is measure-preserving: This can be checked on the π-system of cylinder-sets with a finite-dimensional basis.

(b) Assume now that $T^{-1}(A) = A$ for some $A \in \mathcal{F}$. By iteration we get that $T^{-k}A = A$ for all $k = 1, 2, ...$ In other words

$$A \in \bigcap_{k=1}^{\infty} \sigma(P_k, P_{k+1}, ...)$$

where the P_k are the coordinate functionals $P : M \to \mathbb{R}$. By the $0-1$-law of Kolmogorov we get that $\mu(A) \in \{0,1\}$.

Finally, we remark that the family $(f(T^k))_{k=0}^{\infty}$ forms an iid sequence of random variables having the distribution of f we were starting from.

Proof of Proposition 2.9.2 (see [6]). By normalization we can assume that $\mathbb{E}f = 0$. Let

$$\eta' := \liminf_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k) \le \limsup_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k) =: \eta.$$

The idea is to show that $0 \le \eta' \le \eta \le 0$ a.s. By the symmetry of the problem (replace f by $-f$) it is sufficient to show that

$$\eta \le 0 \quad a.s.$$

Let $\varepsilon > 0$ and

$$\begin{aligned} A_\varepsilon &:= \{\eta > \varepsilon\}, \\ f^* &:= (f-\varepsilon)\chi_{A_\varepsilon}, \\ S_n^* &:= f^* + \cdots + f^*(T^{n-1}), \\ M_n^* &:= \max\{0, f_1^*, ..., f_n^*\}. \end{aligned}$$

From Lemma 2.9.3 it follows that

$$\mathbb{E}f^*\chi_{M_n^*>0} \ge 0.$$

Moreover,

$$\{M_n^* > 0\} \uparrow_n \{\sup_k S_k^* > 0\} = \{\sup_k \frac{S_k^*}{k} > 0\} = \{\sup_k \frac{S_k}{k} > \varepsilon\} \cap A_\varepsilon = A_\varepsilon$$

because of $\sup_k \frac{S_k}{k} \geq \eta$. By Lebesgues dominated convergence,

$$0 \leq \mathbb{E} f^* \chi_{M_n^* > 0} \to_n \mathbb{E} f^* \chi_{A_\varepsilon} = \mathbb{E} f \chi_{A_\varepsilon} - \varepsilon \mathbb{P}(A_\varepsilon).$$

If J is the σ-algebra of T-invariant sets, we get that $A_\varepsilon \in J$ and

$$\mathbb{E} f \chi_{A_\varepsilon} = \mathbb{E}(\chi_{A_\varepsilon} \mathbb{E}(f|J)) = \mathbb{E}(\chi_{A_\varepsilon} 0)$$

as all sets from J have measure 0 or 1. Consequently,

$$\varepsilon \mathbb{P}(A_\varepsilon) \geq 0$$

for all $\varepsilon > 0$ and we are done. □

In the proof we used the following maximal ergodic theorem:

Lemma 2.9.3. *Let $T : \Omega \to \Omega$ be a measure preserving map and $f : \Omega \to \mathbb{R}$ be an integrable random variable. Let*

$$\begin{aligned} S_n &:= f + \cdots + f(T^{n-1}), \\ M_n &:= \max\{0, S_1, ..., S_n\}. \end{aligned}$$

Then $\mathbb{E}(f \chi_{\{M_n > 0\}}) \geq 0$.

Index

Bibliography

[1] H. Bauer. *Probability Theory*, Walter de Gruyter, 1996.

[2] L. Breiman. *Probability*, Addison-Wesley , 1968.

[3] R.M. Dudley. *Real Analysis and Probability*, Cambridge, 2002.

[4] W. Feller. *An introduction to probability theory and its applications*, Volumes I and II, Wiley, 1968 and 1966.

[5] C. Geiss and S. Geiss. *An introduction to probability theory.* Lecture Notes 60, Department of Mathematics and Statistics, University of Jyväskylä, 2009.

[6] A. Shirjaev. *Probability*, Springer, 1996.

[7] D. Williams. *Probability with Martingales*, Cambridge, 1991.