MACS 203b – Part I: Probability

Lecture Notes

P. Bianchi

December 10, 2021

Contents

Foreword

These notes are in two parts.
The first part provides the probabilistic background to establish the different types of convergences of sequences of random variables. The random variables are supposed to take values on some (finite dimensional) Euclidean space, although most of the results still hold in complete and separable metric spaces. In some cases, a proof might be more accessible when it is provided in the specific case of *real-valued* random variables. In such cases, we do not hesitate to provide this simpler proof for pedagogical reasons, and leave the higher dimensions to a Ms-course.
The second part uses the results of the first part in order to establish the asymptotic behavior of useful statistical estimators.

Notations

Throughout these notes, we use the following notations:

- $\mathcal{X} := \mathbb{R}^d$, where d is a positive integer ;
- $\mathcal{Y} := \mathbb{R}^m$, where m is a positive integer ;
- $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is the set of linear operators on $\mathcal{X} \to \mathcal{Y}$;
- $\mathscr{B}(\mathcal{X})$ is the Borel σ-field on $\mathcal{X}$;
- $\mathcal{P}(\mathcal{X})$ is the set of probability measures on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$;
- $C_b(\mathcal{X})$ (resp. $C_b^l(\mathcal{X})$) is the set of continuous (resp. Lipschitz continuous) functions on $\mathcal{X} \to \mathbb{R}$;
- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space ;
- A^c is the complementary set of a set A ;
- $\langle \, . \, , \, . \, \rangle$ and $\| \, . \, \|$ respectively represent the inner product and the norm, both in $\mathcal{X}$ and $\mathcal{Y}$ (the same notation is used for both spaces).

Chapter 1

Convergences of Random Variables

1.1 The Borel-Cantelli Lemma

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. An *event* is an element of $\mathcal{F}$. Let $(A_n : n \in \mathbb{N})$ be a sequence of events. We define

$$\bigcup_n A_n := \{\omega \in \Omega : \exists n, \omega \in A_n\}$$
$$\bigcap_n A_n := \{\omega \in \Omega : \forall n, \omega \in A_n\}$$

The sequence A_n is said *non-decreasing* if $A_n \subset A_{n+1}$ for every n, *non-increasing* if $A_{n+1} \subset A_n$ for every n.

Proposition 1.1. *If (A_n) is non-decreasing, then $\mathbb{P}(\bigcup_n A_n) = \lim_n \mathbb{P}(A_n)$.*
If (A_n) is non-increasing, then $\mathbb{P}(\bigcap_n A_n) = \lim_n \mathbb{P}(A_n)$.
If $\mathbb{P}(A_n) = 0$ for all n, then $\mathbb{P}(\bigcup_n A_n) = 0$.
If $\mathbb{P}(A_n) = 1$ for all n, then $\mathbb{P}(\bigcap_n A_n) = 1$.

We define

$$\limsup_n A_n := \bigcap_n \bigcup_{k \geq n} A_k \,.$$

A point ω lies in $\limsup_n A_n$ if and only if

$$\forall n, \ \exists k \geq n, \ \omega \in A_k \,.$$

In common language, this means that A_n is realized infinitely often.

Lemma 1.2 (Borel-Cantelli)**.** *If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_n A_n) = 0$.*

Proof. The sequence $B_n := \bigcup_{k \geq n} A_k$ is non-increasing. Thus $\lim \mathbb{P}(B_n) = \mathbb{P}(\bigcap_n B_n) = \mathbb{P}(\limsup_n A_n)$. As $\mathbb{P}(B_n) \leq \sum_{k \geq n} \mathbb{P}(A_k)$, the convergence of the series implies that $\mathbb{P}(B_n) \to 0$, hence the result. □

1.2 Definitions and Properties

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
Define $\mathcal{X} := \mathbb{R}^d$, where $d \geq 1$ is an integer. We denote by $\|\,.\,\|$ the natural Euclidean norm on $\mathcal{X}$. We endow $\mathcal{X}$ with its Borel σ-algebra $\mathscr{B}(\mathcal{X})$. If X is a mapping on $\Omega \to \mathcal{X}$, we denote by $X^{(1)}, \ldots, X^{(d)}$ its real components. Recall that X is a random variable if and only if $X^{(1)}, \ldots, X^{(d)}$ are random variables.

Let (X_n), X be r.v. on $\Omega \to \mathcal{X}$.

Definition 1.3.

- The sequence (X_n) converges *in probability* to X, noted $X_n \xrightarrow{P} X$, if

$$\forall \varepsilon > 0,\ \lim_n \mathbb{P}(\|X_n - X\| > \varepsilon) = 0\,.$$

- The sequence (X_n) converges *almost surely* (a.s.) to X, noted $X_n \xrightarrow{a.s.} X$, if $X_n \to X$ $\mathbb{P}$-almost everywhere (a.e.). Otherwise stated, if there exists $A \in \mathcal{F}$ such that (s.t.) $\mathbb{P}(A) = 1$, and s.t. for every $\omega \in A$, $X_n(\omega) \to X(\omega)$.

- Let $p > 0$. The sequence (X_n) converges *in* L_p to X, noted $X_n \xrightarrow{L_p} X$ if $\mathbb{E}(\|X_n\|^p) < \infty$, $\mathbb{E}(\|X\|^p) < \infty$, and if $\lim_n \mathbb{E}(\|X_n - X\|^p) = 0$.

Proposition 1.4. *If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{P} X$.*
If $X_n \xrightarrow{L_p} X$, then $X_n \xrightarrow{P} X$.

Proof. The first point follows from Lebesgue's dominated convergence theorem. The second point follows from Markov's inequality. □

Proposition 1.5. *If it holds that*

$$\forall \varepsilon > 0,\ \sum_n \mathbb{P}(\|X_n - X\| > \varepsilon) < \infty\,,$$

then $X_n \xrightarrow{a.s.} X$.

Proof. By Borel-Cantelli's lemma, $\mathbb{P}(\limsup_n \{\|X_n - X\| > \varepsilon\}) = 0$. Choosing $q \in \mathbb{N}^*$ and setting $\varepsilon = 1/q$, this yields

$$\mathbb{P}\left(\exists n,\, \forall k \geq n,\, \|X_n - X\| \leq 1/q\right) = 1\,,$$

and by taking the intersection of these events for all $q \in \mathbb{N}^*$, it follows that

$$\mathbb{P}\left(\forall q \in \mathbb{N}^*,\, \exists n,\, \forall k \geq n,\, \|X_n - X\| \leq 1/q\right) = 1\,,$$

which reads $\mathbb{P}(\lim_n \|X_n - X\| = 0) = 1$. □

Recall that a sequence $(v_n : n \in \mathbb{N})$ is said to be a *subsequence* of the sequence (u_n) if there exists a strictly increasing function $\varphi : \mathbb{N} \to \mathbb{N}$ s.t. for every $n \in \mathbb{N}$, $v_n = u_{\varphi_n}$.

Remark 1.1. In this course, we will often extract a subsequence from a subsequence. By the definition above, a subsequence of the subsequence (u_{φ_n}) is a sequence of the form $(u_{\varphi_{\psi_n}})$ where $\psi : \mathbb{N} \to \mathbb{N}$ is strictly increasing. This also writes $(u_{(\varphi\circ\psi)_n})$ where $\circ$ stands for the composition.

Proposition 1.6. *The following statements are equivalent:*

i. $X_n \xrightarrow{P} X$ *;*

ii. From any subsequence of (X_n)*, one can extract a further subsequence that converges to* X *almost surely.*

Proof. *i*$\Rightarrow$ *ii.* For every $\varepsilon > 0$, $\mathbb{P}(\|X_n - X\| > \varepsilon) \to 0$. Thus, for every n, there exists φ_n such that $\mathbb{P}(\|X_{\varphi_n} - X\| > \varepsilon) \le 2^{-n}$. This implies that $\sum_n \mathbb{P}(\|X_{\varphi_n} - X\| > \varepsilon) < \infty$. Therefore, $X_{\varphi_n} \xrightarrow{a.s.} X$ by Prop. 1.5. This proves the point *ii.*
ii $\Rightarrow$ *i.* By contradiction, assume that X_n does not converge in probability to X. Thus, there exists some $\varepsilon > 0$ and some subsequence (X_{φ_n}) such that for every n, $\mathbb{P}(\|X_{\varphi_n} - X\| > \varepsilon) > \varepsilon$. By the standing assumption, one can extract a further subsequence $(X_{(\phi\circ\psi)_n})$ which converges a.s. to X. The latter satisfies as well $\mathbb{P}(\|X_{(\phi\circ\psi)_n} - X\| > \varepsilon) > \varepsilon$. But the dominated convergence theorem implies that $\mathbb{P}(\|X_{(\phi\circ\psi)_n} - X\| > \varepsilon) \to 0$, which leads to a contradiction. □

Define $\mathcal{Y} := \mathbb{R}^m$, where m is a positive integer.

Theorem 1.7 (Composition by a continuous mapping)**.** *Let* $h : \mathcal{X} \to \mathcal{Y}$ *be a measurable map. Assume that there exists a set* $C \in \mathscr{B}(\mathcal{X})$ *s.t.* $\mathbb{P}(X \in C) = 1$ *and s.t.* h *is continuous at every point of* C*.*

i. If $X_n \xrightarrow{a.s.} X$*, then* $h \circ X_n \xrightarrow{a.s.} h \circ X$ *;*

ii. If $X_n \xrightarrow{P} X$*, then* $h \circ X_n \xrightarrow{P} h \circ X$*.*

Proof. *i.* Choose $A \in \mathcal{F}$ s.t. $\mathbb{P}(A) = 1$ and $X_n(\omega) \to X(\omega)$ for every $\omega \in A$. For every $\omega \in A \cap X^{-1}(C)$, h is continuous at point $X(\omega)$, hence $h(X_n(\omega)) \to h(X(\omega))$. As $\mathbb{P}(A \cap X^{-1}(C)) = 1$, the statement is proven.
ii. Consider an arbitrary subsequence $(h \circ X_{\varphi_n})$. As $X_n \xrightarrow{P} X$, Prop. 1.6 implies that there exists that one can extract a subsequence from (X_{φ_n}) which converges a.s. to zero. Denote by $(X_{(\varphi\circ\psi)_n})$ this subsequence: $X_{(\varphi\circ\psi)_n} \xrightarrow{a.s.} X$. Using point *i.*, $h \circ X_{(\varphi\circ\psi)_n} \xrightarrow{a.s.} h \circ X$. Hence, we have shown that from any subsequence $(h \circ X_{\varphi_n})$, one can extract a further subsequence which converges a.s. to zero. By Prop. 1.6, the conclusion follows. □

We shall always equipp the product space $\mathcal{X} \times \mathcal{Y}$ with its Borel σ-field $\mathscr{B}(\mathcal{X} \times \mathcal{Y}) = \mathscr{B}(\mathcal{X}) \otimes \mathscr{B}(\mathcal{Y})$.

Proposition 1.8. *Let (X_n), X be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Let (Y_n), Y be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Y}, \mathscr{B}(\mathcal{Y}))$. If $X_n \xrightarrow{a.s.} X$ and $Y_n \xrightarrow{a.s.} Y$, then $(X_n, Y_n) \xrightarrow{a.s.} (X, Y)$. The same holds when the almost sure convergence is replaced by the convergence in probability.*

Proof. The proof is immediate for the almost sure convergence, thus we only prove the statement for the convergence in probability. Assume that $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Consider a subsequence $(X_{\varphi_n}, Y_{\varphi_n})$. By Prop. 1.6, one can extract a subsequence, say $\varphi' := \varphi \circ \psi$ s.t. $X_{\varphi'_n} \xrightarrow{a.s.} X$, and a further subsequence say $\varphi'' := \varphi' \circ \kappa$ s.t. $Y_{\varphi''_n} \xrightarrow{a.s.} Y$. Along this subsequence, $(X_{\varphi''_n}, Y_{\varphi''_n}) \xrightarrow{a.s.} (X, Y)$. This shows $(X_n, Y_n) \xrightarrow{P} (X, Y)$ again by Prop. 1.6. $\square$

Corollary 1.9. *The sequence X_n converges a.s. to X if and only if for every $i = 1 \dots d$, $X_n^{(i)}$ converges a.s. to $X^{(i)}$. The same holds when the almost sure convergence is replaced by the convergence in probability.*

Proof. The proof is immediate for the almost sure convergence, thus we only prove the statement for the convergence in probability. Assume that $X_n \xrightarrow{P} X$. For every $i = 1 \dots d$, the projection map $(x^{(1)}, \dots, x^{(n)}) \mapsto x^{(i)}$ being continuous, it holds that $X_n^{(i)} \xrightarrow{P} X^{(i)}$ by Th. 1.7. Conversely assume that $X_n^{(i)} \xrightarrow{P} X^{(i)}$ for all i. By induction from Prop. 1.8, $(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{P} (X^{(1)}, \dots, X^{(k)})$, and the result follows. $\square$

1.3 Strong Law of Large Numbers (LLN)

The acronym iid stands for independent and identically distributed. If X is a r.v. on $\mathcal{X}$, the notation $\mathbb{E}(X)$ stands for the vector $(\mathbb{E}(X^{(1)}), \dots, \mathbb{E}(X^{(d)}))$, whenever it is well defined.

Theorem 1.10. *Let $(X_n : n \in \mathbb{N}^*)$ be an iid sequence on $\mathcal{X}$. Assume that $\mathbb{E}(\|X_1\|) < \infty$. Then, $n^{-1} \sum_{i=1}^n X_i \xrightarrow{a.s.} \mathbb{E}(X_1)$.*

Proof. By Prop. 1.9, it is sufficient to prove the result when $\mathcal{X} = \mathbb{R}$. It is even sufficient to prove the result in the case where $X_n \geq 0$ for every n. Indeed, if X_n is not necessarily positive, just use the decomposition $X_n = X_n^+ - X_n^-$ where $X_n^\pm = \max(\pm X_n, 0)$, apply the LLN to X_n^+ and X_n^- separately, and conclude using Th. 1.7. We thus assume from now on that $X_n \geq 0$. We set $S_n := \sum_{i=1}^n X_i$. We first provide the proof in the simpler case where $\mathrm{Var}(X_1) < \infty$, and then explain the extension to the general case.

Case 1: Var$(X_1) < \infty$. For all $\varepsilon > 0$, Markov's inequality implies that

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| > \varepsilon\right) \leq \frac{\mathrm{Var}(X_1)}{n\varepsilon^2}\,.$$

Choose $\alpha > 1$ and define for each n, $k_n := \lceil \alpha^n \rceil$, where $\lceil x \rceil$ represents the smallest integer greater than or equal to x. There exists a constant $c_\varepsilon < \infty$ s.t.

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \frac{S_{k_n} - \mathbb{E}S_{k_n}}{k_n} \right| > \varepsilon \right) \leq c_\varepsilon \sum_{n=1}^{\infty} k_n^{-1} < \infty \,.$$

By Prop. 1.5, $\frac{S_{k_n}}{k_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. Choose any $i \in \mathbb{N}^*$. Denote by $n(i)$ an integer s.t.

$$k_{n(i)} \leq i \leq k_{n(i)+1} \,.$$

Such an integer $n(i)$ exists because (k_n) is an increasing sequence. As $X_n \geq 0$ for all n, we deduce that $S_{k_{n(i)}} \leq S_i \leq S_{k_{n(i)}+1}$. Dividing by i,

$$\frac{S_{k_{n(i)}}}{k_{n(i)}} \frac{k_{n(i)}}{i} \leq \frac{S_i}{i} \leq \frac{S_{k_{n(i)+1}}}{k_{n(i)+1}} \frac{k_{n(i)+1}}{i} \,.$$

It is not difficult to show that

$$\frac{k_{n(i)}}{i} \geq \frac{1}{\alpha} \quad \text{and} \quad \frac{k_{n(i)+1}}{i} \leq \alpha + \frac{1}{i} \,.$$

Denote by A_α an event s.t. $\mathbb{P}(A_\alpha) = 1$ and for all $\omega \in A_\alpha$, $\frac{S_{k_n}(\omega)}{k_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. We obtain that for all $\omega \in A_\alpha$,

$$\frac{1}{\alpha}\mathbb{E}(X_1) \leq \liminf_{i \to \infty} \frac{S_i(\omega)}{i} \leq \limsup_{i \to \infty} \frac{S_i}{i} \leq \alpha \mathbb{E}(X_1) \,.$$

We now set α of the form $\alpha = 1 + \frac{1}{q}$ for $q \in \mathbb{N}^*$. For any such q, there exists a probability one event $A_{1+1/q}$ s.t. the above inequality hold on that event. The event $A := \bigcap_{q \in \mathbb{N}^*} A_{1+1/q}$ is s.t. $\mathbb{P}(A) = 1$, and for every $\omega \in A$,

$$\forall q \in \mathbb{N}^*, \quad \frac{1}{1+q^{-1}}\mathbb{E}(X_1) \leq \liminf_{i \to \infty} \frac{S_i(\omega)}{i} \leq \limsup_{i \to \infty} \frac{S_i(\omega)}{i} \leq (1+q^{-1})\mathbb{E}(X_1) \,.$$

Letting $q \to \infty$, we obtain that for all $\omega \in A$, $\lim_i \frac{S_i}{i} = \mathbb{E}(X_1)$.

Case 2: Var$(X_1) = +\infty$. Define $Y_n := X_n \mathbb{1}_{X_n < n}$ and $S_n^\star := \sum_{i \leq n} Y_i$. For any $\alpha > 1$, define k_n as above. For all $\varepsilon > 0$,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \frac{S_{k_n}^\star - \mathbb{E}S_{k_n}^\star}{k_n} \right| > \varepsilon \right) &\leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} \mathrm{Var}(Y_i) \\
&\leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \mathrm{Var}(Y_i) \sum_{n: k_n \geq i} \frac{1}{\alpha^{2n}} \\
&\leq \frac{1}{\varepsilon^2(1-\alpha^{-1})} \sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} \,.
\end{aligned}$$

Using that $\mathbb{E}(Y_i^2) = \sum_{k=0}^{i-1} \mathbb{E}(X_1^2 \mathbb{1}_{k \leq X_1 < k+1})$ and permuting the indices k and i, and setting $c := \frac{1}{\varepsilon^2(1-\alpha^{-1})}$, we obtain

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S^\star_{k_n} - \mathbb{E}S^\star_{k_n}}{k_n}\right| > \varepsilon\right) &\leq c \sum_{k=0}^{\infty} \mathbb{E}(X_1^2 \mathbb{1}_{k \leq X_1 < k+1}) \sum_{i>k} \frac{1}{i^2} \\ &\leq c \sum_{k=0}^{\infty} \mathbb{E}(X_1^2 \mathbb{1}_{k \leq X_1 < k+1}) \frac{1}{k+1} \\ &\leq c \sum_{k=0}^{\infty} \mathbb{E}(X_1 \mathbb{1}_{k \leq X_1 < k+1}) \ \leq c\mathbb{E}(X_1)\,.\end{aligned}$$

By Prop. 1.5, $\frac{S^\star_{k_n}}{k_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. Following the exact same proof as in Case 1, we deduce that $\frac{S^\star_n}{n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. It remains to prove that $\frac{S_n}{n}$ has the same a.s. limit as $\frac{S^\star_n}{n}$, and the proof will be complete. To that end, remark that $\mathbb{P}(X_n \neq Y_n) = \mathbb{P}(X_n > n) = \mathbb{P}(X_1 > n)$. By the same kind of derivations,

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(k+1 \geq X_1 > k) \\ &= \sum_{k=1}^{\infty} k\, \mathbb{P}(k+1 \geq X_1 > k) \\ &\leq \sum_{k=1}^{\infty} \mathbb{E}(X_1 \mathbb{1}_{k+1 \geq X_1 > k}) \ \leq \mathbb{E}(X_1) \ < \infty\,.\end{aligned}$$

By the Borel-Cantelli lemma, we obtain that $\mathbb{P}(\limsup_n \{X_n \neq Y_n\}) = 0$, which means that, $\mathbb{P}$-a.e., $X_n = Y_n$ for all n outside a finite set. Thus, $\mathbb{P}$-a.e., $\lim_n n^{-1} S_n = \lim_n n^{-1} S^\star_n$. This concludes the proof. □

Chapter 2

Narrow Convergence of Probability Measures

Let $(E, \mathcal{E})$, $(E', \mathcal{E}')$ be two measurable spaces. Let μ be a probability measure on $(E, \mathcal{E})$ and $f : E \to E'$ a measurable function. The *image measure* of f is the measure μf^{-1} defined on $(E', \mathcal{E}')$ and given by

$$\mu f^{-1}(A) = \mu(f^{-1}(A))$$

for every $A \in \mathcal{E}'$. In probabilistic language, it is also called the *distribution* or the *law* of the *random variable* f in the probability space $(E, \mathcal{E}, \mu)$. If $f : E \to \mathbb{R}$ is a measurable function, we often use the notation

$$\mu(f) = \int f d\mu$$

whenever the integral is well defined.

2.1 Definition

Denote by $\mathcal{P}(\mathcal{X})$ the set of probability measures on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$, where we recall that $\mathcal{X} := \mathbb{R}^d$. For every $\mu \in \mathcal{P}(\mathcal{X})$, we define the distribution function of μ by

$$F_\mu(x_1, \ldots, x_d) := \mu\left((-\infty, x_1] \times \cdots \times (-\infty, x_d]\right) ,$$

for every $(x_1, \ldots, x_d) \in \mathcal{X}$. We recall the fundamental property:

Theorem 2.1. *Disctinct measures cannot have the same distribution function.*

We say that x is a point of continuity of a function F if F is continuous at x.

Definition 2.2. Consider measures (μ_n), μ in $\mathcal{P}(\mathcal{X})$. We say that (μ_n) converges *narrowly* to μ, noted $\mu_n \Rightarrow \mu$, if

$$\lim_n F_{\mu_n}(x) = F_\mu(x) \tag{2.1}$$

for every x point of continuity of F_μ.

Remark 2.1. In most books, the term *weak convergence* is generally preferred to narrow convergence. But the terminology *weak convergence* is confusing for students having followed a course in functional analysis, because it has nothing to do with a convergence in a weak topology[1]: the term *weak* refers to the fact that the convergence (2.1) is restricted to the points of continuity. In these notes, we prefer the less common but non-ambiguous terminology of *narrow convergence.*

2.2 Properties

The following result will be revealed useful in order to prove important properties of the narrow convergence. This is why we state it as a lemma, although the result is known as the *Skorohod's representation theorem.*

Lemma 2.3 (Skorohod)**.** *Consider measures (μ_n), μ in $\mathcal{P}(\mathcal{X})$ s.t. $\mu_n \Rightarrow \mu$. Then, there exists a certain probability space, and there exist r.v. (Y_n), Y defined on that space into $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$ s.t. for every n, Y_n has distribution μ_n, Y has distribution μ and $Y_n \to Y$ pointwise.*

Before we provide the proof, the following definition will be useful. The *generalized inverse* of a mapping $F : \mathbb{R} \to \mathbb{R}$ is the function F^{-1} defined for every $x \in \mathbb{R}$ by

$$F^{-1}(x) := \inf\{y \in \mathbb{R} : x \leq F(y)\} . \tag{2.2}$$

Proof. We provide the proof in the case $d = 1$. When $d \geq 1$, the proof is more involved (and in fact, not easier than in the case where $\mathcal{X}$ is a general metric space). We refer to [1] for a general proof.
The considered probability space is $([0,1], \mathscr{B}([0,1]), \lambda)$, where $\mathscr{B}([0,1])$ is the Borel σ-algebra on the interval $[0,1]$ and λ is the Lebesgue measure on $[0,1]$ (it is a probability measure). We note for simplicity $F = F_\mu$ and $F_n = F_{\mu_n}$. We define $Y = F^{-1}$ and $Y_n = F_n^{-1}$. For every $\omega \in [0,1]$, $x \in \mathbb{R}$,

$$Y(\omega) \leq x \iff \omega \leq F(x) .$$

In the considered probability space, the distribution of the r.v. Y is λY^{-1}. The corresponding distribution function is $\lambda Y^{-1}((-\infty, x]) = \lambda(\{Y \leq x\}) = \lambda([0, F(x)]) = F(x)$. This means λY^{-1} has the same distribution function as μ, hence $\lambda Y^{-1} = \mu$. Thus, Y has distribution μ, and by the same arguments, Y_n has distribution μ_n. It remains to prove that for every $\omega \in [0,1]$, $Y_n(\omega) \to Y(\omega)$, which we leave as an exercise. □

For every $A \subset \mathcal{X}$, we respectively denote by $\overline{A}$ and $\mathring{A}$ the closure and the interior of A. We denote by $\partial A = \overline{A} \backslash \mathring{A}$ the boundary of A. It coincides with the set of points which can both be written as a limit of a sequence taking its values in A,

[1] For such students, who are able to interpret a probability measure as an element of the dual of bounded continuous functions on $\mathcal{X} \to \mathbb{R}$, the appropriate terminology would be *weak-$\star$ convergence*, as will be clear from Th. 2.4.

and also as a limit of an other sequence taking its values in the complementary set of A.
Recall that a mapping $f : \mathcal{X} \to \mathbb{R}$ is said *Lipschitz continuous* if there exists L s.t. $|f(x) - f(y)| \leq L\|x - y\|$ for every $x, y \in \mathcal{X}$, in which case L is called a Lipschitz constant of f. We denote by $C_b(\mathcal{X})$ the set of bounded continuous functions on $\mathcal{X} \to \mathbb{R}$. We denote by $C_b^l(\mathcal{X})$ the set of bounded Lipschitz continuous functions on $\mathcal{X} \to \mathbb{R}$.
The following result is a (short version of) the so-called *Portmanteau* theorem.

Theorem 2.4. *The following statements are equivalent.*

i. $\mu_n \Rightarrow \mu$ *;*

ii. *For every* $f \in C_b(\mathcal{X})$, $\mu_n(f) \to \mu(f)$ *;*

iii. *For every* $f \in C_b^l(\mathcal{X})$, $\mu_n(f) \to \mu(f)$ *;*

iv. *For every* $A \in \mathscr{B}(\mathcal{X})$ *s.t.* $\mu(\partial A) = 0$, $\mu_n(A) \to \mu(A)$.

Proof. *i* $\Rightarrow$ *ii.* By Lemma 2.3, there exists r.v. Y, (Y_n) defined on some probability space $(E, \mathcal{E}, \mathsf{P})$ with distribution μ and (μ_n) respectively, s.t. $Y_n \to Y$ pointwise. Choose $f \in C_b(\mathcal{X})$. By continuity of f, we obtain $f \circ Y_n \to f \circ Y$ pointwise. As f is bounded, the dominated convergence theorem implies that $\mathsf{P}(f \circ Y_n) \to \mathsf{P}(f \circ Y)$. This reads equivalently $\mathsf{P}Y_n^{-1}(f) \to \mathsf{P}Y^{-1}(f)$. By definition, $\mathsf{P}Y_n^{-1} = \mu_n$ and $\mathsf{P}Y^{-1} = \mu$, hence the result.

ii $\Rightarrow$ *iii.* Evident.

iii $\Rightarrow$ *i.* We provide the proof in the case $d = 1$ (try to generalize it to $d \geq 1$). Choose $\varepsilon > 0$. Choose x a point of continuity of F_μ. Introduce the function $f : \mathbb{R} \to [0, 1]$ as $f(t) = 1$ if $t \leq x$, $f(t) = 1$ if $t \leq x$, $f(t) = (x + \varepsilon - t)/\varepsilon$ if $x < t < x + \varepsilon$ and $f(t) = 0$ otherwise. The function f is Lipschitz continuous. Hence, $\mu_n(f) \leq \mu(f)$. On the otherhand $F_{\mu_n}(x) \leq \mu_n(f)$ and $\mu(f) \leq F_\mu(x+\varepsilon)$. Therefore, $\limsup_n F_{\mu_n}(x) \leq F_\mu(x + \varepsilon)$. Letting $\varepsilon \downarrow 0$ and using the right-continuity of F_μ, it follows that $\limsup_n F_{\mu_n}(x) \leq F_\mu(x)$. The reverse inequality $\liminf_n F_{\mu_n}(x) \geq F_\mu(x)$ is proven in the same manner, using moreover the fact that x is a point of continuity of F_μ.

i $\Rightarrow$ *iv.* The proof is nearly a copy-paste of *i* $\Rightarrow$ *ii.* Do it as an exercise.

iv $\Rightarrow$ *i.* We provide the proof in the case $d = 1$. Choose x a point of continuity of F_μ. It satisfies $\mu(\{x\}) = 0$. Set $A = (-\infty, x]$. We have $\partial A = \{x\}$. By the standing assumption, $\mu_n(A) \to \mu(A)$, which reads $F_{\mu_n}(x) \to F_\mu(x)$. □

Theorem 2.5 (Composition by a continuous mapping)**.** *Let* $m \in \mathbb{N}^*$ *and let* $h : \mathcal{X} \to \mathcal{Y}$ *be a measurable map. Assume that there exists a set* $C \in \mathscr{B}(\mathcal{X})$ *s.t.* $\mu(C) = 1$ *and s.t.* h *is continuous at every point of* C*.*
If $\mu_n \Rightarrow \mu$*, then* $\mu_n h^{-1} \Rightarrow \mu h^{-1}$*.*

Proof. By Lemma 2.3, there exists r.v. Y, (Y_n) defined on some probability space $(E, \mathcal{E}, \mathsf{P})$ with distribution μ and (μ_n) respectively, s.t. $Y_n \to Y$ pointwise.

As $\mathsf{P}Y^{-1} = \mu$, note that $\mathsf{P}(Y^{-1}(C)) = \mu(C) = 1$. For every $\omega \in Y^{-1}(C)$, it holds that h is continuous at $Y(\omega)$. Since $Y_n(\omega) \to Y(\omega)$, it follows that $h(Y_n(\omega)) \to h(Y(\omega))$. We have shown that, on the space $(E, \mathcal{E}, \mathsf{P})$,

$$h \circ Y_n \xrightarrow{a.s.} h \circ Y \,.$$

Choose $f \in C_b(\mathcal{Y})$. Note that

$$\mu_n h^{-1}(f) = \mu_n(f \circ h) = \mathsf{P}Y_n^{-1}(f \circ h) = \mathsf{P}(f \circ h \circ Y_n)\,.$$

By Prop. 1.7,

$$f \circ h \circ Y_n \xrightarrow{a.s.} f \circ h \circ Y \,.$$

Therefore, by the dominated convergence theorem, $\mathsf{P}(f \circ h \circ Y_n) \to \mathsf{P}(f \circ h \circ Y)$. This reads $\mu_n h^{-1}(f) \to \mu h^{-1}(f)$. The proof is complete by Th. 2.4. □

2.3 Tightness and Prokhorov's Theorem

Denote by $\mathscr{F}$ the set of non-decreasing right-continuous functions on $\mathbb{R} \to [0,1]$.

Lemma 2.6 (Helly). *For every sequence (F_n) on $\mathscr{F}$, there exists $F \in \mathscr{F}$ and a subsequence (F_{φ_n}) such that*

$$F_{\varphi_n}(x) \to F(x)$$

for every x that is a point of continuity of F.

Proof. The set $\mathbb{Q}$ of rational numbers being denumerable, we write it as $\mathbb{Q} = \{x_0, x_1, \dots\}$. The sequence $(F_n(x_0))$ being bounded, there exists a value in $[0,1]$, which we call $G(x_0)$, and a strictly increasing map $\psi^0 : \mathbb{N} \to \mathbb{N}$ s.t.

$$F_{\psi^0_n}(x_0) \to G(x_0)\,.$$

The sequence $(F_{\psi^0_n}(x_1))$ being bounded, there exists a value in $[0,1]$, which we call $G(x_1)$, and a strictly increasing map $\psi^1 : \mathbb{N} \to \mathbb{N}$ s.t.

$$F_{(\psi^0 \circ \psi^1)_n}(x_1) \to G(x_1)\,.$$

Continuing the process, be can recursively construct a sequence (ψ^k) of mappings on $\mathbb{N} \to \mathbb{N}$, and a sequence $(G(x_k))$ on $[0,1]$, such that for every k,

$$F_{(\psi^0 \circ \cdots \circ \psi^k)_n}(x_k) \to G(x_k)\,. \tag{2.3}$$

For every n, define $\varphi_n := (\psi^0 \circ \cdots \circ \psi^n)_n$. The mapping $\varphi : \mathbb{N} \to \mathbb{N}$ is stricly increasing. Moreover, for every $k \in \mathbb{N}$ and every $n \geq k$, note that

$$\varphi_n = (\psi^0 \circ \cdots \circ \psi^k)_{k_n}$$

where $k_n := (\psi^{k+1} \circ \cdots \circ \psi^n)_n$ tends to ∞ as $n \to \infty$ (to prove this, show that $k_n \geq n$). Thus for every $n \geq k$, $F_{\varphi_n}(x_k) = F_{(\psi^0 \circ \cdots \circ \psi^k)_{k_n}}(x_k)$ which proves that

$$\forall k \in \mathbb{N},\ F_{\varphi_n}(x_k) \to G(x_k)\,.$$

For every $x \in \mathbb{R}$, define

$$F(x) := \inf\{G(x_k)\,:\, k \in \mathbb{N},\, x_k > x\}\,.$$

Obviously, F is non-decreasing. We prove that it is right-continuous. Set $x \in \mathbb{R}$ and $\varepsilon > 0$. By definition of the infimum, there exists a rational number $x_k > x$ s.t. $G(x_k) < F(x) + \varepsilon$. Consider any point $y \in [x, x_k)$. It holds that $F(y) \leq G(x_k) < F(x) + \varepsilon$. Moreover, $F(x) \leq F(y)$ because F is non-decreasing. We have shown that for every $\varepsilon > 0$, there exists $\delta > 0$ (namely, $\delta = x - x_k$) s.t.

$$\forall y \in [x, x+\delta), \quad |F(y) - F(x)| \leq \varepsilon\,.$$

This proves that F is right-continuous.
Now, we must prove that $F_n(x) \to F(x)$ as $n \to \infty$. As above, chose $\varepsilon > 0$ and a rational number $x_k > x$ s.t. $G(x_k) < F(x) + \varepsilon$. By (2.3), this also reads $\lim_n F_{\varphi_n}(x_k) < F(x) + \varepsilon$. Since F_{φ_n} is non-decreasing, $F_{\varphi_n}(x) \leq F_{\varphi_n}(x_k)$, thus,

$$\limsup_{n\to\infty} F_{\varphi_n}(x) < F(x) + \varepsilon\,. \tag{2.4}$$

On the otherhand, choose any rational number x_ℓ s.t. $x - \varepsilon < x_\ell < x$. By the definition of F as an infimum, $F(x-\varepsilon) \leq G(x_\ell)$. By (2.3), this also reads $F(x-\varepsilon) \leq \lim_n F_{\varphi_n}(x_\ell)$. Since F_{φ_n} is non-decreasing, $F_{\varphi_n}(x_\ell) \leq F_{\varphi_n}(x)$, thus,

$$F(x-\varepsilon) \leq \liminf_{n\to\infty} F_{\varphi_n}(x)\,. \tag{2.5}$$

Putting together (2.4) and (2.5) and letting ε tend to zero, we obtain that $\lim_n F_{\varphi_n}(x) = F(x)$ whenever x is a point of continuity of F. □

Recall the notation A^c for the complementary set of a set A.

Definition 2.7. A family of probability measures $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ is called *tight* if for every $\varepsilon > 0$, there exists a compact set $K \subset \mathcal{X}$ s.t.

$$\sup_{\mu\in\mathcal{M}} \mu(K^c) \leq \varepsilon\,.$$

Proposition 2.8. *The following properties hold true.*

i. Any finite family of probability measures on $\mathcal{X}$ is tight.

ii. A sequence (μ_n) on $\mathcal{P}(\mathcal{X})$ is tight if and only if for every $\varepsilon > 0$, there exists a compact set $K \subset \mathcal{X}$ s.t.

$$\limsup_n \mu_n(K^c) \leq \varepsilon\,.$$

Proof. i. Consider $\mathcal{M}$ of the form $\{\mu_1, \dots, \mu_k\}$. Choose $\varepsilon > 0$. Define $B_r = \{x : \|x\| \le r\}$. Since $\bigcup_{r\in\mathbb{N}} B_r = \mathcal{X}$, Prop. 1.1 implies that for every $i = 1 \dots k$, there exists $r_i \in \mathbb{N}$ s.t. $\mu_i(B_{r_i}) > 1 - \varepsilon$. Set $R = \max\{r_i\}$. The set B_R is compact and satisfies $\mu_i(B_R) > 1 - \varepsilon$ for all i. Hence, $\{\mu_1, \dots, \mu_k\}$ is tight.
ii. Choose $\varepsilon > 0$ and K s.t. $\limsup_n \mu_n(K^c) \le \frac{\varepsilon}{2}$. There exists n_0 s.t. for all $n > n_0$, $\mu_n(K^c) \le \varepsilon$. By point *i.*, the family $(\mu_n : n \le n_0)$ is tight, thus there exists a compact set $\tilde{K}$ s.t. $\mu_n(\tilde{K}^c) \le \varepsilon$ for all $n \le n_0$. Thus, for all $n \in \mathbb{N}$, $\mu_n((K \cup \tilde{K})^c) \le \varepsilon$, which shows that $(\mu_n : n \in \mathbb{N})$ is tight. □

Proposition 2.9. *Any sequence on $\mathcal{P}(\mathcal{X})$ which converges narrowly is tight.*

Proof. Consider a sequence on $\mathcal{P}(\mathcal{X})$ s.t. $\mu_n \Rightarrow \mu$. Define $N(x) := \|x\|$ for every $x \in \mathcal{X}$. Set $\varepsilon > 0$. By Prop. 2.8-i, any single probability measure is tight, thus, there exists $K > 0$ s.t. $\mu N^{-1}([-K, K]^c) > 1 - \varepsilon$. By Th. 2.5, $\mu_n N^{-1} \Rightarrow \mu N^{-1}$. Choose $K' > K$ s.t. $\mu N^{-1}(\{K'\}) = 0$. By Th. 2.4-iv, $\mu_n N^{-1}([-K', K']) \to \mu N^{-1}([-K', K'])$. Thus $\liminf_n \mu_n(N^{-1}([-K', K'])) \ge 1 - \varepsilon$, or equivalently, $\limsup_n \mu_n(N^{-1}([-K', K'])^c) \le \varepsilon$. Hence, (μ_n) is tight by Prop. 2.8-ii. □

Theorem 2.10 (Prokhorov)**.** *A sequence of probability measures on $\mathcal{X}$ is tight if and only if every subsequence admits a further subsequence which converges narrowly.*

Proof. We prove the direct implication. The proof is provided in the case $d = 1$. Denote by F_n the distribution function of μ_n. By Helly's lemma, there exists $F \in \mathscr{F}$ and a subsequence (F_{ψ_n}) s.t. $F_{\psi_n}(x) \to F(x)$ at every point of continuity x of F. If moreover F can be shown to be a distribution function *i.e.* if it satisfies $\lim_{x\to+\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, then, (μ_{ψ_n}) converges narrowly to the measure $\mu := \lambda F^{-1}$, where λ is the lebesgue measure on $[0, 1]$ and where F^{-1} is the generalized inverse defined by Eq. (2.2).
Set $\varepsilon > 0$. As (μ_n) is tight, there exists a compact set, which can be chosen of the form $[-K, K]$ s.t. $\mu_n([-K, K]) > 1 - \varepsilon$. This implies that for every $t > K$, $F_n(t) - F_n(-t) > 1 - \varepsilon$. In particular $F_n(t) > 1 - \varepsilon$. Choosing t as a point of continuity of F, and letting $n \to \infty$ along the subsequence ψ_n, it follows that $F(t) > 1 - \varepsilon$. As F is non decreasing, $\lim_{x\to+\infty} F(x) > 1 - \varepsilon$. By letting $\varepsilon \downarrow 0$, we conclude that $\lim_{x\to+\infty} F(x) = 1$. Finally, the inegality $F_n(-t) < \varepsilon$ leads to $\lim_{x\to-\infty} F(x) = 0$ by the same type of arguments.
Thus, we have shown that if (μ_n) is tight, it admits a subsequence which converges narrowly. To prove the conclusion, consider an arbitrary subsequence (μ_{φ_n}). As (μ_n) is tight, (μ_{φ_n}) is tight as well, and thus admits a further subsequence which converges narrowly.

We prove the converse of the theorem. Assume that every subsequence has a further subsequence which converges narrowly. Assume, by contradiction, that (μ_n) is not tight. Then, there exists $\varepsilon > 0$ s.t. for every compact set $K \subset \mathcal{X}$, $\sup_n \mu_n(K^c) > \varepsilon$. Consider the sequence of compact sets $(B_k : k \in \mathbb{N})$ given by the closed balls $B_k = \{x \in \mathcal{X} : \|x\| \le K\}$. For every k, there exists $n \in \mathbb{N}$ s.t. $\mu_n(B_k^c) > \varepsilon$. Thus, one can construct a strictly increasing $\varphi : \mathbb{N} \to \mathbb{N}$ s.t. for all k, $\mu_{\varphi_k}(B_k^c) > \varepsilon$. By the standing hypothesis, one can extract from (μ_{φ_k}) a further subsequence which converges narrowly. Let us denote this subsequence by (μ_{ψ_k}). By Prop. 2.9, the latter sequence is tight. There exists a compact set K s.t. $\mu_{\psi_k}(K^c) < \varepsilon$ for every k. Choose a particular value of k in such a way that $K \subset B_k$. One has $\mu_{\psi_k}(B_k^c) < \varepsilon$, hence a contradiction. □

Remark 2.2. The students who are already familiar with functional analysis might remember that an arbitrary subset of a metric space has a compact closure if and only if every sequence in this set admits a converging subsequence. Reading Prokhorov's theorem, it is tempting to interpret tightness as a compactness criterion on $\mathcal{P}(\mathcal{X})$. In order to make this claim more precise, one should endow $\mathcal{P}(\mathcal{X})$ with a metric (or at least a topology) that is compatible with the narrow convergence. Interpreting a probability measure as a linear functional on $C_b(\mathcal{X})$, we may write $\mathcal{P}(\mathcal{X}) \subset C_b(\mathcal{X})^\star$ where $C_b(\mathcal{X})^\star$ is the dual of $C_b(\mathcal{X})$ equipped with the uniform norm. By Th. 2.4-ii, the topology of narrow convergence coincides with the trace of the weak-$\star$ topology on $\mathcal{P}(\mathcal{X})$. As a matter of fact, the space $\mathcal{P}(\mathcal{X})$ endowed with the weak-$\star$ topology can be shown to be metrizable, by the so-called Lévy-Prokhorov metric. Therefore, Prokhorov's theorem can be restated as follows: *A subset $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ has a compact closure if and only if it is tight.* We refer to [2] for more details.

Proposition 2.11. *Let (μ_n) be a tight sequence on $\mathcal{P}(\mathcal{X})$ and let $\mu \in \mathcal{P}(\mathcal{X})$. Assume that every narrowly convergent subsequence of (μ_n) converges narrowly to μ. Then, $\mu_n \Rightarrow \mu$.*

Proof. In class. □

2.4 Characteristic Functions, Lévy's Theorem

The *characteristic function* of a probability measure $\mu \in \mathcal{P}(\mathcal{X})$ is the mapping $\phi_\mu : \mathcal{X} \to \mathbb{C}$ defined for every $t \in \mathcal{X}$ by

$$\phi_\mu(t) := \int e^{\imath\langle t,x\rangle} d\mu(x)\,,$$

where $\langle\, ,\, \rangle$ stands for the inner production on $\mathcal{X}$ and where $\imath := \sqrt{-1}$. The characteristic function is well defined, and is continuous by simple application of Lebesgue's dominated convergence theorem.

Theorem 2.12. *Distinct measures cannot have the same characteristic function.*

Proof. As a preliminary, consider the function $S : x \mapsto \int_0^x \frac{\sin u}{u} du$ on $\mathbb{R}_+ \to \mathbb{R}$, It is not difficult to prove that $S(x)$ admits a (positive) limit as $x \to +\infty$, because $\int_{(n-1)\pi}^{n\pi} u^{-1} \sin u \, du$ alternates in sign and its absolute value decreases to zero. You might remember that the limit is equal to $\frac{\pi}{2}$, but since we won't need this, just call the limit $\frac{\pi_0}{2}$ (in fact $\pi_0 = \pi$).

Consider the case $\mathcal{X} = \mathbb{R}$. Let $\mu \in \mathcal{P}(\mathbb{R})$. The proof consists in proving the following inversion formula:

$$\mu((a,b]) = \lim_{T\to\infty} \frac{1}{2\pi_0} \int_{-T}^{T} \frac{e^{-\imath ta} - e^{-\imath tb}}{\imath t} \phi_\mu(t) dt\,, \tag{2.6}$$

for every $a < b$ satisfying $\mu(\{a\}) = \mu(\{b\}) = 0$. Indeed, if Eq. (2.6) holds, then by letting $a \to -\infty$, we obtain that the distribution F_μ is uniquely determined by ϕ_μ, hence the conclusion.
We now prove Eq. (2.6). Denote by I_T the quantity inside the limit in (2.6). By Fubini's theorem,

$$I_T = \frac{1}{2\pi_0} \int \left(\int_{-T}^{T} \frac{e^{\imath t(x-a)} - e^{\imath t(x-b)}}{\imath t} dt \right) d\mu(x)\,.$$

Note that $\int_{-T}^{T} \frac{e^{\imath t(x-a)}}{\imath t} dt = 2\int_0^T \frac{\sin((x-a)t)}{t} dt = 2\text{sign}(x-a) S(T|x-a|)$ where $\text{sign}(u)$ is equal to 1 if $u > 0$, -1 if $u < 0$ and zero otherwise. Therefore,

$$I_T = \frac{1}{\pi_0} \int \left(\text{sign}(x-a)S(T|x-a|) - \text{sign}(x-b)S(T|x-b|)\right) d\mu(x)\,.$$

Using that $\lim_{u\to\infty} S(u) = \frac{\pi_0}{2}$, the integrand is bounded, and converges as $T \to \infty$ to the function

$$\psi_{a,b}(x) = \begin{cases} 0 & \text{if } x < a, \\ 0.5 & \text{if } x = a, \\ 1 & \text{if } a < x < b \\ 0.5 & \text{if } x = b, \\ 0 & \text{if } x > b. \end{cases}$$

By the dominated convergence theorem, Eq. (2.6) holds, as $\mu(\{a\}) = \mu(\{b\}) = 0$. When $\mathcal{X} = \mathbb{R}^d$ with $d \geq 2$, an inversion formula very similar to Eq. (2.6) can be established, and the proof uses the same arguments, see [1, Eq. (29.3)] □

Proposition 2.13. *Consider $\mu \in \mathcal{P}(\mathbb{R})$ and $p \in \mathbb{N}$ s.t. $\int |x|^p d\mu(x) < \infty$. Then, ϕ_μ is p-times continuously differentiable and for every $t \in \mathbb{R}$, its p-th order derivative satisfies $\phi_\mu^{(p)}(t) = \int \imath^p x^p e^{itx} d\mu(x)$.*

Proof. The statement was proven in the first year probability course. The proof is by induction and uses Lebesgue's dominated convergence theorem. □

We recall that the standard Gaussian distribution $\mathcal{N}(0,1)$ is the probability measure on $\mathbb{R}$ having the probability density function $x \mapsto (2\pi)^{-1/2} e^{-x^2/2}$.

Proposition 2.14. *For every $t \in \mathbb{R}$, $\phi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}}$.*

Proof. Denote by f the $\mathcal{N}(0,1)$-density function and note $\phi := \phi_{\mathcal{N}(0,1)}$. By Prop. 2.13, ϕ is continuously differentiable and $\phi'(t) = \int \imath x e^{itx} f(x) dx$. Noting that $f'(x) = -xf(x)$ and integrating by parts, we obtain $\phi'(t) = -t\phi(t)$. The result follows by solving the differential equation and using $\phi(0) = 1$. □

Theorem 2.15 (Lévy)**.** *Let $(\mu_n), \mu \in \mathcal{P}(\mathcal{X})$. The following properties are equivalent:*

i. $\mu_n \Rightarrow \mu$ *;*

ii. For every $t \in \mathcal{X}$*,* $\phi_{\mu_n}(t) \to \phi_\mu(t)$*.*

Proof. *i* $\Rightarrow$ *ii.* The mapping $x \mapsto e^{i\langle t,x\rangle}$ is bounded and continuous. The conclusion therefore follows from Th. 2.4.
ii $\Rightarrow$ *i.* Assume that (μ_n) is tight (this point is proven below). By Th. 2.10, there exists a subsequence (μ_{φ_n}) and a probability measure ν s.t. $\mu_{\varphi_n} \Rightarrow \nu$. By the first point just proven, $\phi_{\varphi_n}(t) \to \phi_\nu(t)$ for every t. By identification of the limit, $\phi_\nu(t) = \phi_\mu(t)$, hence $\mu = \nu$ by Th. 2.12. As a conclusion, (μ_n) is tight and every subsequence converging narrowly converges to μ. This shows that $\mu_n \Rightarrow \mu$ by Prop. 2.11.
Thus, the crux of the proof consists in showing that (μ_n) is tight. We provide the proof in the case $d = 1$. One has for every $a > 0$, by Fubini's theorem

$$\begin{aligned}\frac{1}{a}\int_{-a}^{a}(1-\phi_{\mu_n}(t))dt &= \int\left(\frac{1}{a}\int_{-a}^{a}(1-e^{itx})dt\right)d\mu_n(x)\\ &= 2\int\left(1-\frac{\sin ax}{ax}\right)d\mu_n(x)\,,\end{aligned}$$

and since the integrand is positive, we have for every measurable set $C \subset \mathbb{R}$,

$$\begin{aligned}\frac{1}{a}\int_{-a}^{a}(1-\phi_{\mu_n}(t))dt &\geq 2\int_C\left(1-\frac{\sin ax}{ax}\right)d\mu_n(x)\\ &\geq \int_C 2\left(1-\frac{1}{ax}\right)d\mu_n(x)\,.\end{aligned}$$

Now setting in particular $C := \mathbb{R}\backslash[-\frac{2}{a}, \frac{2}{a}]$, we obtain:

$$\frac{1}{a}\int_{-a}^{a}(1-\phi_{\mu_n}(t))dt \geq \mu_n(\mathbb{R}\backslash[-\tfrac{2}{a}, \tfrac{2}{a}])\,.$$

Consider the lefthand side. By the dominated convergence theorem, it converges to $\frac{1}{a}\int_{-a}^{a}(1-\phi_\mu(t))dt$. Now consider the latter limit. As ϕ_μ is continuous at zero, it is an easy exercise to show that $\frac{1}{a}\int_{-a}^{a}(1-\phi_\mu(t))dt$ converges to zero as $a \to \infty$. For any $\varepsilon > 0$, choose $a > 0$ s.t. $\frac{1}{a}\int_{-a}^{a}(1-\phi_\mu(t))dt < \varepsilon$. Since $\frac{1}{a}\int_{-a}^{a}(1-\phi_{\mu_n}(t))dt \to \frac{1}{a}\int_{-a}^{a}(1-\phi_\mu(t))dt$, we obtain: $\limsup_n \mu_n(\mathbb{R}\backslash[-\frac{2}{a}, \frac{2}{a}]) < \varepsilon$, which concludes the proof by Prop. 2.8-ii. □

Remark 2.3. Examining the proof of Th. 2.15, the key argument is to show that condition *ii* implies the tightness of (μ_n). To prove this, we did not explicitely used the fact that the limit ϕ_μ is a characteristic function: we just needed the continuity of ϕ_μ at zero. Thus, one can state a more general formulation of Lévy's theorem, which you should be able to prove by your own: *If there exists a function* $\psi : \mathcal{X} \to \mathbb{C}$ *continuous at zero s.t.* $\phi_{\mu_n}(t) \to \psi(t)$ *for all* t*, then there exists* $\mu \in \mathcal{P}(\mathcal{X})$ *s.t.* $\phi_\mu = \psi$ *and* $\mu_n \Rightarrow \mu$*.*

2.5 Random Variables, Convergence in Law

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let (X_n), X be r.v. on $\Omega \to \mathcal{X}$.
We recall that $\mathbb{P}X^{-1}$ is the law of X. If $f : \mathcal{X} \to \mathbb{R}$ is a measurable mapping, we recall that the notation $f(X)$ means $f \circ X$. The expectation of $f(X)$ is the quantity

$$\mathbb{E}(f(X)) := \mathbb{P}(f \circ X) \ = \mathbb{P}X^{-1}(f)$$

whenever it is well defined. The *distribution function* of X, noted F_X, is defined as the distribution function of its law, that is $F_X := F_{\mathbb{P}X^{-1}}$. The *characteristic function* of X, noted ϕ_X, is defined as the characteristic function of its law, that is $\phi_X := \phi_{\mathbb{P}X^{-1}}$. With this notation, we can write

$$\phi_X(t) = \mathbb{E}(e^{\imath \langle t, X \rangle}) .$$

We recall that $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ represents the set of linear operators on $\mathcal{X} \to \mathcal{Y}$.

Lemma 2.16. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ and $b \in \mathcal{Y}$. Let X be a r.v. on $\mathcal{X}$ and define $Y := AX + b$. Then, for every $t \in \mathcal{Y}$, $\phi_Y(t) = e^{\imath \langle t, b \rangle} \phi_X(A^* t)$ where A^* is the adjoint of A.*

Proof. Immediate from the definition. □

Definition 2.17. The sequence (X_n) is called *tight*, or *bounded in probability*, if $(\mathbb{P}X_n^{-1})$ is tight.

This means that for every $\varepsilon > 0$, there exists a compact set $\mathcal{K} \subset \mathcal{X}$ such that $\mathbb{P}(X_n \notin K) < \varepsilon$ for all n.

Definition 2.18. The sequence (X_n) converges *in law* to X, noted $X_n \Rightarrow X$, if $\mathbb{P}X_n^{-1} \Rightarrow \mathbb{P}X$.

Remark 2.4. The convergence in law is the narrow convergence of the laws of the random variables. Therefore, only their distributions matter. Actually, all variables (X_n) and X could have been defined on different probability spaces. The introduction of a unique probability space is mainly for notational convenience.

Remark 2.5. Because only the distribution of X matters in the definition (and not X itself), we will also use the following notation. If (X_n) is a sequence of r.v. on $\mathcal{X}$ and if $\mu \in \mathcal{P}(\mathcal{X})$, we will write $X_n \Rightarrow \mu$ as an equivalent to $\mathbb{P}X_n^{-1} \Rightarrow \mu$. Thus, if X_n, X have the distribution μ_n, μ respectively, the notations $X_n \Rightarrow X$, $\mu_n \Rightarrow \mu$ and $X_n \Rightarrow \mu$ mean the same fact.

Remark 2.6. Convergence in law is also called convergence in distribution. Notations $X_n \xrightarrow{\mathcal{L}} X$ and $X_n \xrightarrow{\mathcal{D}} X$ are often used with the same meaning as $X_n \Rightarrow X$.

By Th. 2.4, $X_n \Rightarrow X$ if and only if for all $f \in C_b(\mathcal{X})$, $\mathbb{P}X_n^{-1}(f) \to \mathbb{P}X(f)$, which reads equivalently $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$. By Th. 2.15, this is also equivalent to the convergence of characteristic functions. Therefore, the following corollary is a simple reformulation fo Th. 2.4 and Th. 2.15 along with the definition of narrow convergence.

Corollary 2.19. *The following points are equivalent:*

i. $X_n \Rightarrow X$;

ii. $F_{X_n}(x) \to F_X(x)$ *at every* x *point of continuity of* F_X ;

iii. $\forall f \in C_b(\mathcal{X}),\ \mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$;

iv. $\forall f \in C_b^l(\mathcal{X}),\ \mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$;

v. $\forall A \in \mathscr{B}(\mathcal{X})$ *s.t.* $\mathbb{P}(X \in \partial A) = 0,\ \mathbb{P}(X_n \in A) \to \mathbb{P}(X \in A)$;

vi. $\forall t \in \mathcal{X},\ \phi_{X_n}(t) \to \phi_X(t)$.

Proposition 2.20. *If* $X_n \xrightarrow{P} X$*, then* $X_n \Rightarrow X$*.*

Proof. Let $f \in C_b(\mathcal{X})$. By Th. 1.7, $f(X_n) \xrightarrow{P} f(X)$. The sequence $(\mathbb{E}(f(X_n)))_n$ is bounded. Let ℓ be an accumulation point, that is, $\mathbb{E}(f(X_n)) \to \ell$ along some subsequence. By Prop. 1.6, one can extract a further subsequence such that $f(X_n) \xrightarrow{a.s.} f(X)$ along that subsequence. By the dominated convergence theorem, $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ along that subsequence. Therefore, $\ell = \mathbb{E}(f(X))$. This proves that $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$. The conclusion follows from Cor. 2.19. □

The following lemma shows that the convergence in law of a sequence of random *vectors* can, in a certain sense, be reduced to convergence in law of *real-valued* r.v.: this can be convenient in some proofs.

Lemma 2.21 (Cramer-Wold)**.** *The following points are equivalent:*

i. $X_n \Rightarrow X$;

ii. $\forall t \in \mathcal{X},\ \langle t, X_n \rangle \Rightarrow \langle t, X \rangle$.

Proof. By Cor. 2.19-iv, $X_n \Rightarrow X$ if and only if $\mathbb{E}(e^{\imath\langle t, X_n\rangle}) \to \mathbb{E}(e^{\imath\langle t, X\rangle})$ for all $t \in \mathcal{X}$. This is equivalent to $\mathbb{E}(e^{\imath u\langle t, X_n\rangle}) \to \mathbb{E}(e^{\imath u\langle t, X\rangle})$ for all $t \in \mathcal{X}$ and all $u \in \mathbb{R}$, which reads $\phi_{\langle t, X_n\rangle}(u) \to \phi_{\langle t, X\rangle}(u)$ for all t, u. The conclusion follows by Cor. 2.19 again. □

2.6 Central Limit Theorem

We start with a short reminder.
We refer to the *covariance matrix* of a random vector $X = (X^{(1)}, \dots, X^{(d)})$ on $\mathcal{X}$ as the positive semidefinite linear operator on $\mathcal{X} \to \mathcal{X}$, noted $\mathrm{Cov}(X)$, whose coefficient (i,j) is the covariance $\mathrm{Cov}(X^{(i)}, X^{(j)})$ of $X^{(i)}$ and $X^{(j)}$. If $\mathbb{E}(\|X\|^2) < \infty$, then $\mathrm{Cov}(X)$ is well-defined. By definition, the r.v. X is a Gaussian vector if $\langle t, X\rangle$ is a Gaussian variable for every $t \in \mathcal{X}$. If X is a Gaussian vector, then $\mathbb{E}(\|X\|^2) < \infty$ and for every $t \in \mathcal{X}$,

$$\phi_X(t) = e^{\imath\langle t, \mathbb{E}(X)\rangle} e^{-\frac{1}{2}\langle t, \mathrm{Cov}(X) t\rangle} .$$

In particular, the law of the Gaussian vector X depends only on its expectation $\mathbb{E}(X)$ and its covariance matrix $\mathrm{Cov}(X)$. The notation $\mathcal{N}(c,\Sigma)$ represents the law of a Gaussian vector having expectation c and covariance matrix Σ. By Lemma 2.16, if X has the distribution $\mathcal{N}(c,\Sigma)$, then the r.v. $AX+b$ has the distribution $\mathcal{N}(Ac+b, A\Sigma A^*)$ for every linear operator A and every vector b.

Let $(X_n : n\in\mathbb{N}^*)$ be a sequence of r.v. defined on $(\Omega,\mathcal{F},\mathbb{P})$ into $(\mathcal{X},\mathscr{B}(\mathcal{X}))$.

Theorem 2.22. *Suppose that (X_n) is an iid sequence of random vectors s.t. $\mathbb{E}(\|X_1\|^2)<\infty$. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i-\mathbb{E}(X_1)) \Rightarrow \mathcal{N}(0,\mathrm{Cov}(X_1))\,.$$

Th. 2.22 is a consequence of the more general result stated below in Th. 2.23. A collection of r.v. $(X_{i,n} : 1\le i\le n,\, n\in\mathbb{N}^*)$ on $(\Omega,\mathcal{F},\mathbb{P})$ into $(\mathcal{X},\mathscr{B}(\mathcal{X}))$ will be refer to as a (triangular) *array*. Such an array is said to satisfy the *Lindebergh condition* if

$$\forall\varepsilon>0,\ \lim_n \sum_{i=1}^n \mathbb{E}\left(\|X_{i,n}\|^2 \mathbb{1}_{\|X_{i,n}\|>\varepsilon}\right)=0\,. \tag{2.7}$$

Theorem 2.23. *Suppose that $(X_{i,n} : 1\le i\le n,\, n\in\mathbb{N})$ is an array of random vectors on $\mathcal{X}$. Assume the following.*

i. For every $n\in\mathbb{N}^*$, the r.v. $X_{1,n},\dots,X_{n,n}$ are independent;

ii. For every $n\in\mathbb{N}^*$ and $i=1,\dots,n$, $\mathbb{E}(X_{i,n})=0$;

iii. The Lindebergh condition (2.7) is satisfied;

iv. There exists a matrix Σ s.t.

$$\lim_n \sum_{i=1}^n \mathrm{Cov}(X_{i,n})=\Sigma\,,$$

where the limit is taken componentwise.

Then,

$$\sum_{i=1}^n X_{i,n} \Rightarrow \mathcal{N}(0,\Sigma)\,.$$

Verify by your own that Th. 2.22 is indeed a consequence of Th. 2.23. Before proving Th. 2.23, we need the following technical lemmas.

Lemma 2.24. *For every $z\in\mathbb{C}$, $|e^z-1-z|\le |z|^2e^{|z|}$.*

Proof. Just expand $|e^z-1-z|\le\sum_{k\ge 2}\frac{|z|^k}{k!}$ and the conclusion follows. □

Lemma 2.25. *For every* $x \in \mathbb{R}$, $|1 + \imath x - \frac{x^2}{2} - e^{\imath x}| \leq \min(x^2, |x|^3)$.

Proof. Write the Taylor-Lagrange expansion of $e^{\imath x}$ at the second and third order. □

Lemma 2.26. *Define* $\mathcal{U} := \{z \in \mathbb{C} : |z| \leq 1\}$. *Consider* $n \in \mathbb{N}^*$ *and two sequences* $(z_i : i = 1 \dots n)$, $(w_i : i = 1 \dots n)$ *on* $\mathcal{U}$. *Then,*

$$\left| \prod_{i=1}^n z_i - \prod_{i=1}^n w_i \right| \leq \sum_{i=1}^n |z_i - w_i| .$$

Proof. By induction. □

Proof of Th. 2.23: the scalar case. We first consider the case where $d = 1$ (random variables are real-valued). The vector case will be handled at the end of this section. Then, $\Sigma \geq 0$ is a scalar. Moreover, the function $x \mapsto \sqrt{\Sigma}\, x$ being continuous, Th. 2.5 implies that it is sufficient to prove the result for $\Sigma = 1$. Define $s_{i,n} := \mathrm{Var}(X_{i,n})$, thus

$$\lim_n \sum_{i=1}^n s_{i,n} = 1 . \tag{2.8}$$

We define $S_n := \sum_{i=1}^n X_{i,n}$. Consider a fixed $t \in \mathbb{R}$. By independence, we obtain

$$\phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_{i,n}}(t) .$$

By the triangular inequality,

$$\left| \phi_{S_n}(t) - e^{-t^2/2} \right| \leq A_n + B_n + C_n ,$$

where

$$\begin{aligned} A_n &:= \left| \prod_{i=1}^n \phi_{X_{i,n}}(t) - \prod_{i=1}^n (1 - \frac{s_{i,n} t^2}{2}) \right| \\ B_n &:= \left| \prod_{i=1}^n (1 - \frac{s_{i,n} t^2}{2}) - \prod_{i=1}^n e^{-s_{i,n} t^2/2} \right| \\ C_n &:= \left| \prod_{i=1}^n e^{-s_{i,n} t^2/2} - e^{-t^2/2} \right| . \end{aligned}$$

Now, we must prove that each of the terms A_n, B_n, C_n converges to zero. By Prop. 2.14, this will imply that, for every $t \in \mathbb{R}$, $\phi_{S_n}(t) \to \phi_{\mathcal{N}(0,1)}(t)$. By Th. 2.15, we will conclude that $S_n \Rightarrow \mathcal{N}(0, 1)$.

By Eq. (2.8), the term C_n converges to zero as $n \to \infty$. Note that for all $\varepsilon > 0$,

$$\max_{1 \le i \le n} s_{i,n} \le \varepsilon^2 + \sum_{i=1}^{n} \mathbb{E}\left((X_{i,n})^2 \mathbb{1}_{|X_{i,n}|>\varepsilon}\right) ,$$

which implies by the Lindebergh condition that

$$\lim_n \max_{1 \le i \le n} s_{i,n} = 0 . \tag{2.9}$$

As a consequence, the values $1 - \frac{s_{i,n}t^2}{2}$ are all in the interval $[-1, 1]$ for n large enough. By lemma 2.26,

$$A_n \le \sum_{i=1}^{n} |\phi_{X_{i,n}}(t) - (1 - \frac{s_{i,n}t^2}{2})|$$

$$B_n \le \sum_{i=1}^{n} |1 - \frac{s_{i,n}t^2}{2} - e^{-s_{i,n}t^2/2}| .$$

By Lemma 2.24, $B_n \le \sum_{i=1}^{n} (\frac{s_{i,n}^2 t^4}{4} e^{s_{i,n}t^2/2})$, and thus $B_n \to 0$ by Eq. (2.8) and (2.9). It remains to prove that $A_n \to 0$. Denote $r(x) := e^{\imath x} - (1 + \imath x - \frac{x^2}{2})$. Then, using that $\mathbb{E}(X_{i,n}) = 0$,

$$\phi_{X_{i,n}}(t) = 1 - \frac{s_{i,n}t^2}{2} + \mathbb{E}(r(tX_{i,n})) .$$

Hence, using Lemma 2.25

$$A_n \le \sum_{i=1}^{n} \mathbb{E}(|r(tX_{i,n})|) \ \le \ \sum_{i=1}^{n} \mathbb{E}(\min((tX_{i,n})^2, |tX_{i,n}|^3)) .$$

Note that for every $\varepsilon > 0$ small enough, we have for every $x \in \mathbb{R}$,

$$\min((tx)^2, |tx|^3) \le (tx)^2 \mathbb{1}_{|x|>\varepsilon} + |tx|^3 \mathbb{1}_{|x| \le \varepsilon} .$$

Thus,

$$A_n \le \sum_{i=1}^{n} \mathbb{E}((tX_{i,n})^2 \mathbb{1}_{|X_{i,n}|>\varepsilon}) + \sum_{i=1}^{n} \mathbb{E}(|tX_{i,n}|^3 \mathbb{1}_{|X_{i,n}| \le \varepsilon}) .$$

The first term in the righthand side converges to zero as $n \to \infty$ by the Lindebergh condition. Therefore, $\limsup_n A_n \le \varepsilon$. As this holds for every $\varepsilon > 0$ small enough, $A_n \to 0$. The proof that $S_n \Rightarrow \mathcal{N}(0, 1)$ is complete.

Proof of Th. 2.23: the vector case. Recall that if Z has the distribution $\mathcal{N}(0, \Sigma)$ and if $t \in \mathcal{X}$, then $\langle t, Z \rangle$ has the distribution $\mathcal{N}(0, \langle t, \Sigma t \rangle)$. By Lemma 2.21, it is therefore sufficient that for every $t \in \mathcal{X}$,

$$\sum_{i=1}^{n} \langle t, X_{i,n} \rangle \Rightarrow \mathcal{N}(0, \langle t, \Sigma t \rangle) .$$

The above convergence holds by applying Th. 2.23 to the real array $(\langle t, X_{i,n} \rangle)$.

Chapter 3

Manipulation of Convergences

3.1 Notations o_P, O_P

Proposition 3.1. *Let (X_n), (Y_n), X, Y be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Consider $a \in \mathcal{X}$. The following properties hold true:*

i. $X_n \xrightarrow{P} a$ *if and only if* $X_n \Rightarrow a$ *;*

ii. *If* $X_n \Rightarrow X$ *and* $Y_n - X_n \xrightarrow{P} 0$, *then* $Y_n \Rightarrow X$ *.*

Proof. *i.* The direct implication is a consequence of Prop. 2.20. For the converse, choose $\varepsilon > 0$ and consider the closed ball $B_\varepsilon := \{x \in \mathcal{X} : \|x - c\| \leq \varepsilon\}$. By Cor. 2.19, $\mathbb{P}(\|X_n - c\| > \varepsilon) = \mathbb{P}(X_n \in B_\varepsilon^c)$ converges to $\mathbb{P}(a \in B_\varepsilon^c) = 0$.
ii. Choose $f \in C_b^\ell(\mathcal{X})$ with a Lipschitz constant L and uniform norm $\|f\|_\infty$. By the triangular inequality,

$$|\mathbb{E}(f(Y_n)) - \mathbb{E}(f(X))| \leq \mathbb{E}(|f(Y_n) - f(X_n)|) + |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \,.$$

The second term in the righthand side tends to zero since $X_n \Rightarrow X$. The first term

$$\mathbb{E}(|f(Y_n) - f(X_n)|) \leq \mathbb{E}(\min(L\|Y_n - X_n\|, \|f\|_\infty)) \,.$$

By the first point, $Y_n - X_n \Rightarrow 0$. Thus, the righthand side of the above inequality converges to zero, since the mapping $\min(L\| \,.\, \|, \|f\|_\infty)$ is bounded continuous. This shows that $\mathbb{E}(f(Y_n)) \to \mathbb{E}(f(X))$ and the proof is concluded by Cor. 2.19. □

Proposition 3.2. *Let (X_n), X be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Let (Y_n) be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Y}, \mathscr{B}(\mathcal{Y}))$. Consider $c \in \mathcal{Y}$. If $X_n \Rightarrow X$ and $Y_n \Rightarrow c$, then $(X_n, Y_n) \Rightarrow (X, c)$.*

Proof. Choose $f \in C_b(\mathcal{X} \times \mathcal{Y})$. We have $\mathbb{E}(f(X_n, c)) \to \mathbb{E}(f(X, c))$ because $f(\,.\,, c) \in C_b(\mathcal{X})$. Thus $(X_n, c) \Rightarrow (X, c)$. By the same argument, $(0, Y_n - c) \Rightarrow 0$. By Prop. 3.1-ii, the sum $(X_n, c) + (0, Y_n - c)$, which coincides with (X_n, Y_n), converges in law to (X, c). □

Definition 3.3. Let (X_n) be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Let (Y_n) be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

- We say that (X_n) is a $o_P(1)$, noted $X_n = o_P(1)$, if $X_n \xrightarrow{P} 0$.
- We say that (X_n) is a $o_P(Y_n)$, noted $X_n = o_P(Y_n)$, if there exists a sequence (Z_n) of r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ s.t. $Z_n = o_P(1)$ and for every n, $X_n = Z_n Y_n$.
- We say that (X_n) is a $O_P(1)$, noted $X_n = O_P(1)$, if (X_n) is bounded in probability (*i.e.* is tight).
- We say that (X_n) is a $O_P(Y_n)$, noted $X_n = O_P(Y_n)$, if there exists a sequence (Z_n) of r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ s.t. $Z_n = O_P(1)$ and for every n, $X_n = Z_n Y_n$.

By Prop. 2.9, any sequence of r.v. which converges in law is a $O_P(1)$.

Lemma 3.4 (Slutsky). *Consider r.v. (X_n), X on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ s.t. $X_n \Rightarrow X$.*

i. Consider r.v. (Y_n) on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ and $a \in \mathcal{X}$ s.t. $Y_n \Rightarrow a$. Then, $X_n + Y_n \Rightarrow X + a$.

ii. Consider r.v. (Y_n) on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ and $a \in \mathbb{R}$ s.t. $Y_n \Rightarrow a$. Then, $Y_n X_n \Rightarrow aX$.
If $a \neq 0$, then $X_n / Y_n \Rightarrow X/a$.

Proof. It is a straightforward application of Prop. 3.2-i along with the continuity of the maps $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$ and $(x, y) \mapsto x/y$ (for the last case, continuity indeed holds everywhere outside $\mathbb{R} \times \{0\}$, which is fine by Prop. 2.5). □

In the sequel, we will write abusively $o_P(1)$ (resp. $O_P(1)$) to designate an element from a sequence of r.v. which is a $o_P(1)$ (resp. $O_P(1)$) without even naming the sequence. For instance, the formal identity

$$o_P(1) + O_P(1) = O_P(1)\,,$$

which we will prove below, precisely means the following: *If (X_n) and (Y_n) are r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ s.t. $X_n = o_P(1)$ and $Y_n = O_P(1)$, then $X_n + Y_n = O_P(1)$.* With this formulation in mind, the statement of the following proposition should be clear.

Proposition 3.5. *The following holds true.*

i. $o_P(1) + O_P(1) = O_P(1)$;

ii. $o_P(1)O_P(1) = o_P(1)$;

iii. $o_P(1) + o_P(1) = o_P(1)$;

iv. $(1 + o_P(1))^{-1} = O_P(1)$.

Proof. i. Consider r.v. (X_n) and (Y_n) on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$ s.t. $X_n = o_P(1)$ and $Y_n = O_P(1)$. Consider an arbitrary $\varphi : \mathbb{N} \to \mathbb{N}$ strictly increasing. As (Y_{φ_n}) is tight, there exists a further subsequence (Y_{ψ_n}) and some $\mu \in \mathcal{P}(\mathcal{X})$ s.t. $Y_{\psi_n} \Rightarrow \mu$, by Th. 2.10 By Slutsky's lemma, $X_{\psi_n} + Y_{\psi_n} \Rightarrow \mu$. Thus, every subsequence of $(X_n + Y_n)$ admits a further subsequence which converges in law. By the converse in Th. 2.10, $(X_n + Y_n)$ is tight.

ii. Choose (X_n), (Y_n) as above but assume this time that X_n is real valued. For the same reason (X_nY_n) is tight. Choose any subsequence converging in law, say $X_{\varphi_n}Y_{\varphi_n} \Rightarrow \nu$ for some $\nu \in \mathcal{P}(\mathcal{X})$. Extract a further subsequence, say ψ_n, s.t. $Y_{\psi_n} \Rightarrow \mu$ for some $\mu \in \mathcal{P}(\mathcal{X})$. By Slutsky's lemma, it is clear that $\nu = \delta_0$ *i.e.* $X_{\varphi_n}Y_{\varphi_n} \Rightarrow 0$. By Prop. 2.11, we obtain that $X_nY_n \Rightarrow 0$, which also reads $X_nY_n \xrightarrow{P} 0$.

iii. This point is a direct consequence of Slutsky's lemma.

iv. Consider real valued r.v. (X_n) s.t. $X_n = o_P(1)$, or equivalently, $X_n \Rightarrow 0$. By Th. 1.7, $(1 + X_n)^{-1} \Rightarrow 1$. By Prop. 2.9, $(1 + X_n)^{-1}$ forms a tight sequence. □

3.2 Delta-Method

Recall that a mapping $g : \mathcal{X} \to \mathcal{Y}$ is said to be *differentiable* at the point $x \in \mathcal{X}$ if there exists $J \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ s.t.

$$\lim_{u \to 0} \frac{\|g(x + u) - g(x) - Ju\|}{\|u\|} = 0 \,.$$

The map J will be noted $\nabla g(x)$ and refered to as the *Jacobian matrix* of g at x.

Theorem 3.6 (Delta-method). *Consider r.v. (X_n), X on $\mathcal{X}$, consider $g : \mathcal{X} \to \mathcal{Y}$ and $x \in \mathcal{X}$. Let $(r_n : n \in \mathbb{N})$ be a positive sequence. Assume the following:*

i. *g is differentiable at x ;*

ii. $\lim_n r_n = +\infty$;

iii. $r_n(X_n - x) \Rightarrow X$.

Then, $r_n(g(X_n) - g(x)) \Rightarrow \nabla g(x)X$.

Proof. In class. □

Bibliography

[1] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1995.

[2] R. M. Dudley. *Real Analysis and Probability.* Wadsworth and Brooks/Cole, 1989.