PT = CONVERGENCE & LIMIT THEOREMS

I _ MODES OF CONVERGENCE

I.1. Definitions.

For a sequence $\{x_n\}$ of real numbers, we say that x_n converges to $x \in \mathbb{R}$ as $n \to +\infty$, and we write "$x_n \to x$" if $\forall \varepsilon > 0 \ \exists n_\varepsilon \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ for any $n \geq n_\varepsilon$.
In other words, x_n should be in a small neighborhood of x for large n.

But now, what does $X_n \to X$, for a sequence of RVs $\{X_n\}$? Random Variables are (measurable) functions and not real numbers. When considering convergence of RVs, we are thus interested in convergence of functions. (don't get fooled by the fact that $X_n \in \mathbb{R}$; it actually means that $\forall \omega \in \Omega$, $X_n(\omega) \in \mathbb{R}$)

The weakest notion of convergence in real analysis is point-wise convergence. Naturally, we would ask for convergence of $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. However, it turns out that this is asking too much, and for most practical purposes, it is enough to restrict ourselves to set set of probability one:

Def (ALMOST SURE CONVERGENCE aka WITH PROBA 1)
For a sequence $\{X_n\}$ of RVs, we say that X_n converges almost surely to X, and we write $X_n \xrightarrow{a.s.} X$ as $n \to +\infty$, if $\exists A \in \mathcal{F}$, $\mathbb{P}(A) = 1$, such that $\forall \omega \in A$, $X_n(\omega) \longrightarrow X(\omega)$, as $n \to +\infty$.

Note that all X_n and X must be defined on a same proba. space

As we shall see shortly, even if point-wise convergence is a weak convergence in real analysis, it turns out to be one of the strongest notion of convergence in probability theory.

Remark: The set $\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\} =: A$ is indeed an event since it can be written

$$A = \bigcap_{k\geq 1} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \{\omega\in\Omega \mid \underbrace{|X_n(\omega) - X(\omega)| < \tfrac{1}{k}}_{\text{measurable function}}\} \in \mathcal{F}$$

$\Updownarrow \quad \Updownarrow \quad \Updownarrow$

set of ω s.t. $\forall k \geq 1 \ \exists m \geq 1 \ \forall n \geq m \quad |X_n(\omega) - X(\omega)| < \frac{1}{k}$

$\Leftrightarrow \ \forall \varepsilon > 0 \ \exists m \geq 1 \ \forall n \geq m \quad |X_n(\omega) - X(\omega)| < \varepsilon$

Recall from continuity of the probability measure that for a sequence of events $A_1 \supset A_2 \supset \dots$, $\lim_{i\to+\infty} \mathbb{P}(A_i) = \mathbb{P}\left(\bigcap_{k\geq 1} A_k\right)$, so that $\mathbb{P}\left(\bigcap_{k\geq 1} A_k\right) = 1$ is equivalent to $\mathbb{P}(A_k) = 1 \ \forall k \geq 1$ (see page 13 in PT = SOLID FOUNDATIONS)

In our definition of A above, we have $A = \bigcap_{k\geq 1} A_k$, with $A_k := \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \{|X_n - X| < \frac{1}{k}\}$, and we see that $A_1 \supset A_2 \supset \dots$ Thus, convergence $X_n \xrightarrow{a.s.} X$ is equivalent to

$$\mathbb{P}(A) = 1 = \mathbb{P}\Big(\bigcap_{k\geq 1} A_k\Big)$$

$$\Updownarrow \quad \forall k \geq 1 \quad \mathbb{P}\Big(\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} |X_n - X| < \tfrac{1}{k}\Big) = 1$$

$$\Updownarrow \quad \forall \varepsilon > 0 \quad \mathbb{P}\Big(\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} |X_n - X| < \varepsilon\Big) = 1$$

take complement

$$\Updownarrow \quad \forall \varepsilon > 0 \quad \mathbb{P}\Big(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} |X_n - X| \geq \varepsilon\Big) = 0$$

The last written equation is equivalent to:

$\forall \varepsilon > 0 \quad \mathbb{P}\left(|X_n - X| \geq \varepsilon \quad \text{i.o.} \right) = 0$

(recall the definition of "infinitely often" page 14 in PT = SOLID FOUNDATIONS)

Now, recall Borel-Cantelli lemma, and apply it to the sequence of events $A_n := \{ |X_n - X| \geq \varepsilon \}$. We see that if $\sum_{n\geq 1} \mathbb{P}(A_n) = \sum_{n \geq 1} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty$, then $\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(|X_n - X| \geq \varepsilon \text{ i.o.}) = 0$.

In other words, $X_n \xrightarrow{a.s.} X$, as $n \to +\infty$. true $\forall \varepsilon > 0$

This leads to the following criterion for almost sure convergence:

BOREL-CANTELLI for A.S. CONVERGENCE.

Let $\{X_n\}$ = sequence of RVs.

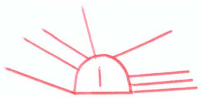
If $\forall \varepsilon > 0$, $\sum_{n \geq 1} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty$, then $X_n \xrightarrow{a.s.} X$, as $n \to +\infty$.

To get almost-sure convergence, it is not enough that $\mathbb{P}(|X_n - X| \geq \varepsilon) \to 0$ as $n \to +\infty$; this sequence of probabilities must converge sufficiently fast to zero. Otherwise, we only have ...

Def (CONVERGENCE in PROBABILITY)

We say that X_n converges in probability to X, and we write $X_n \xrightarrow{P} X$ as $n \to +\infty$ if

$\forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$



(a) Prove the second Borel-Cantelli lemma:

If $\{A_n\}_{n \geq 1}$ are independent events such that $\sum_{n\geq 1} \mathbb{P}(A_n) = +\infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$

(b) Deduce from (a) that if $\{X_n\}_{n\geq 1}$ is a sequence of independent RVs,

$X_n \xrightarrow{a.s} 0 \iff \sum_{n \geq 1} \mathbb{P}(|X_n| \geq \varepsilon) < +\infty$

(c) Suppose that $X_n \sim B(p_n)$, independent. Show that

- $X_n \xrightarrow{P} 0$ if $p_n \to 0$
- $X_n \xrightarrow{a.s} 0$ if $\sum_{n \geq 1} p_n < +\infty$

Remarks: (i) We immediately conclude that convergence in probability is weaker than almost sure convergence.

(ii) If it exists, the limit in probability is unique: if $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} X'$, then $X = X'$ a.s.

Indeed, note that $\mathbb{P}(X \neq X') = \mathbb{P}\left(\bigcup_{k \geq 1} |X - X'| > \frac{1}{k} \right)$

We want to show that

$\forall \varepsilon > 0 \quad \mathbb{P}(|X - X'| > \varepsilon) = 0$. It follows from:

$\{|X - X'| > \varepsilon\} \subseteq \{|X_n - X| > \frac{\varepsilon}{2}\} \cup \{|X_n - X'| > \frac{\varepsilon}{2}\}$

$|X_n - X'| + |X_n - X| \geq |X - X_n + X_n - X'| = |X - X'|$ and if $|X - X'| > \varepsilon$, then either $|X_n - X'|$ or $|X_n - X|$ must be $> \frac{\varepsilon}{2}$.

Thus

$\mathbb{P}(|X - X'| > \varepsilon) \leq \underbrace{\mathbb{P}(|X_n - X| > \tfrac{\varepsilon}{2})}_{\to 0} + \underbrace{\mathbb{P}(|X_n - X'| > \tfrac{\varepsilon}{2})}_{\to 0}$

(by assumption).

Def (CONVERGENCE in QUADRATIC MEAN)

We say that a sequence $\{X_n\}$ of RVs converges in quadratic mean, and we write $X_n \xrightarrow{\mathcal{L}^2} X$, as $n \to +\infty$, if $E(X_n - X)^2 \to 0$, as $n \to +\infty$.

Recall the definition of the space of square integrable functions $\mathcal{L}^2 = \{X : \Omega \to \mathbb{R} \mid \|X\|_2 < \infty\}$, where $\|X\|_2 = \sqrt{EX^2}$ is induced by the inner product $E(XY) = \langle X, Y \rangle_{\mathcal{L}^2}$, and defining the metric $d_2(X,Y) = \|X - Y\|_2$ (p.16 in PT-INTEGRAL & EXPECTATIONS)

Convergence of RVs in quadratic mean $\equiv$ convergence of RVs in $\mathcal{L}^2$.

Alternatively, we define:

Def (CONVERGENCE in MEAN)

A sequence of integrable RVs $\{X_n\}$ converges in mean to $X \in \mathcal{L}^1$ as $n \to +\infty$, and we write $X_n \xrightarrow{\mathcal{L}^1} X$, if $E|X_n - X| \to 0$, as $n \to +\infty$.

Note that convergence in mean and quadratic mean are for RVs (and their limit) defined on a common probability space.

→ More generally, we define convergence in p-norm if $E|X_n - X|^p \to 0$ as $n \to +\infty$, $p \geq 1$.

× Example = let $\{X_n\}$ be such that $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$, $\mathbb{P}(X_n = a_n) = \frac{1}{n}$, for some sequence $\{a_n\} \in \mathbb{R}$.

Since $\mathbb{P}(|X_n| \geq \varepsilon) = \mathbb{P}(X_n = a_n) = \frac{1}{n} \to 0$, we see that $X_n \xrightarrow{P} 0 \equiv X$

Moreover,

$E|X_n - X| = E|X_n| = \frac{|a_n|}{n}$, which does not necessarily tend to zero:

- $a_n = 2n$, then $E|X_n| = 2 \not\to 0$
- $a_n = n^2$, then $E|X_n| \to \infty$
- $a_n = \sqrt{n}$, then $E|X_n| \to 0$ $\quad X_n \xrightarrow{\mathcal{L}^1} 0$
  but $EX_n^2 = 1 \not\to 0$ $\quad X_n \not\xrightarrow{\mathcal{L}^2} 0$

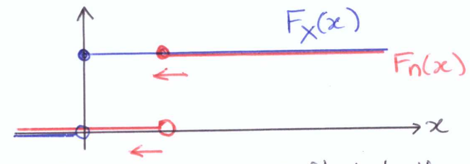
Def (CONVERGENCE in DISTRIBUTION aka WEAK CV)

We say that X_n converges in distribution to X, and we write $X_n \xrightarrow{d} X$, if $\lim_{n\to\infty} F_n(x) = F_X(x)$ at all points $x \in \mathbb{R}$ where $F_X(x)$ is continuous.

For this notion of convergence, the RVs X_n and X need not to be defined on a common probability space.

if X is AC, then we can drop the requirement "where $F_X(x)$ is continuous", which is automatically satisfied.

Remarks (i) In the definition, we have a restriction to continuity points of F_X. It turns out that requiring convergence of the F_n at any x would be too restrictive. Consider the following example, where $X \equiv 0$ and $X_n \equiv \frac{1}{n}$ are constant RVs. Their distributions are:



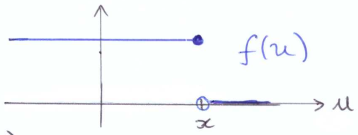
As $n \nearrow$, the function F_n is shifted to the left, and it seems natural to claim that in the limit, X_n and X have the same distribution function. However, at 0,

$F_n(0) = 0 \not\to 1 = F_X(0)$ $(\forall n)$ $\Rightarrow$ We do not have convergence at zero, which is a discontinuity point of F_X.

However $F_n(x) \to F_X(x)$ $\forall x \neq 0$.

So, would you like to exclude such situations?

(ii) Put $f_x(u) = \mathbb{1}(u \leq x)$.



Then $X_n \xrightarrow{d} X$ if $\forall x$, continuity point of F_X, holds

$$\begin{array}{ccc} \mathbb{P}(X_n \leq x) & \longrightarrow & \mathbb{P}(X \leq x) \\ \| & & \| \\ \mathbb{E}\{\mathbb{1}(X_n \leq x)\} & & \mathbb{E}\{\mathbb{1}(X \leq x)\} \\ \| & & \| \\ \mathbb{E}\{f_x(X_n)\} & & \mathbb{E}\{f_x(X)\} \end{array}$$

That is $\mathbb{E}\{f_x(X_n)\} \to \mathbb{E}\{f_x(X)\}$ for f_x = indicator function (a bounded function, & discontinuous).

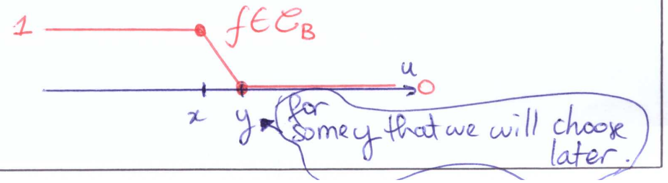
It turns out that we can restrict ourselves to the class of bounded + continuous functions (denoted $\mathcal{C}_B$) to get weak convergence, so that we have the equivalent definition:

Def / Thm: We say that X_n converges in distribution to X if and only if $\forall f \in \mathcal{C}_B$ = space of continuous & bounded functions, $\mathbb{E}\{f(X_n)\} \to \mathbb{E}\{f(X)\}$, as $n \to +\infty$.

proof $\boxed{\Leftarrow}$ Suppose that $\forall f \in \mathcal{C}_B$, $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$.

Let x = continuity point of F_X.

Step I. Consider the function f defined as:



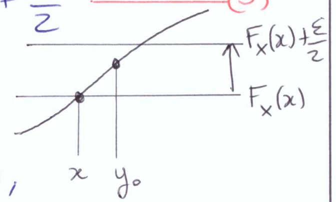
Then $\mathbb{1}(u \leq x) \leq f(u) \leq \mathbb{1}(u \leq y)$

$\mathbb{P}(X_n \leq x) \leq \mathbb{E} f(X_n) \leq \mathbb{P}(X_n \leq y)$

$F_n(x) \leq \mathbb{E} f(X_n) \leq F_n(y)$ —— (1)

Similarly $F_X(x) \leq \mathbb{E} f(X) \leq F_X(y)$. —— (2)

Next, let $\varepsilon > 0$. Since x is a continuity point of F_X, there exist y_0 (depending on ε) such that $F_X(x) \leq F_X(y_0) \leq F_X(x) + \frac{\varepsilon}{2}$ —— (3)



Select f_0 defined previously with this particular y_0.

Since $\mathbb{E}[f_0(X_n)] \to \mathbb{E} f_0(X)$,

$\exists n_0$ s.t. $\forall n \geq n_0$, $\mathbb{E} f_0(X_n) \leq \mathbb{E} f_0(X) + \frac{\varepsilon}{2}$. —— (4)

We conclude that $\forall n \geq n_0$,

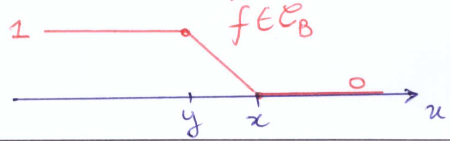
$$F_n(x) \underset{(1)}{\leq} \mathbb{E} f_0(X_n) \underset{(4)}{\leq} \mathbb{E} f_0(X) + \frac{\varepsilon}{2} \underset{(3)}{\leq} F_X(y_0) + \frac{\varepsilon}{2} \leq F_X(x) + \varepsilon$$

$\forall n \geq n_0 \quad F_n(x) \leq F_X(x) + \varepsilon$ —— (5)

Step II. We proceed analogously for $y < x$, first by constructing functions $\tilde{f}$ s.t.

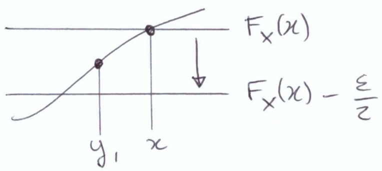
$F_n(y) \leq \mathbb{E} \tilde{f}(X_n) \leq F_n(x)$ —— (6)

$F(y) \leq \mathbb{E} \tilde{f}(X) \leq F(x)$ —— (7)



Since x is a continuity point of F_X, we see that there exists $y_1 < x$ s.t.

$$F_X(x) - \frac{\varepsilon}{2} \leqslant F_X(y_1) \leqslant F_X(x) \qquad (8)$$



Select $\tilde{f}$ for this y_1, call it $\tilde{f}_1$.

Since $\mathbb{E}\,\tilde{f}_1(X_n) \longrightarrow \mathbb{E}\,\tilde{f}_1(X)$ as $n \to +\infty$,

$\exists n_1$ s.t. $\forall n \geqslant n_1$, $\quad \mathbb{E}\,\tilde{f}_1(X_n) \geqslant \mathbb{E}\,\tilde{f}_1(X) - \frac{\varepsilon}{2} \qquad (9)$

Thus, $\forall n \geqslant n_1$,

$$F_n(x) \underset{(6)}{\geqslant} \mathbb{E}\,\tilde{f}_1(X_n) \underset{(9)}{\geqslant} \mathbb{E}\,\tilde{f}_1(X) - \frac{\varepsilon}{2} \underset{(7)}{\geqslant} F_X(y_1) - \frac{\varepsilon}{2} \underset{(8)}{\geqslant} F_X(x) - \varepsilon$$

Summarizing, $\forall n \geqslant \max(n_1, n_2)$,

$F(x) - \varepsilon \leqslant F_n(x) \leqslant F_X(x) + \varepsilon$, so that we established that $F_n(x) \longrightarrow F_X(x)$.

$\boxed{\Rightarrow}$ Admitted.

One interest of weak convergence is to approximate unknown distributions (or known distributions analytically intractable) of X_n by the (often simpler) distribution of its limit.

Arithmetic properties of weak convergence.

If $X_n \xrightarrow{d} X$, then it is not true that $X_n - X \xrightarrow{d} 0$ in general. Find a counterexample.

Since $\mathbb{E}\, f(X_n) \to \mathbb{E}\, f(X)$ for all $f \in \mathcal{C}_B$, it is true in particular for $f(x) = e^{itx}$, $t \in \mathbb{R}$, so that

$$\mathbb{E}\, e^{itX_n} = \varphi_{X_n}(t) \longrightarrow \mathbb{E}\, e^{itX} = \varphi_X(t)$$

the ChF of X_n

$\Rightarrow$ If $X_n \xrightarrow{d} X$, then $\varphi_{X_n}(t) \longrightarrow \varphi_X(t)$ $\forall t \in \mathbb{R}$.

It turns out that the reciprocal is true:

Theorem: (Paul Levy)

If for any $t \in \mathbb{R}$ $\varphi_n(t) \longrightarrow \varphi(t)$ as $n \to +\infty$, and the limit function φ is continuous at $t = 0$, then

- $\varphi(t)$ is the ChF of some RV X
- $X_n \xrightarrow{d} X$

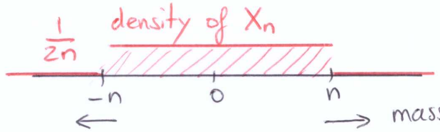
Note that this is a very general result: we do not assume that the function φ is the ChF of some RV: this fact follows from a simple hypothesis on the limit.

Remark: If the limiting φ is not continuous at $t = 0$, it usually means that some probability escapes to infinity. For example, take $X_n \sim \mathcal{U}(-n, n)$ uniformly distributed on the interval $[-n, n]$.

Then $\varphi_n(t) = \mathbb{E}\, e^{itX_n} = \dfrac{\sin nt}{nt} \longrightarrow \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases} = \varphi(t)$

see p. 8 in PT = CHARACTERISTIC FUNCTIONS.

Here $\varphi(t)$ is discontinuous at 0. Moreover, we see that



mass escapes at ∞.

Theorem (Slutsky)

Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c \in \mathbb{R}$. Then

(i) $X_n + Y_n \xrightarrow{d} X + c$

(ii) $X_n Y_n \xrightarrow{d} cX$

We are not specifying a joint distribution for (X_n, Y_n). We only assume that the marginal converge, not the joint structure.

proof = (i) We may assume $c=0$. Let x be a continuity point of F_X, and let $\varepsilon > 0$. Then

$$\mathbb{P}(X_n + Y_n \leq x) = \mathbb{P}(X_n + Y_n \leq x, |Y_n| \leq \varepsilon) + \mathbb{P}(X_n + Y_n \leq x, |Y_n| > \varepsilon)$$
$$\leq \mathbb{P}(X_n \leq x + \varepsilon) + \mathbb{P}(|Y_n| > \varepsilon)$$

Similarly,

$$\mathbb{P}(X_n \leq x - \varepsilon) \leq \mathbb{P}(X_n + Y_n \leq x) + \mathbb{P}(|Y_n| > \varepsilon)$$

$$\Big[\mathbb{P}(X_n \leq x - \varepsilon) = \mathbb{P}(X_n \leq x-\varepsilon, |Y_n| \leq \varepsilon) + \mathbb{P}(X_n \leq x - \varepsilon, |Y_n| > \varepsilon)\Big]$$

Thus,

$$\mathbb{P}(X_n \leq x - \varepsilon) - \mathbb{P}(|Y_n| > \varepsilon) \leq \mathbb{P}(X_n + Y_n \leq x) \leq \mathbb{P}(X_n \leq x + \varepsilon) + \mathbb{P}(|Y_n| > \varepsilon)$$

Taking $n \to \infty$ and then $\varepsilon \to 0$ proves (i), since x is a continuity point of F_X.

(ii) We take $c=0$ without loss of generality: if we can prove that $X_n Z_n \xrightarrow{d} 0$ for $Z_n \xrightarrow{d} 0$, then putting $Y_n = Z_n + c \xrightarrow{d} c$ (from (i)), we get that

$$X_n Y_n = \underbrace{X_n Z_n}_{\xrightarrow{d}\, 0} + \underbrace{c X_n}_{\xrightarrow{d}\, cX} \xrightarrow{d} cX \quad \text{(from (i))}$$

We have that

$$\mathbb{P}(|X_n Y_n| > \varepsilon) \leq \mathbb{P}\left(|X_n Y_n| > \varepsilon, |Y_n| \leq \tfrac{1}{M}\right) + \mathbb{P}\left(|Y_n| > \tfrac{1}{M}\right)$$
$$\leq \mathbb{P}(|X_n| > \varepsilon M) + \mathbb{P}\left(|Y_n| > \tfrac{1}{M}\right)$$
$$\xrightarrow{n\to\infty} \underbrace{\mathbb{P}(|X_n| > \varepsilon M)}_{\text{can be made arbitrarily small by taking } M \text{ sufficiently large.}} + 0$$

∎

We conclude this section with an important result, stating that convergence almost sure, in probability and in distribution are preserved under continuous transformations:

Theorem (CONTINUOUS MAPPING THEOREM)

Let $g: \mathbb{R} \to \mathbb{R}$ be a continuous function. Then

(i) If $X_n \xrightarrow{a.s.} X$, then $g(X_n) \xrightarrow{a.s.} g(X)$

(ii) If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$

(iii) If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$

This theorem holds more generally for random vectors $X_n \in \mathbb{R}^d$ and continuous functions $g: \mathbb{R}^d \to \mathbb{R}$.

proof (i) Obvious

(ii) Let
- $\varepsilon > 0$
- $A_{n,\varepsilon} := \{|g(X_n) - g(X)| > \varepsilon\}$
- $\forall N > 0 \quad I_N := [-N, N]$

Then

$$\mathbb{P}(A_{n,\varepsilon}) = \mathbb{P}(A_{n,\varepsilon}, X \in I_N) + \mathbb{P}(A_{n,\varepsilon}, X \notin I_N)$$
$$\leq \mathbb{P}(A_{n,\varepsilon}, X \in I_N) + \mathbb{P}(X \notin I_N)$$
$$\leq \mathbb{P}(A_{n,\varepsilon}, X \in I_N, |X_n - X| \leq \delta) \quad \text{I}$$
$$+ \mathbb{P}(|X_n - X| > \delta) \quad \text{II}$$
$$+ \mathbb{P}(X \notin I_N) \quad \text{III}$$

- For any $\eta > 0$, choose N large enough to make (III) less than $\eta/2$.
- Since any function continuous on a closed interval is uniformly continuous, g is uniformly continuous on I_{N+1}, so that we can choose $\delta < 1$ small enough so that if $x, y \in I_{N+1}$, $|x-y| < \delta$, then $|g(x) - g(y)| < \varepsilon$. This yields (I) $= 0$.
- Finally, since $X_n \xrightarrow{P} X$, for n large enough, $\mathbb{P}(|X_n - X| > \delta) = \text{(II)} < \frac{\eta}{2}$.

We conclude that $\mathbb{P}(A_{n,\varepsilon}) < \eta$.

(iii) We make use of the definition of weak convergence of page 7: since for $f \in \mathcal{C}_B$, the composition $f \circ g$ with g continuous is also in $\mathcal{C}_B$, we get

$$\mathbb{E} f(Y_n) \underset{\substack{\uparrow \\ Y_n = g(X_n)}}{=} \mathbb{E}\, f\circ g(X_n) \underset{\substack{\uparrow \\ \text{since } X_n \xrightarrow{d} X}}{\longrightarrow} \mathbb{E}\, f\circ g(X) = \mathbb{E} f(Y)$$

<u>Remark</u>: As noted before, a multivariate version of the continuous mapping theorem holds, so that if $\underline{X}_n \in \mathbb{R}^d \xrightarrow{d} \underline{X}$, then $\forall$ continuous g, $g(\underline{X}_n) \xrightarrow{d} g(\underline{X})$. However, for this result to hold, we need convergence of the joint distribution of $\underline{X}_n$. In general however

$$\left.\begin{array}{l} X_{n,1} \xrightarrow{d} X_1 \\ X_{n,2} \xrightarrow{d} X_2 \end{array}\right. \not\Rightarrow g(X_{n,1}, X_{n,2}) \xrightarrow{d} g(X_1, X_2),$$

except in the case $X_{n,2} \xrightarrow{d}$ constant (<u>Slutsky theorem</u>).

I.2. Relationship between the modes of convergence.

The following graph summarizes the relationship between the different notions of convergence:

$$\begin{array}{ccc} & & \xrightarrow{\mathcal{L}^2} \\ & & \Downarrow \\ & & \xrightarrow{\mathcal{L}^1} \\ & & \Downarrow \\ \xrightarrow{a.s.} & \Rightarrow & \xrightarrow{P} \\ & & \Downarrow \\ & & \xrightarrow{d} \end{array}$$

- <u>Almost sure cv $\Rightarrow$ cv in probability</u>.

Suppose $X_n \xrightarrow{a.s.} X$, so that $\exists A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$, such that $\forall \omega \in A$, $X_n(\omega) \longrightarrow X(\omega)$.

Let $\varepsilon > 0$ and put $A_n := \{ |X_n - X| > \varepsilon \}$.

For $\omega \in A$, we cannot have A_n i.o. since $X_n(\omega) \to X(\omega)$.

Thus, $\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k \subset A^c$

$$\Rightarrow 0 = \mathbb{P}(A^c) \geq \mathbb{P}\Big(\bigcap_{n\geq 1} \underbrace{\bigcup_{k\geq n} A_k}_{= B_n}\Big) \underset{\substack{\uparrow \\ \text{continuity of proba measure}}}{=} \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{k \geq n} A_k\Big).$$

$$\mathbb{P}\Big(\bigcup_{k\geq n} A_k\Big) \geq \mathbb{P}(A_n) = \mathbb{P}(|X_n - X| > \varepsilon),$$

which concludes the proof.

- <u>cv in probability $\Rightarrow$ cv in distribution</u>.

Let x be a continuity point of F_X, and let $\varepsilon > 0$.

Then $F_n(x) = \mathbb{P}(X_n \le x)$

$$= \underbrace{\mathbb{P}(X_n \le x,\ |X_n - X| \le \varepsilon)}_{\subset \{X \le x + \varepsilon\}} + \mathbb{P}(X_n \le x,\ |X_n - X| > \varepsilon)$$

$\subset \{X \le x + \varepsilon\}$ since necessarily $X \le X_n + \varepsilon$ & $X_n \le x$

$$\le \mathbb{P}(X \le x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

$$\Rightarrow \boxed{F_n(x) \le F_X(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)}.$$

likewise, $F_X(x-\varepsilon) = \mathbb{P}(X \le x - \varepsilon)$

$$= \underbrace{\mathbb{P}(X \le x-\varepsilon,\ |X_n - X| \le \varepsilon)}_{\subset \{X_n \le x\}} + \mathbb{P}(X \le x - \varepsilon,\ |X_n - X| > \varepsilon)$$

$\subset \{X_n \le x\}$ since $X \le x - \varepsilon$, $X_n \le X + \varepsilon$

$$\le \mathbb{P}(X_n \le x) + \mathbb{P}(|X_n - X| > \varepsilon)$$

$$\boxed{F_X(x-\varepsilon) \le F_n(x) + \mathbb{P}(|X_n - X| > \varepsilon)}.$$

Putting things together, we obtain

$$F_X(x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \le F_n(x) \le F_X(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

$\forall \varepsilon$ these guys tend to 0 as $n \to \infty$ since $X_n \xrightarrow{P} X$ by assumption.

Then can make ε arbitrarily small + using the continuity of F_X at x to conclude that $F_n(x) \to F(x)$ as $n \to \infty$.

- cv in mean $\Rightarrow$ cv in probability.

Follows directly from Markov inequality:

$$\mathbb{P}(|X_n - X| \ge \varepsilon) \le \frac{\mathbb{E}|X_n - X|}{\varepsilon}.$$

- cv in $\mathcal{L}^2 \Rightarrow$ cv in $\mathcal{L}^1$

Follows from Lyapunov inequality (p19 in PT: INTEGRALS & EXPECTATIONS).

More generally if $p \ge q$, then $\mathcal{L}^p$ convergence implies $\mathcal{L}^q$ convergence since $\mathbb{E}|X|^q \le (\mathbb{E}|X|^p)^{q/p}$.

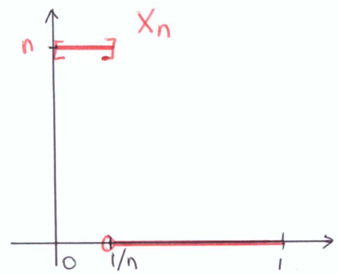
Examples and Counterexamples.

(a) Convergence in proba $\not\Rightarrow$ cv in mean.

Take $\Omega = [0,1]$
$\mathcal{F} = \mathcal{B}([0,1])$
$\mathbb{P} = \mathcal{U}(0,1)$

$X_n := n \mathbb{1}_{[0, \frac{1}{n}]}$

$X \equiv 0$



- $X_n \xrightarrow{a.s} X$ since $\forall \omega > 0$ $X_n(\omega) \to 0$.
- This implies cv in proba & in distribution. (in fact, $\forall \varepsilon > 0$ $\mathbb{P}(|X_n| \ge \varepsilon) = \mathbb{P}(X_n = n) = \frac{1}{n}$)
- However $\mathbb{E}|X_n - X| = \mathbb{E}|X_n| = 1 \not\to 0$ so X_n does not converge in mean to X.

(b) cv in mean $\not\Rightarrow$ cv in quadratic mean.

Previous $(\Omega, \mathcal{F}, \mathbb{P})$ with $Z_n := \sqrt{n}\, \mathbb{1}_{[0, \frac{1}{n}]}$.

We still have a.s & p & d convergence to $Z \equiv 0$.

In addition,

$\mathbb{E}|Z_n - Z| = \mathbb{E}|Z_n| = \frac{1}{\sqrt{n}} \to 0$ so $Z_n \xrightarrow{\mathcal{L}^1} Z$.

However,

$\mathbb{E}(Z_n - Z)^2 = \mathbb{E} Z_n^2 = 1 \not\to 0$ so Z_n does not cv in $\mathcal{L}^2$.

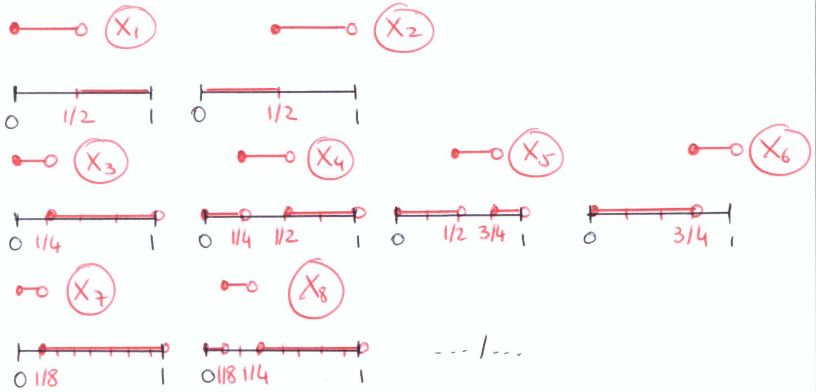
3 Generalize the previous example so that Z_n converges in $\mathcal{L}^q$ but not in $\mathcal{L}^p$, $p > q$.

(c) cv in proba $\not\Rightarrow$ cv almost sure.

Take $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P} = \mathcal{U}(0,1)$

$X \equiv 0$

The sequence $\{X_n\}$ of RVs is defined as:



$\forall \omega \in (0,1)$ $X_n(\omega)$ take values 0 an 1 infinitely often so $X_n \overset{a.s.}{\not\to} X$ a.s.

However,

$\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(X_n = 1) = \left(\frac{1}{2}\right)^{k_n} \to 0$,

(the probability decreases by 2 every power of 2)

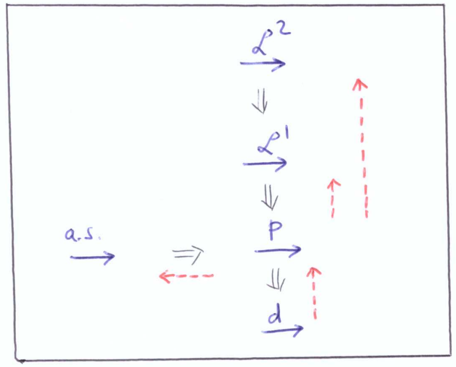
so that $X_n \xrightarrow{P} X$.

Note also that $X_n \xrightarrow{\mathcal{L}^1} X$ and $X_n \xrightarrow{\mathcal{L}^2} X$ so that in particular, $\mathcal{L}^2$-cv does not apply a.s. cv.

4 Provide a counterexample to show that convergence in distribution does not imply cv in probability.

I.3. Reverse partial implications.

Under certain specific assumptions, there exists reverse partial implications between the different modes of convergence:



$\Rightarrow$ always holds

$- - \to$ holds partially

• Cv in distribution & Cv in probability.

Thm: If for a sequence of RVs $\{X_n\}$ defined on a common probability space holds $X_n \xrightarrow{d} c = \text{constant}$, then $X_n \xrightarrow{P} c$, as $n \to \infty$.

proof: let $\varepsilon > 0$.

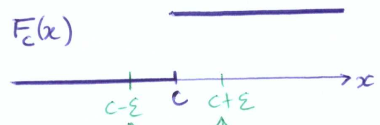
$\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(\{X_n - c > \varepsilon\} \cup \{X_n - c < -\varepsilon\})$

disjoint events

$= \mathbb{P}(X_n - c > \varepsilon) + \mathbb{P}(X_n - c < -\varepsilon)$

$= 1 - F_n(c+\varepsilon) + F_n(c-\varepsilon)$.

$\downarrow$ 1 $\qquad \downarrow$ 0, as $n \to \infty$



$\forall \varepsilon > 0$, continuity points of F_c

$\Rightarrow \mathbb{P}(|X_n - c| > \varepsilon) \to 0$, as $n \to \infty$

- **Cv in proba & Almost Sure cv.**

We have already encounted on page 3 a special case when convergence in probability implies almost sure convergence: if the cv in probability happens "fast enough", so that $\sum_{n \geq 1} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty$. As a corollary, we get:

Corollary: Suppose that $X_n \xrightarrow{P} X$, as $n \to \infty$. Then there exists a subsequence $n_k \nearrow \infty$ such that $X_{n_k} \xrightarrow{a.s.} X$, as $k \to \infty$.

proof = Since $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, as $n \to \infty$

In other words,

$\forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists N \in \mathbb{N} \quad \forall n \geq N \quad \mathbb{P}(|X_n - X| > \varepsilon) < \delta$.

Take $\varepsilon = 1/k$

Choose $N = n_1$ such that $\mathbb{P}(|X_{n_1} - X| > 1) < 1/2$

Then, let n_k be such that $\mathbb{P}(|X_{n_k} - X| > \frac{1}{k}) \leq \frac{1}{2^k}$.

Borel-Cantelli lemma applied to $\{A_k = |X_{n_k} - X| > \frac{1}{k}\}$ ensures that $\mathbb{P}(A_k \text{ i.o.}) = 0$, so that almost surely only a finite number of events A_k are realized. In other words, there exists k_0 such that $\forall k \geq k_0$, $|X_{n_k} - X| \leq 1/k$. letting $k \to \infty$, we see that almost surely, $X_{n_k} \to X$. ■

- **Cv in proba & cv in $\mathcal{L}^P$.**

We next derive conditions under which $\xrightarrow{P}$ implies $\xrightarrow{\mathcal{L}^P}$. We first introduce two fundamental results, which are extremely useful: the MONOTONE CONVERGENCE THEOREM & FATOU'S LEMMA.

Theorem (MONOTONE CONVERGENCE THEOREM)

Let $\{X_n\}$ be a sequence of RVs such that $0 \leq X_1 \leq X_2 \leq \dots$ a.s. — OK if $= \infty$

Then $\mathbb{E} X_n \uparrow \mathbb{E}(\lim_{n\to\infty} X_n)$

Note that when X_n are simple functions, then the theorem is trivial, since this is just the definition of expectation.

FATOU'S LEMMA

Let $\{X_n\}$ be a sequence of a.s. nonnegative RVs.

Then $\mathbb{E}(\liminf X_n) \leq \liminf \mathbb{E} X_n$.

If there exists Y such that $X_n \leq Y$ a.s $\forall n$, and $\mathbb{E} Y < \infty$, then $\mathbb{E}(\limsup X_n) \geq \limsup \mathbb{E} X_n$.

Use the monotone convergence theorem to $Z_n := \inf_{k \geq n} X_k$ to prove the first part of Fatou's lemma.

Theorem (DOMINATED CONVERGENCE THEOREM)

Let $X_n \xrightarrow{P} X$.

Suppose there exists $Y \in \mathcal{L}^P$, $p \geq 1$, such that $|X_n| \leq Y$ a.s $\forall n$.

Then
- X and X_n are in $\mathcal{L}^P$
- $X_n \xrightarrow{\mathcal{L}^P} X$
- $\mathbb{E} X_n \to \mathbb{E} X$

Consequence = If the sequence $\{X_n\}$ is uniformly bounded, so that $|X_n| \leq K < \infty$ $\forall n$, then $X_n \xrightarrow{P} X$ implies that $\mathbb{E} X_n \to \mathbb{E} X$.

<u>Remark</u>: Relationship between convergence in $\mathcal{L}^p$ and in probability can also be established using the notion of uniform integrability + Vitali's theorem:

A sequence of RVs $\{X_n\}_{n\geq 1}$ is said to be <u>UNIFORMLY INTEGRABLE (u.i.)</u> if

$$\lim_{c\to\infty} \sup_{n\geq 1} \mathbb{E}\left\{ |X_n| \mathbb{1}_{\{|X_n|>c\}} \right\} = 0$$

<u>Ex</u>: (i) Family of RVs containing a single element (integrable) is u.i.

(ii) A sequence of RVs dominated by an integrable RV Z is u.i.

Indeed, since $|X_n| \mathbb{1}_{\{|X_n|>c\}} \leq |Z| \mathbb{1}_{\{|Z|>c\}}$,

$$\lim_{c\to\infty} \sup_{n\geq 1} \mathbb{E}|X_n| \mathbb{1}_{\{|X_n|>c\}} \leq \lim_{c\to\infty} \mathbb{E}|Z| \mathbb{1}_{\{|Z|>c\}} = 0$$

Theorem (VITALI)

Let $\{X_n\}$ be a sequence of integrable RVs. The following statements are equivalent:

(i) $\{X_n\}_{n\geq 1}$ converges in $\mathcal{L}^1$.

(ii) $\{X_n\}_{n\geq 1}$ is a Cauchy sequence in $\mathcal{L}^1$:
$\forall \varepsilon>0 \ \exists n_0 \in \mathbb{N} \ \forall n,m \geq n_0 \ \ \mathbb{E}|X_n - X_m| < \varepsilon$.

(iii) $\{X_n\}_{n\geq 1}$ is u.i. & X_n converges in probability.

The theorem holds as well in $\mathcal{L}^p$

Note that implication (ii) $\Rightarrow$ (i) shows that $\mathcal{L}^1/\mathcal{L}^p$ is complete for the norm $\|.\|_{1/p}$.

II. LIMIT THEOREMS

II.1. Law of Large Numbers.

In this section, we turn our attention to the asymptotic behaviour of sums of independent random variables $S_n = X_1 + \dots + X_n$. The main results concern the behavior of S_n/n as $n\to\infty$ (law of large numbers). In the next subsection, we state the central limit theorem, which shows that the speed of convergence of S_n/n towards $\mathbb{E}X_1$ is $\sqrt{n}$.

There are two versions of the law of large numbers. The first version establishes convergence in probability. We first prove it under square integrable assumptions, which is a simple consequence of Chebychev inequality; and then under integrability conditions.

Theorem: <u>WEAK LAW OF LARGE NUMBER in $\mathcal{L}^2$</u>

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent & identically distributed RVs in $\mathcal{L}^2$, and put $S_n = X_1 + \dots + X_n$.

Then $\dfrac{S_n}{n} \xrightarrow{d} \mathbb{E}X_1$, as $n\to\infty$

<u>proof</u>: We want to show that

$$\forall \varepsilon>0 \quad \mathbb{P}\left(\left| \frac{S_n}{n} - \mathbb{E}X_1 \right| \geq \varepsilon \right) \to 0, \text{ as } n\to\infty.$$

Note that
- $\mathbb{E}\left(\frac{S_n}{n}\right) = \mathbb{E}X_1$
- $\mathrm{Var}\left(\frac{S_n}{n}\right) = \dfrac{\mathrm{Var}\,X_1}{n}$

Applying Chebychev inequality to S_n/n yields:

$$\mathbb{P}(|S_n/n - \mathbb{E}X_1| \geq \varepsilon) \leq \frac{\mathrm{Var}(S_n/n)}{\varepsilon^2} = \frac{\mathrm{Var}\,X_1}{n\varepsilon^2} \to 0 \text{ as } n\to\infty.$$

∎

Next we use characteristic functions to prove another version of the weak law of large number, under the assumption that $X_n \in \mathcal{L}^1$. (23)

Theorem: WEAK LAW OF LARGE NUMBERS (WLLN)

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed RVs in $\mathcal{L}^1$. Put $S_n := X_1 + \cdots + X_n$. Then

$$\frac{S_n}{n} \xrightarrow{d} \mathbb{E}X_1, \quad \text{as } n \to \infty.$$

Proof. Let $t \in \mathbb{R}$, and denote by $\varphi_X(t)$ the ChF of X_1.

Then $\varphi_{\frac{S_n}{n}}(t) = \varphi_{S_n}\left(\frac{t}{n}\right) = \left[\varphi_X\left(\frac{t}{n}\right)\right]^n$.

Expanding φ_X around 0 to the first order (which is justified since $X_1 \in \mathcal{L}^1$, so that φ_X is differentiable and $\varphi_X'(0) = i\,\mathbb{E}X_1$), we have

$$\varphi_X(s) = \underbrace{\varphi_X(0)}_{=1} + \varphi_X'(0)\, s\,(1 + o(1))$$

$$\Rightarrow \varphi_{\frac{S_n}{n}}(t) = \left[1 + i\,\mathbb{E}X_1\left(\frac{t}{n}\right)(1 + o(1))\right]^n$$

$$\longrightarrow e^{i(\mathbb{E}X_1)t} = \text{ChF of a degenerate RV equal to } \mathbb{E}X_1 \quad \blacksquare$$

Remark: Since the limit is a constant, it follows from the theorem on page 18 that convergence in probability holds as well:

$$\forall \varepsilon > 0 \quad \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X_1\right| \geq \varepsilon\right) \longrightarrow 0 \text{ as } n \to \infty$$

The probability that $\frac{S_n}{n}$ deviates outside $I_\varepsilon = [\mathbb{E}X_1 - \varepsilon, \mathbb{E}X_1 + \varepsilon]$ tends to zero. However $\frac{S_n}{n}$ could still regularly visit I_ε^c,

so that $\forall \varepsilon > 0$, $\mathbb{P}(A_{n,\varepsilon} \text{ i.o.}) > 0$, (24)

where $A_{n,\varepsilon} := \{|X_n - X| \geq \varepsilon\}$. The next result shows that this cannot occur.

Theorem: STRONG LAW OF LARGE NUMBERS (SLLN)

Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. RVs such that $\mathbb{E}|X_1| < \infty$. Put $S_n := X_1 + \cdots + X_n$. Then

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}X_1, \quad \text{as } n \to \infty.$$

There exists different proofs of the SLLN. Historically, the first proof was obtained under the assumption that $\mathbb{E}|X_1|^4 < \infty$, before being generalized to the weaker assumption $\mathbb{E}|X_1| < \infty$.

Consequence: $\forall \varepsilon > 0$, $\mathbb{P}(A_{n,\varepsilon} \text{ i.o.}) = 0$.

Back to frequency interpretation: let A be some event, and put

$$X_j = \begin{cases} 1 & \text{if } A \text{ occurs in the } j\text{-th replication of the random experiment} \\ 0 & \text{otherwise.} \end{cases}$$

Then $\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} = \frac{n_A}{n} = \frac{\#\text{ times } A \text{ occured}}{\#\text{ trials}}$.

Moreover, $\mathbb{E}X_1 = \mathbb{P}(A)$.

The SLLN states that $\frac{n_A}{n} \xrightarrow{a.s.} \mathbb{P}(A)$

Wow! This is exactly what we wanted to model: the main empirical fact about randome experiments: convergence of relative frequencies $\Rightarrow$ our theory must be very powerful; starting with only 3 axioms, we manage to reproduce a fundamental law of nature.

The SLLN states that $S_n = n\,\mathbb{E}X_1 + \underset{a.s.}{o}(n)$. Next, we would like to refine this result, and obtain an idea of the size of the error $\frac{S_n}{n} - \mathbb{E}X_1$.

$\Rightarrow$ We need a magnifying glass: scale the difference, and consider

$$b_n \left(\frac{S_n}{n} - \mathbb{E}X_1 \right) = \frac{b_n}{n} \left(S_n - n\mathbb{E}X_1 \right).$$

$\{b_n\}$ = diverging sequence with n. Which choice for $\{b_n\}$? It turns out that when the X_j are square integrable, the right choice is $b_n \sim n^{1/2}$.

II. 2. Central Limit Theorem.

Theorem (CENTRAL LIMIT THEOREM – CLT)

Let $\{X_n\}_{n\geq 1}$ = sequence of iid RVs with mean $\mu = \mathbb{E}X_1$ and variance $\sigma^2 = \mathrm{Var}\, X_1 < \infty$. Put $S_n := X_1 + \dots + X_n$.
Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $n \to +\infty$.

The normal distribution arises as a result of a "combined collective efforts" of a large number of "small contributions".

proof: First, standardize the variables and put $U_j = \frac{X_j - \mu}{\sigma}$.
Then the U_j are iid with zero mean and unit variance.
Moreover, $\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{U_1 + \dots + U_n}{\sqrt{n}}$.

Thus

$$\varphi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = \varphi_{\frac{U_1 + \dots + U_n}{\sqrt{n}}}(t) = \varphi_{U_1 + \dots + U_n}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left[\varphi_{U_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Expand $\varphi_{U_1}(s)$ around 0:

$$\varphi_{U_1}(s) = \underbrace{\varphi_{U_1}(0)}_{1} + \underbrace{\varphi'_{U_1}(0)}_{i\mathbb{E}U_1 = 0}\, s + \frac{1}{2}\underbrace{\varphi''_{U_1}(0)}_{-\mathbb{E}U_1^2 = -1}\, s^2 (1 + o(1))$$

$$= 1 - \frac{s^2}{2}(1 + o(1)).$$

$$\Rightarrow \varphi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = \left[1 - \frac{t^2}{2n}(1 + o(1))\right]^n$$

$$\to \exp\left(-\frac{t^2}{2}\right) = \text{ChF of a } \mathcal{N}(0,1) \quad \blacksquare$$

Remarks

(i) In practice, when n is large enough, we approximate the law of $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ by the law of a $\mathcal{N}(0,1)$ RV.
Likewise, we make use of $S_n \overset{d}{\approx} \mathcal{N}(n\mu, n\sigma^2)$.

(ii) In particular, when the X_j are $B(p)$, so that $S_n = X_1 + \dots + X_n \sim Bi(n,p)$, we obtain the approximation $Bi(n,p) \approx \mathcal{N}(np, np(1-p))$. This result, the first CLT, is known as the Moivre–Laplace theorem, and was proved by Moivre in 1733 for $p = 1/2$, and generalized by Laplace in 1812 for all $p \in (0,1)$.

(iii) This result is important in Mathematical Statistics to construct confidence intervals, or to perform hypothesis testing.

(iv) Various extensions of this version of the CLT hold by relaxing the assumption of "independent" to "not too correlated" or "identically distributed".

$\rightarrow$ LINDEBERG CLT: $\{X_n\}_{n\geq 1}$ = sequence of independent RVs, $\mathbb{E}X_n = 0$ and $\sigma_n^2 = \mathrm{Var}\, X_n < \infty$.
Put $s_n^2 = \sum_{k=1}^{n} \sigma_k^2$. If $\forall \varepsilon > 0$, holds

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left(X_k^2 \, \mathbb{1}\{|X_k| > \varepsilon s_n\}\right) = 0,$$

then $\frac{1}{s_n}(X_1 + \cdots + X_n) \xrightarrow{d} \mathcal{N}(0,1)$.

(v) A multivariate version of the CLT holds: if $\underline{X} = (X_1, \ldots, X_d)^t \in \mathbb{R}^d$ is a random vector with mean $\mathbb{E}\underline{X}$ and covariance matrix Σ, and if $\underline{X}_1, \ldots, \underline{X}_n$ is a sequence of independent random vectors all distributed like $\underline{X}$, then

$$\frac{1}{\sqrt{n}}\left(S_n - n\mathbb{E}\underline{X}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma), \text{ as } n \to \infty.$$

(vi) If $\mathbb{E}X_1^2 = \infty$, then we can still get a CLT-type result, but the magnifying sequence $\{b_n\}$ in $b_n\left(\frac{S_n}{n} - \mathbb{E}X_1\right)$ must be chosen differently. In addition, we require regular variation of the tails of F_X:

$H(x) := F_X(-x) + \left(1 - F_X(x)\right) = x^{-\alpha} L(x) \quad \alpha \in [1,2]$,

where $L(x)$ is a slowly varying function, satisfying $\forall v \; \frac{L(vx)}{L(x)} \to 1$. If in addition $\frac{F_X(-x)}{H(x)} \to$ constant, then the scaling should be $b_n = n^{1-\frac{1}{\alpha}} \ell(n)$, where ℓ is also slowly varying. The limiting distribution will be one of the STABLE LAWS, for which $\forall c_1, c_2 > 0$, $c_1 X_1 + c_2 X_2 \stackrel{d}{=} aX + b$ for some a, b, and X_1, X_2 are independent copies of X. Ex: Cauchy distrib.

In this configuration, since the tails have a power function behaviour, the main contributions to S_n come from a small number of the X_j's: the largest ones. This is the main difference with the CLT.

(vii) The rate of convergence in the CLT is provided by the Berry-Essen theorem, provided the third moment is finite:

Thm (BERRY-ESSEN)
Let $\{X_n\}_{n\geq 1}$ be a sequence of iid RVs with finite third moment: $\mathbb{E}X_1 = 0$, $\operatorname{Var} X_1 = \sigma^2$, $\mathbb{E}|X_1|^3 = \beta < \infty$. Put $S_n := X_1 + \cdots + X_n$, and let $F_n(x)$ denotes the distribution function of $\frac{S_n}{\sigma\sqrt{n}}$. Then there exists a constant $C < \infty$ such that

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{C\beta}{\sigma^3\sqrt{n}},$$

where Φ denotes the normal distribution function.

(viii) The next result is very useful in statistics, when dealing with the asymptotic normal distributions of estimators.

Thm (DELTA METHOD)
Let
- $Z_n := a_n(X_n - \theta) \xrightarrow{d} Z$, where $a_n \uparrow \infty$, $\theta \in \mathbb{R}$
- g = continuously differentiable function at θ.

Then
$$a_n\left(g(X_n) - g(\theta)\right) \xrightarrow{d} g'(\theta) Z, \text{ as } n \to \infty.$$

proof = Taylor expanding g around θ gives
$g(X_n) - g(\theta) = g'(\theta_n^*)(X_n - \theta)$, for some X_n^* between X_n and θ. We have

$$|\theta_n^* - \theta| \underset{\text{by def of } \theta_n^*}{<} |X_n - \theta| = a_n^{-1}|a_n(X_n - \theta)| = a_n^{-1}|Z_n| \xrightarrow{P} 0, \text{ where}$$

this follows from Slutsky theorem with $a_n^{-1} \xrightarrow{d} 0$, $|Z_n| \xrightarrow{d} |Z|$

Thus $\theta_n^* \xrightarrow{P} \theta$.

The continuous mapping theorem ensures that

$$g'(\theta_n^*) \xrightarrow{P} g'(\theta).$$

Thus

$$a_n(g(X_n) - g(\theta)) = \underbrace{a_n \, \underbrace{g'(\theta_n^*)}_{\substack{\downarrow P \\ g'(\theta)}} \, (X_n - \theta)}_{\substack{\downarrow d \\ Z}} \xrightarrow{d} g'(\theta) Z$$