III. Convergence of Random Variables.

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with

$$X_n : \Omega \longrightarrow \mathcal{X} \quad \text{state space}$$

↳ some metric space endowed with the Borel σ-algebra. e.g.: $\mathcal{X} = \mathbb{R}^n$.

There are various notions of convergence for X_n.

objective:
- describe these convergences and their connections
- have useful criterions of convergence.
- prove a few convergence theorems.

Chapter III. Convergence of Random Variables.

1. Almost sure convergence

Definition The sequence $(X_n)_{n\in\mathbb{N}}$ converges a.s. to a random variable X with values in $(\mathcal{X}, d)$ iff:

$$\mathbb{P}\left[X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega) \right] = \mathbb{P}\left[d(X_n(\omega), X(\omega)) \xrightarrow[n\to+\infty]{} 0 \right] = 1.$$

Since the variables X_n, X might be defined up to a negligible subset of Ω (e.g., X_n or $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$), this is the strongest possible notion.

It is easy to see that a.s. convergence is kept by operations (addition, multiplication, etc.) and by continuous maps $f : (\mathcal{X}, d_{\mathcal{X}}) \longrightarrow (\mathcal{Y}, d_{\mathcal{Y}})$.

The first and main example of a.s. convergence is:

Theorem (law of large numbers)

Let $(X_n)_{n\geq 1}$ be a sequence of independent real random variables, all with the same distribution $\mathbb{P}_{X_1} = \mathbb{P}_{X_n}$ $\forall n \geq 1$.

If $X_1 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then $M_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow[a.s.]{} \mathbb{E}[X_1]$.

We shall try to give a proof which might extend to more general temporal means
$\longrightarrow$ ergodic theorem.

<u>setting</u>: $(\Omega, \mathcal{F}, \mathbb{P})$ some probability space.
$T: \Omega \longrightarrow \Omega$ a measurable map (time shift) such that $T_* \mathbb{P} = \mathbb{P}$.

An event $A \in \mathcal{F}$ is said invariant if $T^{-1}(A) = A$. The invariant events form a σ-subalgebra $\mathcal{I} \subset \mathcal{F}$.

<u>Definition</u>: The transformation T is called ergodic if the σ-subalgebra $\mathcal{I}$ is $\mathbb{P}$-trivial: $\forall A \in \mathcal{I}$, $\mathbb{P}[A] \in \{0, 1\}$.

<u>Theorem</u> Consider some distribution $\mathbb{P}_X$ on $\mathbb{R}$, and

$$(\Omega, \mathcal{F}, \mathbb{P}) = \left(\mathbb{R}^{\otimes \mathbb{N}^*}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}^*}, \mathbb{P}_X^{\otimes \mathbb{N}^*}\right).$$

The coordinate maps $X_1, X_2, \ldots$ on $(\Omega, \mathcal{F}, \mathbb{P})$ yield a sequence of i i D random variables with law $\mathbb{P}_X$.

Consider $T: (x_1, x_2, \ldots) \longrightarrow (x_2, x_3, \ldots)$ (time shift, or Bernoulli shift)
This is an ergodic transformation.

<u>Proof</u>: ⓐ $T_* \mathbb{P} = \mathbb{P}$. Notice that the set of subsets $A \in \mathcal{F}$ such that $\mathbb{P}[T^{-1}(A)] = \mathbb{P}[A]$ is a monotone class.

Therefore, it suffices to prove the identity for cylinders $A = A_1 \times \cdots \times A_n \times \mathbb{R}^\infty$
Hower, $\mathbb{P}[T^{-1}(A)] = \mathbb{P}[\mathbb{R} \times A_1 \times \cdots \times A_n \times \mathbb{R}^\infty]$

$$= \prod_{i=1}^{n} \mathbb{P}_X[A_i] = \mathbb{P}[A].$$

② Let $\mathcal{A} \subset \mathcal{F}$ be the set of subsets C with the approximation property: $\forall \varepsilon > 0$, $\exists n \geq 1$, $\exists C_n \in \mathcal{B}(\mathbb{R})^{\otimes n} \otimes \{\phi, \mathbb{R}^{[\![n+1, +\infty]\!]}\}$

such that $\mathbb{P}[C \Delta C_n] = \mathbb{P}[C \setminus C \cap C_n] + \mathbb{P}[C_n \setminus C \cap C_n] \leq \varepsilon$

Then, $\mathcal{A}$ is stable:

- by complementation: if C_n approximates C, then $\Omega \setminus C_n$ approximates $\Omega \setminus C$:

$$\mathbb{P}[\Omega \setminus C_n \,\Delta\, \Omega \setminus C] = \mathbb{E}[\mathbb{1}_{\Omega \setminus C_n} + \mathbb{1}_{\Omega \setminus C} - 2\,\mathbb{1}_{\Omega \setminus C_n} \mathbb{1}_{\Omega \setminus C}]$$
$$= \mathbb{E}[1 - \mathbb{1}_{C_n} + 1 - \mathbb{1}_C - 2 + 2\mathbb{1}_{C_n} + 2\mathbb{1}_{C} - 2\,\mathbb{1}_{C_n} \mathbb{1}_C]$$
$$= \mathbb{E}[\mathbb{1}_{C_n} + \mathbb{1}_C - 2\,\mathbb{1}_{C_n} \mathbb{1}_C] = \mathbb{P}[C_n \Delta C].$$

- by increasing union: if $C = \bigcup_{m \geq 1}\uparrow C^{(m)}$ with the $C^{(m)} \in \mathcal{A}$, then $\mathbb{P}[C \setminus C^{(m)}] \leq \frac{\varepsilon}{2}$ for m large enough, and $\mathbb{P}[C^{(m)} \Delta C_n] \leq \frac{\varepsilon}{2}$ for some $C_n \in \mathcal{B}(\mathbb{R})^{\otimes n} \otimes \{\phi, \mathbb{R}^{[\![n+1, +\infty]\!]}\}$, so

$$\mathbb{P}[C \Delta C_n] = \mathbb{E}[\mathbb{1}_C + \mathbb{1}_{C_n} - 2\,\mathbb{1}_C \mathbb{1}_{C_n}]$$
$$\leq \frac{\varepsilon}{2} + \mathbb{E}[\mathbb{1}_{C^{(m)}} + \mathbb{1}_{C_n} - 2\,\mathbb{1}_{C^{(m)}} \mathbb{1}_{C_n}]$$
$$\leq \varepsilon.$$

- by simple intersection: if C and D are approximated by C_n and D_m, we can assume without loss of generality that $n = m$, and then $C_n \cap D_n$ approximate $C \cap D$:

$$\mathbb{P}[(C \cap D) \Delta (C_n \cap D_n)] = \mathbb{E}[\mathbb{1}_C \mathbb{1}_D + \mathbb{1}_{C_n} \mathbb{1}_{D_n} - 2\,\mathbb{1}_C \mathbb{1}_{C_n} \mathbb{1}_D \mathbb{1}_{D_n}]$$
$$= \mathbb{E}[\mathbb{1}_C (\mathbb{1}_{D \Delta D_n})] + \mathbb{E}[\mathbb{1}_{D_n} (\mathbb{1}_{C_n} - \mathbb{1}_C + 2\,\mathbb{1}_C \mathbb{1}_D - 2\,\mathbb{1}_C \mathbb{1}_D \mathbb{1}_{C_n})]$$
$$= \mathbb{E}[\mathbb{1}_C (\mathbb{1}_{D \Delta D_n})] + \mathbb{E}[\mathbb{1}_{D_n} (\mathbb{1}_{C \Delta C_n})]$$
$$- 2\,\mathbb{E}[\mathbb{1}_C \mathbb{1}_{D_n} (1 - \mathbb{1}_D)(1 - \mathbb{1}_{C_n})].$$
$$\leq \mathbb{P}[D \Delta D_n] + \mathbb{P}[C \Delta C_n].$$

So, $\mathcal{A} = \mathcal{F}$.

③ Let A be an invariant event, and A_n approximating A up to $\varepsilon/2$.
We have

$$\mathbb{P}[A] = \mathbb{P}[A \cap T^{-k}(A)] \leq \mathbb{P}[A \cap T^{-k}(A) \,\Delta\, A_n \cap T^{-k}(A_n)] + \mathbb{P}[A_n \cap T^{-k}(A_n)]$$

$$\leq \varepsilon + \mathbb{P}[A_n \cap T^{-k}(A_n)]$$

$$= \mathbb{P}[A_n]\,\mathbb{P}[T^{-k}(A_n)] \text{ if } k \geq n.$$

$$= \mathbb{P}[A_n]^2$$

$$\leq \varepsilon + (\mathbb{P}[A] + \mathbb{P}[A \Delta A_n])^2 \leq \frac{\varepsilon^2}{4} + 2\varepsilon + \mathbb{P}[A]^2.$$

By taking the limit, $\mathbb{P}[A] \leq \mathbb{P}[A]^2$, so $\mathbb{P}[A] \in \{0, 1\}$ $\square$.

In the setting of ergodic transformations of a probability space, the ergodic theorems are limiting results for the means

$$M_n(X) = \frac{1}{n}\sum_{k=0}^{n-1} X \circ T^k \quad \text{with } X \in L^p(\Omega, \mathcal{F}, \mathbb{P}).$$

(they might converge to something not constant...)

A. Square integrable variables.

Let T be a transformation of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

$H = \{ X \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \mid X = X \circ T \}$.

Theorem ① $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $M_n(X) \xrightarrow[L^2]{} \Pi_H(X)$, where Π_H is the orthogonal projection on the closed vector subspace H.

② If the transformation is ergodic, then $H = \mathbb{R}$ is the space of a.s. constant functions, and $\Pi_H = \mathbb{E}$.

Proof ① We denote $U_T(X) = X \circ T$. U_T is a continuous linear map $L^p(\Omega, \mathcal{F}, \mathbb{P}) \to L^p(\Omega, \mathcal{F}, \mathbb{P})$ which preserves the L^p norm, $\forall p \geq 1$.

For instance, in L^1, $\| U_T(X)\|_{L^1} = \mathbb{E}[|U_T(X)|]$

$$= \int_\Omega |X \circ T(\omega)| \, \mathbb{P}[d\omega] = \int_\Omega |X(\omega')| \, T_* \mathbb{P}[d\omega']$$

$$= \int_\Omega |X(\omega')| \, \mathbb{P}[d\omega'] = \|X\|_{L^1} .$$

In particular, U_T is an isometry of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$, and H is the closed subspace $\ker(U_T - \mathrm{id}_{L^2})$.
Set $I = \mathrm{im}(U_T - \mathrm{id}_{L^2})$.

- if $x_1 = U_T(y) - y \in I$ and $x_2 \in H$, then

$$\langle x_1 | x_2 \rangle = \mathbb{E}[x_1 x_2] = \mathbb{E}[U_T(y) x_2 - y x_2]$$
$$= \mathbb{E}[U_T(y) U_T(x_2)] - \mathbb{E}[y x_2] \text{ since } x_2 \in H$$
$$= \mathbb{E}[y x_2] - \mathbb{E}[y x_2] = 0$$

So $H \subset I^\perp$

- conversely, if $x \in I^\perp$, then

$$\mathbb{E}[(U_T(x) - x)^2] = \mathbb{E}[U_T(x)^2] + \mathbb{E}[x^2] - 2\mathbb{E}[U_T(x) x]$$
$$= \mathbb{E}[x(x - U_T(x))] = 0 \text{ since } x \in I^\perp .$$

so $x \in H$.

Conclusion: $H = I^\perp$; we have the decomposition in orthogonal subspaces:
$L^2(\Omega, \mathcal{F}, \mathbb{P}) = H \oplus H^\perp = H \oplus \overline{I}$.

Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. We decompose $X = \pi_H(X) + R$ with $R \in \overline{I}$

$R = \lim_{n \to +\infty} R_m$

$R_m \in I$.

Then, $M_n(X) = \pi_H(X) + M_n(R)$

$$= \pi_H(X) + \underbrace{M_n(R - R_m)}_{\leq \varepsilon \text{ in } L^2 \text{ norm.}} + M_n(R_m), \quad R_m \in I$$

- $\|M_n(R - R_m)\|_{L^2} \leq \frac{1}{n} \sum_{k=0}^{n-1} \|(R - R_m) \circ T^k\|_{L^2} = \|R - R_m\|_{L^2} \leq \varepsilon .$
- $\|M_n(R_m)\|_{L^2} = \frac{1}{n} \|\sum_{k=0}^{n-1} (V \circ T - V) \circ T^k\|_{L^2} = \frac{1}{n} \|V \circ T^n - V\|_{L^2} \leq \frac{2\|V\|_{L^2}}{n}$

for some V

So $\lim_{n\to+\infty} M_n(X) = \Pi_H(X)$ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

② If T is ergodic, notice that the property $\mathbb{P}[A] \in \{0,1\}$ extends to almost invariant events : $\mathbb{P}[A \Delta T^{-1}(A)] = 0$.

Then, if $X \in H$, the events $\{X \leq a\}$ are almost invariant, so they all have probability 0 or 1. This implies that X is a.s. constant.
In this case, for $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $\Pi_H(X)$ is a constant, equal to $\mathbb{E}[\Pi_H(X)] = \lim_{n\to+\infty} \mathbb{E}[M_n(X)] = \mathbb{E}[X]$; $\Pi_H = \mathbb{E}$. $\square$

__Example__: if $(X_n)_{n\geq 1}$ is a sequence of L^2 iid variables, then
$\frac{X_1 + \dots + X_n}{n} = M_n(X_1) \xrightarrow[L^2]{} \mathbb{E}[X]$. This can be proven with a simple computation of variance.

B. Integrable variables.

__Corollary__: Given a transformation T of a probability space, and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $M_n(X)$ converges in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ towards a variable Y s.t. $Y = Y \circ T$.
If the transformation is ergodic, then $Y = \mathbb{E}[X]$ is constant.

__Proof__: On the one hand, M_n is a contraction of the Banach space L^1.
On the other hand, $\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $M_n(X)$ converges towards a T-invariant variable in L^2

Let $X \in L^1$; by density of L^2 in L^1, $\exists Z \in L^2$ s.t. $\|X - Z\|_{L^1} \leq \varepsilon$.
Then,
$$\|M_{n_1}(X) - M_{n_2}(X)\|_{L^1} \leq 2\varepsilon + \|M_{n_1}(Z) - M_{n_2}(Z)\|_{L^2}$$

so $(M_n(X))_{n\geq 1}$ is a Cauchy sequence in $L^1(\Omega, \mathcal{F}, \mathbb{P})$, and converges to some $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.
Moreover, $\|M_n(X) \circ T - M_n(X)\|_{L^1} = \frac{1}{n}\|X \circ T^n - X\|_{L^1} \leq \frac{2\|X\|_{L^1}}{n} \to 0$.

So by taking the limit, $Y \circ T = Y$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. The ergodic case is then analogous to what has been done for square integrable variables.

C. Birkhoff theorem.

<u>Theorem</u> We also have for $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$: $M_n(X) \xrightarrow[a.s.]{} Y$.

This immediately implies the law of large numbers, since we have in this case an ergodic transformation.

<u>Lemma</u> (maximal inequality). For $L \in \mathbb{R}$, set $B_L = \{\omega \in \Omega \mid \sup_{n \geq 1} M_n(X)(\omega) > L\}$
We have:

$$L\, \mathbb{P}[B_L] \leq \mathbb{E}[X \mathbb{1}_{B_L}].$$

More generally, for any invariant event C,

$$L\, \mathbb{P}[B_L \cap C] \leq \mathbb{E}[X \mathbb{1}_{B_L \cap C}].$$

<u>Proof</u> Set $Y = X - L$; $S_n = \sum_{k=0}^{n-1} Y \circ T^k$; $R_n = \max_{0 \leq k \leq n} S_k \geq 0$.

We have $B_L = \bigcup_{n \in \mathbb{N}} \uparrow \{R_n > 0\}$.

- If $H \geq 0$ a.s., then $U_T(H) \geq 0$ a.s., because
  $\mathbb{P}[U_T(H) \geq 0] = (T_* \mathbb{P})[H \geq 0] = \mathbb{P}[H \geq 0] = 1$.
- Therefore,
  $$U_T(R_n) + Y \geq \max_{0 \leq k \leq n} (U_T(S_k) + Y) = \max_{0 \leq k \leq n} S_{k+1} \geq \max_{1 \leq k \leq n} S_k.$$
  (take $H = R_n - S_k$)

On the event $\{R_n > 0\}$, the RHS is R_n since $S_0 = 0$, so $Y \geq R_n - U_T(R_n)$.
Therefore,

$$\mathbb{E}[Y \mathbb{1}_{\{R_n > 0\}}] \geq \mathbb{E}[R_n \mathbb{1}_{R_n > 0}] - \mathbb{E}[U_T(R_n) \mathbb{1}_{R_n > 0}]$$
$$\geq \mathbb{E}[R_n] - \mathbb{E}[U_T(R_n)] = 0.$$

By dominated convergence, $\mathbb{E}[Y \mathbb{1}_{B_L}] \geq 0$, whence the first inequality.
For the second inequality, we can assume $\mathbb{P}[C] > 0$, and then work in $\left(C, \mathcal{F}_{|C}, \frac{\mathbb{P}[\cdot]}{\mathbb{P}[C]}\right)$.

$\square$.

Final proof: Set $Y_- = \liminf_{n \to +\infty} M_n(X)$, $Y_+ = \lim_{n \to +\infty} M_n(X)$.
Since $U_T \circ M_n = M_n + \frac{U_T^{\circ n} - id}{n}$, Y_- and Y_+ are T-invariant.
It suffices to show that $\forall \alpha < \beta \in \mathbb{Q}$, $E_{\alpha,\beta} = \{Y_- < \alpha \text{ and } Y_+ > \beta\}$ has probability 0.

Then, $Y_- = Y_+ = Y = \lim_{n \to +\infty} M_n(X)$ a.s. (and this must also be the limit in $L^1(\Omega, \mathcal{F}, \mathbb{P})$).
We have $E_{\alpha,\beta} \subset B_\beta = \{\sup M_n(X) > \beta\}$.
Therefore, $\beta\, \mathbb{P}[E_{\alpha,\beta}] \leq \mathbb{E}[Y \mathbb{1}_{E_{\alpha,\beta} \cap B_\beta}] = \mathbb{E}[Y \mathbb{1}_{E_{\alpha,\beta}}]$.
Replacing X by $-X$, we also have $\alpha\, \mathbb{P}[E_{\alpha,\beta}] \geq \mathbb{E}[Y \mathbb{1}_{E_{\alpha,\beta}}]$.
So, $(\alpha - \beta)\, \mathbb{P}[E_{\alpha,\beta}] \geq 0 \implies \mathbb{P}[E_{\alpha,\beta}] = 0$ $\square$.

2. Convergence in probability and in L^1.

We have encountered above another notion of convergence of random variables:

Definition: Let $(X_n)_{n \in \mathbb{N}}$, $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.
We have $X_n \xrightarrow[L^1]{} X$ if $\mathbb{E}[|X_n - X|] \xrightarrow[n \to +\infty]{} 0$.

In general, if $X_n \xrightarrow[a.s.]{} X$, we need a control on the sequence $(X_n)_{n \in \mathbb{N}}$ in order to also have $X_n \xrightarrow[L^1]{} X$, which would imply in particular that
$$\mathbb{E}[X] = \lim_{n \to +\infty} \mathbb{E}[X_n].$$

For instance, if $\left| \begin{array}{l} X_n \xrightarrow[a.s.]{} X \\ |X_n| \le R \quad \forall n \in \mathbb{N} \end{array} \right.$ with $R \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ nonnegative, then by the dominated convergence theorem $X_n \xrightarrow[\mathcal{L}^1]{} X$.

__Definition__ Let $(X_n)_{n\in\mathbb{N}}$, X be random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to a metric space $(\mathcal{X}, d)$.
The sequence **converges in probability** to X if, $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}[d(X_n, X) \ge \varepsilon] = 0. \qquad \text{Notation: } X_n \xrightarrow[\mathbb{P}]{} X.$$

__Proposition__ If $X_n \xrightarrow[a.s.]{} X$, then $X_n \xrightarrow[\mathbb{P}]{} X$.

Indeed, η, ε being fixed, we have $\lim_{n\to+\infty} d(X_n, X) = 0$ with probability 1, so $\mathbb{P}\left[\bigcup_{n\in\mathbb{N}} \nearrow \bigcap_{m\ge n} \{d(X_m, X) < \varepsilon\}\right] = 1$. For n large enough,

$$\mathbb{P}\left[\bigcap_{m\ge n} \{d(X_m, X) < \varepsilon\}\right] \ge 1 - \eta, \quad \text{so } \mathbb{P}[d(X_n, X) \ge \varepsilon] \le \eta.$$

$$\Rightarrow \lim_{n\to+\infty} \mathbb{P}[d(X_n, X) \ge \varepsilon] = 0. \qquad \square$$

__Proposition__ If the X_n and X are in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \xrightarrow[\mathcal{L}^1]{} X$, then $X_n \xrightarrow[\mathbb{P}]{} X$.

__Proof__ $\mathbb{P}[|X_n - X| \ge \varepsilon] = \mathbb{E}[\mathbb{1}_{|X_n - X| \ge \varepsilon}] \le \mathbb{E}\left[\mathbb{1}_{|X_n - X| \ge \varepsilon} \dfrac{|X_n - X|}{\varepsilon}\right]$

(Markov inequality) $\qquad \le \dfrac{\mathbb{E}[|X_n - X|]}{\varepsilon} \xrightarrow[n\to+\infty]{} 0, \quad \forall \varepsilon > 0. \quad \square$

So,

$$\begin{tikzcd}
X_n \xrightarrow[a.s.]{} X \arrow[rr, "\text{with a domination hypothesis}"] \arrow[dr] & & X_n \xrightarrow[\mathcal{L}^1]{} X \arrow[dl] \\
& X_n \xrightarrow[\mathbb{P}]{} X &
\end{tikzcd}$$

Can we go up?

__Definition__ A family $(X_i)_{i\in I}$ is **uniformly integrable** if

$$\lim_{M \to +\infty} \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{|X_i| \geq M}] = 0.$$

examples : ① a single random variable which is integrable is uniformly integrable.
(by DCT).

② a union of two UI families $(X_i)_{i \in I}$ and $(Y_j)_{j \in J}$ is UI.

Proposition $(X_i)_{i \in I}$ is UI
$\Longleftrightarrow$ $(X_i)_{i \in I}$ is bounded in L^1, and $\lim_{\varepsilon \to 0} \left(\sup_{\substack{A : \mathbb{P}[A] \leq \varepsilon \\ i \in I}} \mathbb{E}[|X_i| \mathbb{1}_A] \right) = 0$

Proof:

$\Longleftarrow$ If $K = \sup_{i \in I} \mathbb{E}[|X_i|]$, then $\mathbb{P}[|X_i| \geq M] \leq \frac{K}{M}$.
Fix $\eta > 0$ and ε such that $\sup_{\mathbb{P}[A] \leq \varepsilon,\, i \in I} \mathbb{E}[|X_i| \mathbb{1}_A] \leq \eta$.

If $M \geq \frac{K}{\varepsilon}$, then $\sup_i \mathbb{E}[|X_i| \mathbb{1}_{|X_i| \geq M}] \leq \eta$.

$\Longrightarrow$ If $\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{|X_i| \geq M}] \leq 1$, then $\sup_{i \in I} \mathbb{E}[|X_i|] \leq M + 1$
$\Rightarrow$ bounded sequence.

Then, $\sup_{\substack{A : \mathbb{P}[A] \leq \varepsilon \\ i \in I}} \mathbb{E}[|X_i| \mathbb{1}_A] \leq \sup_{A, i} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| < M\} \cap A}]$
$+ \sup_i \mathbb{E}[|X_i| \mathbb{1}_{|X_i| \geq M}]$

$\leq M\varepsilon + \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{|X_i| \geq M}]$. Take $M = \frac{1}{\sqrt{\varepsilon}}$ and let $\varepsilon \to 0$.
$\square$.

Theorem The following are equivalent:

① $X_n \xrightarrow[L^1]{} X$

② $X_n \xrightarrow[\mathbb{P}]{} X$ and $(X_n)_{n \in \mathbb{N}}$ is UI.

Proof ① $\Rightarrow$ ② We already know that $X_n \xrightarrow[\mathbb{P}]{} X$. Moreover, if $\mathbb{P}[A] \leq \varepsilon$, then $\mathbb{E}[|X_n| \mathbb{1}_A] \leq \mathbb{E}[|X_n - X|] + \mathbb{E}[|X| \mathbb{1}_A]$

$\leq \eta$ for $n \geq n_0$ $+ \eta$ for ε small enough. ((X) is ui).

So, $\sup_{\substack{n \geq n_0 \\ \mathbb{P}[A] \leq \varepsilon}} \mathbb{E}[|X_n| \mathbb{1}_A] \leq 2\eta$. Since $(X_n)_{n \leq n_0 - 1}$ is also ui,
by diminishing the value of ε, we can remove the assumption $n \geq n_0$.
So, $(X_n)_{n \in \mathbb{N}}$ is ui.

② $\Rightarrow$ ① If $(X_n)_{n \in \mathbb{N}}$ is ui, then $(X_n - X)$ also, because
$\mathbb{E}[|X_n - X| \mathbb{1}_A] \leq \mathbb{E}[|X_n| \mathbb{1}_A] + \mathbb{E}[|X| \mathbb{1}_A]$.

Then, $\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - X| \mathbb{1}_{\varepsilon \leq |X_n - X| < M}] + \mathbb{E}[|X_n - X| \mathbb{1}_{|X_n - X| \geq M}]$
$+ \mathbb{E}[|X_n - X| \mathbb{1}_{|X_n - X| < \varepsilon}]$
$\leq M \, \mathbb{P}[|X_n - X| \geq \varepsilon] + \varepsilon + \varepsilon$ for M large enough

$\square$.

3. Convergence in distribution

Let $(\mathcal{X}, d)$ be a metric space and $(X_n)_{n \in \mathbb{N}}$ be a sequence of $\mathcal{X}$-valued random variables.

Definition One says that $(X_n)_{n \in \mathbb{N}}$ converges in law or in distribution if, $\forall f$ continuous bounded on $\mathcal{X}$, $\mathbb{E}[f(X_n)] \xrightarrow[n \to +\infty]{} \mathbb{E}[f(X)]$.
Notice that this is a statement about the laws $\mathbb{P}_{X_n}$ and $\mathbb{P}_X$:

$$\forall f \in \mathcal{C}_b(\mathcal{X}), \int_{\mathcal{X}} f(x) \, \mathbb{P}_{X_n}[dx] \longrightarrow \int_{\mathcal{X}} f(x) \, \mathbb{P}_X[dx].$$

More generally, given probability measures μ_n, μ on $\mathcal{X}$ w.r.t. the Borel σ-algebra, we say that $\mu_n \xrightarrow[n\to+\infty]{} \mu$ if $\int_{\mathcal{X}} f(x)\, \mu_n(dx) \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}} f(x)\, \mu(dx)$. $\forall f \in C_b(\mathcal{X})$

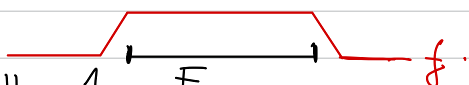
<u>Theorem</u> (Portmanteau) The following are equivalent for probability measures μ_n, μ:

1. $\mu_n \to \mu$.
2. $\forall f$ Lipschitz bounded on $\mathcal{X}$, $\mu_n(f) \to \mu(f)$
3. $\forall U$ open subset, $\liminf_{n\to+\infty} \mu_n(U) \geq \mu(U)$
4. $\forall F$ closed subset, $\limsup_{n\to+\infty} \mu_n(F) \leq \mu(F)$.
5. $\forall A$ Borel subset with $\mu(\partial A) = 0$, $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.

<u>Proof</u> 1 $\Rightarrow$ 2 is obvious.

3 $\Leftrightarrow$ 4 ———

2 $\Rightarrow$ 4. Set $f(x) = \max(0, 1 - K d(x, F))$



This function is K lipschitz, larger than $\mathbb{1}_F$.

By monotone convergence, $\forall \varepsilon > 0$, $\exists K$ s.t. $\mu(F) \geq \mu(f) - \varepsilon$.

$$\geq \limsup_{n\to+\infty} \mu_n(f) - \varepsilon$$
$$\geq \limsup \mu_n(F) - \varepsilon.$$

So, $\mu(F) \geq \liminf_{n\to+\infty} \mu_n(F)$.

3 + 4 $\Rightarrow$ 5 : $\limsup \mu_n(A) \leq \limsup \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(\mathring{A})$

$$\leq \liminf \mu_n(\mathring{A}) \leq \liminf \mu_n(A)$$

so all the terms are equal to $\lim_{n\to+\infty} \mu_n(A)$.

5 $\Rightarrow$ 1. Let $f \in C_b(\mathcal{X})$. Up to an affine transformation, we can assume $0 \leq f(x) \leq 1$ $\forall x \in \mathcal{X}$.

$$\mu_n(f) = \int_{\mathcal{X}} \int_0^1 \mathbb{1}_{y \le f(x)}\, dy\, \mu_n(dx) = \int_0^1 \mu_n(\{x \in \mathcal{X} \mid f(x) \ge y\})\, dy.$$

The map $y \underset{g}{\longmapsto} \mu(\underbrace{\{x \in \mathcal{X} \mid f(x) \ge y\}}_{F_y \text{ closed subset of } \mathcal{X}\,;\ \mathring{F}_y \supset \{x \in \mathcal{X} \mid f(x) > y\}})$ is non increasing, so it has at most a countable set of discontinuities.

If g is continuous on y, then $\mu(F_y) = \mu(\mathring{F}_y) \Longrightarrow \lim\limits_{n \to +\infty} \mu_n(F_y) = \mu(F_y)$.

So, $g_n(y) = \mu_n(F_y)$ converges almost everywhere on $[0,1]$ to $g(y)$.

By dominated convergence, $\lim\limits_{n \to +\infty} \mu_n(f) = \int_0^1 \mu(F_y)\, dy = \mu(f)$ $\square$.

<u>Corollary</u>: if $X_n \xrightarrow[\mathbb{P}]{} X$, then $X_n \rightharpoonup X$.

<u>Proof</u>: Let F be a closed subset of $\mathcal{X}$. We have:

$$\begin{aligned}
\limsup_{n \to +\infty} \mathbb{P}[X_n \in F] &= \limsup_{n \to +\infty} \mathbb{P}\Big[\bigcap_{m \ge 1} \{d(X_n, F) \le \tfrac{1}{m}\}\Big] \\
&\le \limsup_{n \to +\infty} \mathbb{P}\big[d(X_n, F) \le \tfrac{1}{2m}\big] \\
&\le \limsup_{n \to +\infty} \mathbb{P}\big[d(X_n, F) \le \tfrac{1}{2m} \text{ and } d(X_n, X) \le \tfrac{1}{2m}\big] \\
\scriptstyle (X_n \xrightarrow[\mathbb{P}]{} X) \quad &\le \mathbb{P}\big[d(X, F) \le \tfrac{1}{2m}\big]. \quad \forall m \ge 1.
\end{aligned}$$

So, $\limsup\limits_{n \to +\infty} \mathbb{P}[X_n \in F] \le \mathbb{P}[X \in F]$ $\forall F$ closed subset $\square$.

Scheme of convergences:

$$\begin{tikzcd}
X_n \xrightarrow[a.s]{} X \arrow[rr, dashed, "\text{with domination}"] \arrow[dr, Rightarrow] & & X_n \xrightarrow[L^1]{} X \\
& X_n \xrightarrow[\mathbb{P}]{} X \arrow[ur, Rightarrow] \arrow[ur, dashed, bend right, "\text{with UI.}"'] \arrow[d, Rightarrow] & \\
& X_n \rightharpoonup X &
\end{tikzcd}$$

<u>Theorem</u> Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real random variables. We have $X_n \xrightarrow[n\to+\infty]{} X$ iff $\forall t$ point of continuity of F_X,

$$F_{X_n}(t) \xrightarrow[n\to+\infty]{} F_X(t).$$

<u>Proof</u>: $\Rightarrow$ this is Portmanteau ⑤ applied to $(-\infty, t] = A$.

$\Leftarrow$ let U be an open subset of $\mathbb{R}$. We have a countable disjoint union $U = \bigcup_{i\in I} (a_i, b_i)$.

$$\text{Thus, } \mathbb{P}_X(U) \leq \sum_{i=1}^{m} \mathbb{P}_X((a_i, b_i)) + \varepsilon = \left(\sum_{i=1}^{m} F_X(b_i^-) - F_X(a_i)\right) + \varepsilon$$

$$\leq \left(\sum_{i=1}^{m} F_X(c_i) - F_X(d_i)\right) + 2\varepsilon$$

with $c_i \leq b_i$, $d_i \geq a_i$, $\underline{F_X \text{ continuous at } c_i}$ d_i

$$\leq \lim_{n\to+\infty} \left(\sum_{i=1}^{m} F_{X_n}(c_i) - F_{X_n}(d_i)\right) + 2\varepsilon$$

$$\leq \liminf \mathbb{P}_{X_n}(U) + 2\varepsilon. \qquad \square$$

The other essential criterion for real valued random variables is:

<u>Theorem</u> (Lévy) We have $X_n \xrightarrow[n\to+\infty]{} X$ iff $\forall \xi \in \mathbb{R}$, $\mathbb{E}[e^{i\xi X_n}] \xrightarrow[n\to+\infty]{} \mathbb{E}[e^{i\xi X}]$

The implication $\Rightarrow$ is obvious (split $e^{i\xi x}$ in real and imaginary parts).

<u>Lemma 1.</u> The map $\mathcal{M}^1(\mathbb{R}) \longrightarrow C_b(\mathbb{R})$ is injective.

$$\mu \longmapsto \left(\hat{\mu} : \xi \to \int e^{i\xi x}\mu(dx)\right)$$

<u>Proof</u>: Let $X \sim \mu$ and $Y_n \sim \mathcal{N}(0, \frac{1}{n})$ independent.

$X_n = X + Y_n$

We have $\mathbb{P}[|X-X_n| \geq \varepsilon] = \mathbb{P}[|Y_n| \geq \varepsilon] = \mathbb{P}[|Y_1| \geq \sqrt{n}\,\varepsilon] \underset{n \to +\infty}{\longrightarrow} 0$.

So $X_n \underset{\mathbb{P}}{\longrightarrow} X$, and in particular $X_n \underset{p.s.}{\longrightarrow} X$.

So, it suffices to show that $\hat{\mu}$ determines the law of the variables X_n.

$$\mathbb{P}[X_n \in K] = \int_{\mathbb{R}^2} \mathbb{1}_{x+y \in K}\, \mu(dx)\, \underbrace{\frac{1}{\sqrt{\frac{2\pi}{n}}} e^{-\frac{ny^2}{2}}}_{\frac{1}{\sqrt{2\pi n}} \int e^{-\frac{s^2}{2n} - isy} ds}\, dy.$$

K bounded.

$$= \frac{1}{2\pi} \int_{\mathbb{R}^3} \mathbb{1}_{x+y \in K}\, e^{-\frac{s^2}{2n^2} - isy}\, \mu(dx)$$

$$= \frac{1}{2\pi} \int_{\substack{s \in \mathbb{R} \\ v \in K}} e^{-\frac{s^2}{2n^2} - isv}\, \hat{\mu}(s)\, ds \qquad \square.$$

Therefore, it suffices to prove that there is a subsequence of $\mu_n = \mathbb{P}_{X_n}$ which converges in distribution to $\mu = \mathbb{P}_X$. (compacity + unicity of the limit).

<u>Lemma 2</u> Under the assumptions of Lévy's criterion, $\forall \varepsilon > 0$, $\exists K$ compact such that $\inf_{n \in \mathbb{N}} \mu_n(K) \geq 1 - \varepsilon$.

<u>Proof</u>: $\forall \delta > 0$,

$$\mathbb{1}_{|\delta x| > 2} \leq 2\left(1 - \frac{\sin \delta x}{\delta x}\right) = \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos \xi x)\, d\xi$$

Integration / $\mu_n(dx)$ yields:

$$\mu_n([-\tfrac{2}{\delta}, \tfrac{2}{\delta}]^c) = \mathbb{P}[|X_n| > \tfrac{2}{\delta}] \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos \xi X_n)\, d\xi$$

$$\leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \hat{\mu}_n(\xi))\, d\xi$$

Since $\hat{\mu}$ is continuous and $\hat{\mu}(0) = 1$, for δ small enough,
$$\frac{1}{\delta}\int_{-\delta}^{\delta} |1 - \hat{\mu}(\xi)|\, d\xi \leq \frac{\varepsilon}{2}.$$

By dominated convergence, for n large enough,
$$\frac{1}{\delta}\int_{-\delta}^{\delta} |1 - \hat{\mu}_n(\xi)|\, d\xi \leq \varepsilon. \implies \mathbb{P}\left[X_n \in \left[-\frac{2}{\delta}, \frac{2}{\delta}\right]\right] \geq 1-\varepsilon.$$
By modifying the value of δ, we can include the small n values $\square$.

<u>Lemma 3</u> Let K be a compact metric space. Given $(\mu_n)_{n\in\mathbb{N}}$ sequence of probability measures on K, $\exists$ subsequence μ_{n_k} s.t. $\mu_{n_k} \xrightarrow[k\to+\infty]{} \mu$.
a probability measure μ

<u>Proof</u>: On the one hand, by Riesz representation theorem, a continuous linear form on $C(K)$ writes uniquely as $\mu = \mu_+ - \mu_-$, a finite signed measure.

So, the μ_n are elements of $\underbrace{C(K)}_{\text{Banach space}}{}^{*}$] dual.

Then, the convergence in distribution is the restriction to $\mathcal{P}^1(K) \subset C(K)^*$ of the $*$-weak topology:
$$\mu_n \to \mu \iff \mu_n \underset{*}{\rightharpoonup} \mu.$$

On the other hand, by the Banach Alaoglu theorem, the unit ball $\overline{B}_{C(K)^*}(0,1)$ is compact for the $*$-weak topology.
$\underbrace{}$ endowed with the norm $|||\phi||| = \sup_{f\in C(K),\, \|f\|_\infty \leq 1} |\phi(f)|$

By the Stone-Weierstrass theorem, $C(K)$ admits a countable dense subset $(f_n)_{n\in\mathbb{N}}$.

Then, $d(\phi, \phi') = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \min(1, |\phi(f_n) - \phi(f_n')|)$ is a distance on $\overline{B}_{C(K^*)}(0,1)$ which corresponds to the $*$weak topology.

Finally, $\mathcal{M}^1(K) \subset \overline{B}_{C(K^*)}(0,1)$ is closed in this compact metric space:
$$= \left\{ \phi \in \overline{B}_{C(K^*)}(0,1) \text{ s.t. } \begin{array}{l} \phi(1) = 1 \\ \phi(f) \geq 0 \text{ if } f \geq 0 \end{array} \right\}.$$

Conclusion: $\mathcal{M}^1(K)$ is a compact metric space for the $*$weak topology $\square$.

<u>Proof of Lévy's theorem</u> (which obviously adapts to $\mathbb{R}^{n\geq 1}$):
We endow $\mathbb{R}$ in a compact space, by using e.g.
$$\begin{aligned} \varphi : \mathbb{R} &\longrightarrow [-1, 1]. \\ x &\longmapsto \frac{2}{\pi} \arctan x \end{aligned}$$
By Lemma 3, $(\varphi_* \mu_n)_{n\in\mathbb{N}}$ admits a convergent subsequence: $\exists$ extraction $(n_k)_{k\in\mathbb{N}}$ such that $\varphi_* \mu_{n_k} \rightharpoonup \nu$ probability measure on $[-1, 1]$.

By Lemma 2, given $\varepsilon > 0$, $\mu_n([-M, M]^c) \leq \varepsilon$ $\forall n \in \mathbb{N}$ for M large enough.
$$\Longrightarrow \varphi_* \mu_n([-(1-\eta), (1-\eta)]^c) \leq \varepsilon \quad \text{for } \eta \text{ small enough}.$$
By Portmanteau theorem,
$$\liminf \varphi_* \mu_{n_k}([-(1-\eta), (1-\eta)]^c) \geq \nu([-(1-\eta), (1-\eta)]^c)$$
So, $\nu(\{-1\}) = \nu(\{1\}) = 0$. Thus ν is supported on $(-1, 1)$ and we can use φ^{-1}!
$$\mu_{n_k} \underset{k\to+\infty}{\rightharpoonup} \mu = \varphi^{-1}_* \nu$$
(the convergence in law is compatible with continuous maps)

this must be μ by Lemma 1 and injectivity of the Fourier transform.

$\square$.

<u>Application</u> the central limit theorem.
Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables in $L^2(\Omega, \mathcal{F}, \mathbb{P})$

$m = \mathbb{E}[X_1]$, $\sigma^2 = var(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$

We have $\dfrac{1}{\sqrt{n}} \dfrac{X_1 + \dots + X_n - nm}{\sigma} \xrightarrow[n \to +\infty]{} \mathcal{N}(0,1)$.

<u>Proof</u> If G_n is the left hand side, then:

$$\mathbb{E}[e^{i\xi(X_1+\dots+X_n)}] = \prod_{j=1}^{n} \mathbb{E}[e^{i\xi X_j}] = \hat{\mu}(\xi)^n, \quad \mu = \mathbb{P}_{X_1}.$$

$$\mathbb{E}[e^{i\xi G_n}] = \left(\hat{\mu}\left(\frac{\xi}{\sqrt{n}\,\sigma}\right)\right)^n e^{-i\frac{\sqrt{n}\,m\xi}{\sigma}}.$$

However, $\xi \to \underset{C(\xi).}{\overset{}{\log \mathbb{E}[e^{i\xi X_1}]}}$ is | well defined in a neighborhood of 0
| C^2.

($C(\xi) = \log \mathbb{E}[e^{i\xi X_1}]$.)

$$C(\xi) = im\xi - \frac{\sigma^2\xi^2}{2} + o(\xi^2).$$

So, $\left(\hat{\mu}\left(\frac{\xi}{\sqrt{n}\,\sigma}\right)\right)^n = \exp\left(\frac{i\sqrt{n}\,m\xi}{\sigma} - \frac{\xi^2}{2} + o(1)\right)$ with ξ fixed.

$$\mathbb{E}[e^{i\xi G_n}] = \exp\left(-\frac{\xi^2}{2} + o(1)\right) \to \text{Fourier transform of } \mathcal{N}(0,1) \ \square$$

multidimensional version: $(\vec{X}_n)_{n\geq 1}$ iid sequence in $L^2(\Omega, \mathcal{F}, \mathbb{P}) \otimes \mathbb{R}^d$.

$$\vec{G}_n = \frac{1}{\sqrt{n}}(\vec{X}_1 + \dots + \vec{X}_n - nM) \text{ with } M = \mathbb{E}[\vec{X}_1] = \begin{pmatrix} \mathbb{E}[\vec{X}_{1,1}] \\ \vdots \\ \mathbb{E}[\vec{X}_{1,d}] \end{pmatrix}$$

For $\vec{\xi} \in \mathbb{R}^d$ we have

$$\mathbb{E}[e^{i\langle \vec{\xi} \mid \vec{G}_n \rangle}] \xrightarrow[n \to +\infty]{} \exp\left(-\frac{\vec{\xi}^{\,t}\Sigma\vec{\xi}}{2}\right) \text{ with } \Sigma = \left(cov(\vec{X}_{1,i}, \vec{X}_{1,j})\right)_{i,j}$$

So $\vec{G}_n \xrightarrow[n \to +\infty]{} \mathcal{N}_{\mathbb{R}^d}(0, \Sigma)$ Gaussian vector.

Another important convergence in law :

$$S_n = \sum_{i=1}^{n} \underbrace{B(p_{n,i})}_{\substack{\text{independent} \\ \text{Bernoulli variables}}} \quad \text{with} \quad \sum_{i=1}^{n} p_{n,i} \xrightarrow[n\to+\infty]{} \lambda$$

$$\sum_{i=1}^{n} p_{n,i}^2 = o(1).$$

Then, $S_n \xrightarrow[n\to+\infty]{} \text{Poisson}(\lambda)$.

<u>Proof</u>: if $X \sim \text{Poisson}(\lambda)$, then $\mathbb{E}[e^{i\xi X}] = \sum_{k=0}^{\infty} e^{i\xi k} \frac{\lambda^k}{k!} e^{-\lambda}$

$$= e^{\lambda(e^{i\xi}-1)}.$$

On the other hand, $\mathbb{E}[e^{i\xi B(p)}] = 1 + p(e^{i\xi}-1)$.

So $\log \mathbb{E}[e^{i\xi S_n}] = \sum_{i=1}^{n} \log(1 + p_{n,i}(e^{i\xi}-1))$

$$= \left(\sum_{i=1}^{n} p_{n,i}\right)(e^{i\xi}-1) + O\left(\sum p_{n,i}^2\right)$$

$$\longrightarrow \lambda(e^{i\xi}-1).$$

$\square$.