Probability

8. Asymptotic Convergence

Ewen Gallic

Bachelor in Economics
International Program in Economics and Management
Academic Year 2026

amU FEG
Aix Marseille Université

Foreword

- In many applications, we study sequences of random variables:

$$(X_n)_{n\geq 1}.$$

- For example:
  - when we compute the empirical mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,
  - when we compute partial sums $S_n = \sum_{i=1}^{n} X_i$,
  - or, more generally, when we build an estimator $\hat{\theta}_n$ from a sample of size n.
- For each n, X_n has a distribution F_{X_n} and is a function of ω.
- The central question we ask is as follows: **how does the distribution of X_n evolve as n increases?**
  - does X_n concentrate around a constant?
  - does its variance shrink?
  - does its distribution approach a Gaussian law?
- **Asymptotic analysis** studies the **the limiting behaviour of the law of X_n as $n \to \infty$.**

Illustration

- Suppose we want to estimate the average height of students in this university.
- We collect a sample of size n and compute:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- If we increase n, does $\bar{X}_n$ stabilize? In other words:
  - does its variance decrease?
  - do large deviations from one possible random sample to another become unlikely?
  - does the distribution concentrate around a deterministic value (the population mean)?
- Through these questions, we seek to characterize the **limiting behaviour of the distribution** of $\bar{X}_n$ as $n \to \infty$.

Deterministic vs Random Convergence

To understand the notion of **asymptotic convergence**, let us first compare the **deterministic** and the **random** settings:

1. **Deterministic case:** for a sequence of real numbers (x_n),

$$x_n \to x \quad \Longleftrightarrow \quad |x_n - x| \to 0.$$

The distance $|x_n - x|$ is a number. Convergence simply means that this number becomes arbitrarily small.

2. **Random case:** for random variables (X_n),

$$|X_n - X| \quad \text{is itself a random variable.}$$

This "error" is no longer a number but a random variable with its own distribution.

So what does it mean for a **random error** to become small?

How Can a Random Error Become Small?

- When studying the convergence of X_n toward X, we define the **error** or **deviation** at step n as the random variable:

$$E_n := X_n - X.$$

- Its **magnitude** is measured by $|E_n| = |X_n - X|$.
- Since $|X_n - X|$ is random, smallness can be understood in different ways:

1. **Small with high probability:** large deviations become unlikely:

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0.$$

2. **Small on average:** the expected magnitude of the error vanishes:

$$\mathbb{E}(|X_n - X|^r) \to 0.$$

3. **Small in distribution:** the distribution (or law) of X_n approaches the distribution of X.

- These are different notions of convergence that will be developed in this chapter.

Roadmap of the Chapter

- In this chapter, we will investigate how random quantities behave asymptotically as $n \to \infty$. We will study:
  1. **Modes of convergence**
     - Convergence in probability
     - Almost sure convergence
     - Convergence in mean (L^r)
     - Convergence in distribution
  2. **Hierarchy of convergences**
  3. **Two fundamental limit theorems**
     - Weak law of large numbers
     - Central Limit Theorem
- These slides are adapted from lecture notes by Mathieu Faure, kindly shared.
- Supplementary reading: Chapter 7 of Hansen (2022), chapter 17 of Miller (2017), chapters 6 and 7 of Garnier et al. (2021).

1. Inequalities

A Little Detour

- In asymptotic analysis, as previously mentioned, we often want to control probabilities of the form:
$$\mathbb{P}(|X_n - X| > \varepsilon).$$
- Direct computation is usually impossible:
  - we may not know the full distribution,
  - the density may be complicated,
  - X_n may depend on many random variables.
- Instead, we use **inequalities** to bound probabilities using simpler quantities like:
  - expectations,
  - variances.
- These tools will allow us to prove convergence results without computing distributions explicitly.

1.1. Markov Inequality

Markov Inequality

Definition 1. Markov Inequality

Let Y be a nonnegative random variable ($Y \geq 0$) with finite expectation. Then, for all $\varepsilon > 0$,

$$\mathbb{P}(Y \geq \varepsilon) \leq \frac{\mathbb{E}(Y)}{\varepsilon}.$$

- Suppose Y measures a physical quantity (e.g., a height). Then ε has the same unit as Y, and $\frac{\mathbb{E}(Y)}{\varepsilon}$ is dimensionless, which is coherent, since it bounds a probability.
- If $\varepsilon < \mathbb{E}(Y)$, then $\frac{\mathbb{E}(Y)}{\varepsilon} > 1$. Since probabilities are always ≤ 1, the bound becomes trivial. In that case, Markov's inequality does not provide useful information.
- If $Y = \varepsilon$ a.s., then $\mathbb{P}(Y \geq \varepsilon) = 1$ and $\frac{\mathbb{E}(Y)}{\varepsilon} = \frac{\varepsilon}{\varepsilon} = 1$. In this case, Markov's inequality holds with equality.
- If $\varepsilon \gg \mathbb{E}(Y)$, then $\frac{\mathbb{E}(Y)}{\varepsilon}$ is small. Large deviations become unlikely.

Proof of Markov's Inequality (Continuous Case)

Assume $Y \geq 0$ is a continuous random variable with density f_Y, and let $\varepsilon > 0$. Note that if $y \geq \varepsilon$, then $y/\varepsilon \geq 1$.

$$\begin{aligned}
\mathbb{P}(Y \geq \varepsilon) &= \int_{\varepsilon}^{\infty} f_Y(y)\, dy = \int_{\varepsilon}^{\infty} 1 \cdot f_Y(y)\, dy \\
&\leq \int_{\varepsilon}^{\infty} \frac{y}{\varepsilon} f_Y(y)\, dy \quad \text{since } y \geq \varepsilon \Rightarrow \frac{y}{\varepsilon} \geq 1 \\
&= \frac{1}{\varepsilon} \int_{\varepsilon}^{\infty} y f_Y(y)\, dy \\
&\leq \frac{1}{\varepsilon} \int_{0}^{\infty} y f_Y(y)\, dy = \frac{\mathbb{E}(Y)}{\varepsilon}.
\end{aligned}$$

Thus,

$$\mathbb{P}(Y \geq \varepsilon) \leq \frac{\mathbb{E}(Y)}{\varepsilon}.$$

Example: Annual Household Income

Application of Markov's Inequality

Let Y be the annual income (in euros) of a randomly chosen household. Assume that:

$$\mathbb{E}(Y) = 30{,}000.$$

1. *What is the probability that an individual, selected at random, earns more than €60,000?*
2. *At least €1,000,000?*

Solution

1. We apply Markov with $\varepsilon = 60{,}000$:

$$\mathbb{P}(Y \geq 60{,}000) \leq \frac{30{,}000}{60{,}000} = \frac{1}{2}.$$

Thus,

$$\mathbb{P}(Y \geq 60{,}000) \leq 50\%.$$

2. We now apply Markov with $\varepsilon = 1{,}000{,}000$:

$$\mathbb{P}(Y \geq 1{,}000{,}000) \leq \frac{30{,}000}{1{,}000{,}000} = 0.03.$$

So,

$$\mathbb{P}(Y \geq 1{,}000{,}000) \leq 3\%.$$

1.2. Bienaymé–Chebyshev Inequality

Bienaymé–Chebyshev Inequality

- Markov's inequality bounds tail probabilities using only $\mathbb{E}(Y)$ (and requires $Y \geq 0$).
- If we also know the **variance**, we can control deviations around the mean more effectively, and obtain tighter bounds.
- We will not assume, however, that the random variable is nonnegative.

Definition 2. Bienaymé–Чебышёв Inequality

Let X be a random variable with finite mean and finite variance. Then, for any $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Interpretation

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

- The inequality controls the probability of **large deviations** from the mean.
- The bound depends on two quantities:
  - the variance (size of fluctuations),
  - the deviation level ε.
- If $\mathrm{Var}(X)$ is small, large deviations must be rare.
- If ε increases, the upper bound decreases like $\frac{1}{\varepsilon^2}$.
- Note that Bienaymé–Chebyshev inequality is distribution-free: it holds for any random variable with finite variance.

When Is the Bound Informative?

- If ε is small,
$$\frac{\mathrm{Var}(X)}{\varepsilon^2}$$
may exceed 1, so the bound is trivial (a probability is at most 1).
- The inequality becomes informative when
$$\varepsilon^2 > \mathrm{Var}(X).$$
- Chebyshev becomes powerful when variance shrinks, and this is an important characteristic that we will use for proving convergence results.

Proof of Bienaymé–Chebyshev Inequality

- Let X be a random variable with finite mean ($\mathbb{E}(X) < \infty$) and finite variance ($\mathrm{Var}(X) < \infty$). Fix $\varepsilon > 0$. Set
$$Y := (X - \mathbb{E}(X))^2.$$
- Then $Y \geq 0$ a.s. and $\mathbb{E}(Y) = \mathrm{Var}(X)$.
- Note that
$$|X - \mathbb{E}(X)| \geq \varepsilon \quad \Longleftrightarrow \quad (X - \mathbb{E}(X))^2 \geq \varepsilon^2.$$
- Hence,
$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) = \mathbb{P}\big((X - \mathbb{E}(X))^2 \geq \varepsilon^2\big) = \mathbb{P}(Y \geq \varepsilon^2).$$
- Since $Y \geq 0$, Markov's inequality gives
$$\mathbb{P}(Y \geq \varepsilon^2) \leq \frac{\mathbb{E}(Y)}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$
- Thus,
$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

1.3. Jansen Inequality

Why Jensen's Inequality?

- Markov controls probabilities using expectations.
- Bienaymé–Chebyshev controls deviations using variance.
- Jensen's inequality relates **convexity** and **expectation**.
- It allows us to compare:

$$\mathbb{E}(\varphi(X)) \quad \text{and} \quad \varphi(\mathbb{E}(X)).$$

- We will rely on this inequality for studying L^r convergence (convergence in mean).

Jensen's Inequality

Definition 3. Jensen's Inequality

*Let X be an integrable random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ a **convex** function. Then,*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

*If φ is **concave**, the inequality is reversed:*

$$\mathbb{E}(\varphi(X)) \leq \varphi(\mathbb{E}(X)).$$

Note: The Square Function

Special case: The Square Function

- *Consider* $\varphi(x) = x^2$, *which is convex.*
- *Jensen gives:*
$$(\mathbb{E}(X))^2 \leq \mathbb{E}(X^2).$$
- *Rewriting:*
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \geq 0.$$
- *Jensen immediately implies that variance is nonnegative.*

Geometric Interpretation

- A convex function lies above its tangent lines.
- The expectation is a weighted average.
- Jensen's inequality says:

Applying a convex function after averaging

is smaller than

averaging after applying the convex function.

Proof of Jensen's Inequality (Convex Case)

- Let $\mu = \mathbb{E}(X)$. Since g is convex, it lies above any of its tangent lines.
- Let
$$\ell(x) = a + bx$$
be the tangent line to g at $x = \mu$. Then
$$g(x) \geq \ell(x) = a + bx \qquad \text{for all } x \in \mathbb{R},$$
and
$$g(\mu) = \ell(\mu) = a + b\mu.$$
- Apply this inequality with $x = X$ and take expectations:
$$\mathbb{E}\big[g(X)\big] \geq \mathbb{E}[a + bX] = a + b\,\mathbb{E}[X] = a + b\mu = g(\mu) = g\big(\mathbb{E}[X]\big).$$
- Hence $g\big(\mathbb{E}[X]\big) \leq \mathbb{E}\big[g(X)\big]$.

2. Modes of Convergence

2.1. Convergence in Probability

Definition

Definition 4. Convergence in Probability

A sequence (X_n) converges in probability to X if

$$\forall \varepsilon > 0, \qquad \lim_{n \to \infty} \mathbb{P}\big(|X_n - X| > \varepsilon\big) = 0.$$

Intuitively, this means that large deviations become unlikely.

Notation

We will write:

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

Convergence in probability also appear, in some textbooks, using the plim operator:

$$\operatorname*{plim}_{n \to \infty} X_n = X.$$

Interpretation

How should we interpret this definition?

- Fix $\varepsilon > 0$.
- $\mathbb{P}(|X_n - X| > \varepsilon)$ is the probability that the error exceeds ε.
- Convergence in probability means that, for any precision level ε, the probability of being far from X goes to zero.
- The randomness does not disappear, but large errors become rare.

Example 1: Convergence in Probability (1/2)

Example

Let (X_n) be a sequence of random variables defined by

$$X_n = \begin{cases} 1 & \text{with probability } \frac{1}{n}, \\ 0 & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

The sequence X_n converges in probability to 0.

- Intuitively, the event $X_n = 1$ becomes rarer and rarer as n increases.
- Let $\varepsilon > 0$. We evaluate $\mathbb{P}(|X_n - 0| > \varepsilon)$.
- Since $|X_n - 0| = |X_n|$, the event becomes $|X_n| > \varepsilon$.
- Because $X_n \in \{0, 1\}$, if $0 < \varepsilon < 1$:
  - if $X_n = 0$, then $0 > \varepsilon$ is false,
  - if $X_n = 1$, then $1 > \varepsilon$ is true.

Example 1: Convergence in Probability (2/2)

- Hence
$$\mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(X_n = 1) = \frac{1}{n}.$$
- Since
$$\frac{1}{n} \longrightarrow 0,$$
we conclude that
$$X_n \xrightarrow{P} 0.$$
- This means that:
  - Most of the time, $X_n = 0$.
  - The probability that X_n differs from 0 becomes smaller and smaller as n increases.
  - Convergence in probability means that, as n increases,
$$\mathbb{P}(|X_n - 0| > \varepsilon) \to 0.$$

Example 2: Convergence in Probability to 0

Example

Let $(X_n)_{n\geq 1}$ be a sequence of random variables with density

$$f_{X_n}(x) = \begin{cases} ne^{-nx} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Let us show that $X_n \xrightarrow{P} 0$.

- Let $\varepsilon > 0$. Since $X_n \geq 0$ (the density is 0 for all $x < 0$, hence $\mathbb{P}(X \geq 0) = 1$),
$$\mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(X_n > \varepsilon) = \int_{\varepsilon}^{+\infty} ne^{-nx}\, dx.$$
- Compute:
$$\int_{\varepsilon}^{+\infty} ne^{-nx}\, dx = \left[-e^{-nx}\right]_{\varepsilon}^{+\infty} = e^{-n\varepsilon}.$$
- Since $e^{-n\varepsilon} \to 0$ as $n \to +\infty$, we conclude that
$$\mathbb{P}(|X_n| > \varepsilon) \to 0, \quad \text{hence} \quad X_n \xrightarrow{P} 0.$$

A Useful Criterion for Convergence in Probability

Proposition 1. Convergence via Mean and Variance

Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that

$$\lim_{n\to+\infty} \mathbb{E}(X_n) = a, \qquad \lim_{n\to+\infty} \mathrm{Var}(X_n) = 0.$$

Then

$$X_n \xrightarrow{P} a.$$

Moment-Based Convergence Criterion

If the mean and variance exist, this result gives a practical criterion for establishing convergence in probability of a sequence $(X_n)_n$ based solely on its first two moments.

Proof (at home)

- Fix $\varepsilon > 0$.
- Write (triangle inequality: $|u+v| \leq |u| + |v|$)
$$|X_n - a| \leq |X_n - \mathbb{E}(X_n)| + |\mathbb{E}(X_n) - a|.$$
- Hence,
$$\mathbb{P}(|X_n - a| > \varepsilon) \leq \mathbb{P}\left(|X_n - \mathbb{E}(X_n)| > \frac{\varepsilon}{2}\right) \quad \text{for } n \text{ large enough.}$$
- By Chebyshev's inequality,
$$\mathbb{P}\left(|X_n - \mathbb{E}(X_n)| > \frac{\varepsilon}{2}\right) \leq \frac{\mathrm{Var}(X_n)}{(\varepsilon/2)^2} = \frac{4\ \mathrm{Var}(X_n)}{\varepsilon^2}.$$
- Since $\mathrm{Var}(X_n) \to 0$, the right-hand side tends to 0.
- Therefore,
$$\mathbb{P}(|X_n - a| > \varepsilon) \to 0,$$
- which proves that
$$X_n \xrightarrow{P} a.$$

Corollary

Corollary 1. Moment Criterion Around a Limit

Let $(X_n)_{n\geq 1}$ *and* X *be random variables such that*

$$\lim_{n\to+\infty} \mathbb{E}(X_n - X) = 0, \qquad \lim_{n\to+\infty} \mathrm{Var}(X_n - X) = 0.$$

Then

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

Proof

- Set $Y_n := X_n - X$. Then, $\mathbb{E}(Y_n) \to 0$ and $\mathrm{Var}(Y_n) \to 0$.
- By the previous proposition, $Y_n \xrightarrow{P} 0$.
- Since $X_n - X \xrightarrow{P} 0$ is equivalent to $X_n \xrightarrow{P} X$, the result follows.

□

Remark: Translation Invariance

For any random variables X_n and X,

$$X_n \xrightarrow{P} X \quad \Longleftrightarrow \quad X_n - X \xrightarrow{P} 0,$$

Indeed, if $X_n \xrightarrow{P} X$, by definition, $\forall \varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. But $|(X_n - X) - 0| = |X_n - X|$. Hence, $\mathbb{P}(|(X_n - X) - 0| > \varepsilon) = \mathbb{P}(|X_n - X| > \varepsilon)$.

Slutsky's Theorem

Theorem 1. Slutsky's Theorem (Basic Version)

Let $(X_n)_n$ and $(Y_n)_n$ be sequences of random variables such that

$$X_n \xrightarrow{P} X, \qquad Y_n \xrightarrow{P} c,$$

where c is a constant.
Then:

$$X_n + Y_n \xrightarrow{P} X + c,$$

$$X_n Y_n \xrightarrow{P} cX,$$

and, if $c \neq 0$,

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{c}.$$

Proof of the Addition Case (1/2)

Fix $\varepsilon > 0$. By the triangle inequality,

$$|(X_n + Y_n) - (X + c)| = |(X_n - X) + (Y_n - c)| \leq |X_n - X| + |Y_n - c|.$$

Therefore, if $|(X_n + Y_n) - (X + c)| > \varepsilon$, then necessarily

$$|X_n - X| + |Y_n - c| > \varepsilon.$$

Since the sum is larger than ε, at least one of the two terms need to be larger than $\varepsilon/2$. Hence, either $|X_n - X| > \varepsilon/2$ or $|Y_n - c| > \varepsilon/2$. Hence, we have the inclusion of events:

$$\Big\{ |(X_n + Y_n) - (X + c)| > \varepsilon \Big\} \subset \Big\{ |X_n - X| > \varepsilon/2 \Big\} \cup \Big\{ |Y_n - c| > \varepsilon/2 \Big\}.$$

Taking probabilities and using $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$,

$$\mathbb{P}\Big(|(X_n + Y_n) - (X + c)| > \varepsilon \Big) \leq \mathbb{P}\Big(|X_n - X| > \varepsilon/2 \Big) + \mathbb{P}\Big(|Y_n - c| > \varepsilon/2 \Big).$$

Proof of the Addition Case (2/2)

Now, since $X_n \xrightarrow{\mathbb{P}} X$,

$$\lim_{n\to\infty} \mathbb{P}\Big(|X_n - X| > \varepsilon/2 \Big) = 0,$$

and since $Y_n \xrightarrow{\mathbb{P}} c$,

$$\lim_{n\to\infty} \mathbb{P}\Big(| Y_n - c | > \varepsilon/2 \Big) = 0.$$

Hence the sum of these two real numbers converges to 0, and by the bound above,

$$\lim_{n\to\infty} \mathbb{P}\Big(| (X_n + Y_n) - (X + c) | > \varepsilon \Big) = 0.$$

This is exactly $X_n + Y_n \xrightarrow{\mathbb{P}} X + c$.

2.2. Almost Sure Convergence

Almost Sure Convergence

- An event that may depend on randomness but occurs with probability 1 is said to occur **almost surely**.

Definition 5. Almost Sure Convergence

A sequence of random variables $(X_n) \in \mathbb{R}$ converges almost surely (or almost everywhere, or with probability 1) to X if

$$\mathbb{P}\Big(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\Big) = 1.$$

Notation

We note this $X_n \xrightarrow{a.s.} X$

- In other words, for *almost every* outcome ω, the sequence of real numbers $X_1(\omega), X_2(\omega), \ldots$ converges to $X(\omega)$.

Example: Estimating a Probability (1/3)

Example

Suppose we repeatedly perform the same experiment (e.g., tossing a coin). Let p denote the probability of Heads.
The empirical proportion of Heads, $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, converges a.s. to p.

- Here,

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th toss is Heads,} \\ 0 & \text{otherwise.} \end{cases}$$

- For one particular sequence of tosses, we obtain numbers

$$X_1(\omega),\, X_2(\omega),\, X_3(\omega), \ldots$$

and compute the sequence of proportions

$$\hat{p}_1(\omega),\, \hat{p}_2(\omega),\, \hat{p}_3(\omega), \ldots$$

Example: Estimating a Probability (2/3)

- For example, if the outcome ω is $(H, T, H, H, \ldots)$), we obtain:

$$1, 0, 1, 1, \ldots$$

- For the first four values, the numerical sequence we have is:

$$\hat{p}_1(\omega) = 1,\ \hat{p}_2(\omega) = 0.5,\ \hat{p}_3(\omega) = \tfrac{2}{3}, \hat{p}_4(\omega) = \tfrac{3}{4}, \ldots$$

- As we keep increasing n, the empirical proportions may look like:

n	empirical proportion $\hat{p}_n$
1	1.00
5	0.60
10	0.50
20	0.55
50	0.52
100	0.51
500	0.503

Example: Estimating a Probability (3/3)

- As n increases, the empirical proportion tends to get closer to the true probability p.
- Almost sure convergence means that for almost every possible sequence of tosses ω,

$$\hat{p}_n(\omega) \longrightarrow p,$$

- In other words, this stabilization occurs for **almost every sequence** of tosses, except for extremely pathological ones whose probability is zero
  - for instance a sequence consisting only of Heads,
  - almost sure convergence allows the property to fail on such sequences.

Almost Sure vs. Convergence in Probability

Proposition 2.

If

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

then

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

- Almost sure convergence is therefore a **stronger notion** than convergence in probability.

Some Reminders

Before turning to the proof of the proposition, some reminders about the convergence of numerical sequences may be useful.

Reminder: Convergence of numerical sequences

If a sequence of numbers a_n converges to a, then for every $\varepsilon > 0$ there exists an index N such that for all $n \geq N$,

$$|a_n - a| \leq \varepsilon.$$

In other words, once n is large enough, the sequence stays within a distance ε of its limit.

Some Reminders: Example

Example

Consider the numerical sequence

$$a_n = \frac{1}{n}.$$

We know that $a_n \to 0$.
For example, if $\varepsilon = 0.1$, *we can choose* $N = 10$ *since*

$$\left| \frac{1}{n} - 0 \right| \leq 0.1 \quad \textit{for all } n \geq 10.$$

Proof: $X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{P} X$ (1/2)

Assume that $X_n \xrightarrow{\text{a.s.}} X$.

- By definition, the event

$$A = \{\omega \in \Omega : \ X_n(\omega) \to X(\omega)\}$$

  satisfies $\mathbb{P}(A) = 1$.
- Fix $\varepsilon > 0$. For any $\omega \in A$, the numerical sequence

$$X_1(\omega),\, X_2(\omega),\, X_3(\omega), \ldots$$

  converges to $X(\omega)$.
- Using the reminder on numerical sequences, this implies that for each $\omega \in A$ there exists an index $N(\omega)$ such that for all $n \geq N(\omega)$,

$$|X_n(\omega) - X(\omega)| \leq \varepsilon.$$

Proof: $X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{P} X$ (2/2)

- Equivalently, for each $\omega \in A$, the event

$$|X_n(\omega) - X(\omega)| > \varepsilon$$

can only occur for finitely many n (it may happen a few times at the beginning, for small n, but eventually stops happening).

- Therefore, the probability of observing a deviation larger than ε must vanish:

$$\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0.$$

Since this holds for every $\varepsilon > 0$, we conclude that

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

□

2.3. Convergence in Mean (L^r)

Convergence in Mean

- Another way to measure how close two random variables are is to look at the **expected size of their difference**.
- Instead of controlling probabilities such as

$$\mathbb{P}(|X_n - X| > \varepsilon),$$

we study the expected error:

$$\mathbb{E}(|X_n - X|^r)\,.$$

- This leads to the notion of **convergence in mean** (or L^r **convergence**).

Definition: Convergence in Mean

Definition 6. Convergence in Mean of Order p

Let $r \geq 1$. A sequence (X_n) converges in mean of order r to X if

$$\lim_{n \to \infty} \mathbb{E}(|X_n - X|^r) = 0.$$

Notation

We write

$$X_n \xrightarrow{L^r} X.$$

Example

Let $X_n = \frac{U}{n}$, where $U \sim \text{Uniform}(0,1)$. We can show that $X_n \xrightarrow{L^2} 0$.

- We have
$$|X_n - 0|^2 = \frac{U^2}{n^2}.$$
- Hence
$$\mathbb{E}(X_n^2) = \frac{\mathbb{E}(U^2)}{n^2}.$$
- Since $\mathbb{E}(U^2) = \int_0^1 u^2 \cdot \frac{1}{b-a}\, du = \int_0^1 u^2\, du = \left[\frac{u^3}{3}\right]_0^1 = \frac{1}{3}$,
$$\mathbb{E}(X_n^2) = \frac{1}{3n^2} \to 0.$$
- Therefore
$$X_n \xrightarrow{L^2} 0.$$

Relation with Convergence in Probability

Proposition 3.

If

$$X_n \xrightarrow{L^r} X$$

for some $r \geq 1$, *then*

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

Convergence in Mean Square

We will particularly be interested in the case where X_n *converges in* r*-th mean to* X *for* $r = 2$.
In such a case, we say that X_n *converges in mean square (or in quadratic mean) to* X.

Proof: L^r Convergence $\Rightarrow$ Convergence in Probability

$$\text{Assume} \quad X_n \xrightarrow{L^r} X, \qquad \text{i.e.,} \qquad \mathbb{E}(|X_n - X|^r) \xrightarrow[n\to+\infty]{} 0.$$

- Fix $\varepsilon > 0$. Then, consider the nonnegative random variable $Y_n = |X_n - X|^r$.
- Since $Y_n \geq 0$, we can apply **Markov's inequality**, using $|X_n - X| \geq \varepsilon \iff |X_n - X|^r \geq \varepsilon^r$:

$$\mathbb{P}(Y_n \geq \varepsilon^r) \leq \frac{\mathbb{E}(Y_n)}{\varepsilon^r}.$$

- But,
$$\mathbb{P}(Y_n \geq \varepsilon^r) = \mathbb{P}(|X_n - X|^r \geq \varepsilon^r) = \mathbb{P}(|X_n - X| \geq \varepsilon).$$
- Therefore
$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|^r)}{\varepsilon^r}.$$
- Since $\mathbb{E}(|X_n - X|^r) \to 0$, the right-hand side tends to 0.
- Hence,
$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0, \qquad \text{so} \qquad X_n \xrightarrow{P} X. \quad \square$$

Exercise

Exercise

Let $(X_n)_n$ be a sequence of random variables defined by

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}, \qquad \mathbb{P}(X_n = n) = \frac{1}{n}.$$

Study the following modes of convergence of $(X_n)_n$ toward 0:

1. *Convergence in probability;*
2. *Convergence in quadratic mean (that is, in L^2).*

Solution: Convergence in Probability

1. We study the convergence in probability: we want to know if X_n converges to 0, in probability. Let $\varepsilon > 0$. Since X_n only takes the values 0 and n, for $n > \varepsilon$ we have

$$\{|X_n - 0| > \varepsilon\} = \{X_n = n\}.$$

Hence

$$\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = n) = \frac{1}{n}.$$

Since

$$\frac{1}{n} \longrightarrow 0,$$

we conclude that

$$X_n \xrightarrow{P} 0.$$

Solution: Convergence in Quadratic Mean

2 Now, we study the convergence in quadratic mean. We want to know if X_n converges to 0 in quadratic mean.

We compute

$$\mathbb{E}(|\ X_n\ |^2) = \mathbb{E}(X_n^2) = 0^2 \left(1 - \frac{1}{n}\right) + n^2 \cdot \frac{1}{n} = n.$$

Therefore

$$\mathbb{E}\big((X_n - 0)^2\big) = \mathbb{E}(X_n^2) = n.$$

Since

$$n \not\to 0,$$

we conclude that $(X_n)_n$ does not converge to 0 in quadratic mean.

From Quadratic Mean to Mean Convergence

Proposition 4.

If

$$X_n \xrightarrow{L^2} X,$$

then

$$X_n \xrightarrow{L^1} X.$$

Remark

More generally, if $r' \geq r \geq 1$,

$$X_n \xrightarrow{L^{r'}} X \quad \Longrightarrow \quad X_n \xrightarrow{L^r} X.$$

Proof of the Proposition

We assume that

$$X_n \xrightarrow{L^2} X, \qquad \text{i.e.} \qquad \mathbb{E}\Big((X_n - X)^2\Big) \xrightarrow[n\to+\infty]{} 0.$$

- Apply Jensen's inequality to the convex function $\varphi(x) = x^2$ and to the nonnegative random variable $|X_n - X|$:

$$(\mathbb{E}(|X_n - X|))^2 \leq \mathbb{E}\Big((X_n - X)^2\Big).$$

- Hence

$$\mathbb{E}(|X_n - X|) \leq \sqrt{\mathbb{E}((X_n - X)^2)}.$$

- Since $\mathbb{E}\big((X_n - X)^2\big) \to 0$, the right-hand side tends to 0.
- Therefore

$$\mathbb{E}(|X_n - X|) \to 0,$$

which means that $X_n \xrightarrow{L^1} X$.

Counterexample: L^1 Does Not Imply L^2

Example

Let $(X_n)_n$ be a sequence of random variables defined by

$$\mathbb{P}(X_n = n) = \frac{1}{n^2}, \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2}.$$

We study the convergence of $(X_n)_n$ toward 0.

- **Convergence in mean.**

$$\mathbb{E}(|X_n|) = n \cdot \frac{1}{n^2} + 0 \cdot \left(1 - \frac{1}{n^2}\right) = \frac{1}{n} \longrightarrow 0.$$

Hence

$$X_n \xrightarrow{L^1} 0.$$

- **Convergence in quadratic mean.**

$$\mathbb{E}(X_n^2) = n^2 \cdot \frac{1}{n^2} + 0^2 \cdot \left(1 - \frac{1}{n^2}\right) = 1.$$

Therefore, $\mathbb{E}((X_n - 0)^2) = 1$, which does not tend to 0.
Thus, $X_n \xrightarrow{L^2} 0$.

2.4. Convergence in Distribution

Convergence in Distribution

Definition 7. Convergence in distribution

Let $(X_n)_{n\geq 1}$ and X be real random variables with distribution functions F_{X_n} and F_X. We say that X_n converges in distribution to X if

$$F_{X_n}(x) \to F_X(x) \quad \text{for every continuity point } x \text{ of } F_X.$$

Notation

We write

$$X_n \xrightarrow{d} X \qquad \text{or} \qquad X_n \xrightarrow{\mathcal{L}} X.$$

This means that the distributions of X_n approach the distribution of X.

Example: Convergence in Distribution (1/2)

Let

$$X_n \sim \mathcal{U}(-1/n, 1/n).$$

Then,

$$F_{X_n}(x) = \begin{cases} 0 & x < -\frac{1}{n} \\ \dfrac{x + \frac{1}{n}}{2/n} & -\frac{1}{n} \leq x \leq \frac{1}{n} \\ 1 & x > \frac{1}{n} \end{cases}$$

Let $X = 0$ (a degenerate random variable). Its CDF is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

We can show that $F_{X_n}(x) \to F_X(x)$ for every continuity point of F_X, hence, $X_n \xrightarrow{d} 0$.

Example: Convergence in Distribution (2/2)

- If $x < 0$, then for n large enough we have $x < -\frac{1}{n}$. Hence
$$F_{X_n}(x) = 0 \quad \text{for all sufficiently large } n,$$
so $F_{X_n}(x) \to 0 = F_X(x)$.
- If $x > 0$, then for n large enough we have $x > \frac{1}{n}$. Hence
$$F_{X_n}(x) = 1 \quad \text{for all sufficiently large } n,$$
so $F_{X_n}(x) \to 1 = F_X(x)$.
- At $x = 0$, we have
$$F_{X_n}(0) = \frac{0 + \frac{1}{n}}{2/n} = \frac{1}{2},$$
whereas $F_X(0) = 1$. So there is no convergence at $x = 0$. However, this is not a problem, because F_X is discontinuous at $x = 0$.

Since $F_{X_n}(x) \to F_X(x)$ for every continuity point x of F_X, we conclude that $X_n \xrightarrow{d} 0$.

Relation with Other Modes of Convergence

Proposition 5. Convergence in Probability ⇒ Convergence in Distribution

If

$$X_n \xrightarrow{P} X,$$

then

$$X_n \xrightarrow{d} X.$$

- We will admit this proposition.
- The converse is generally false (see example on next slides)
- Convergence in distribution is the **weakest** of the common modes of convergence.

Example (1/3)

Example: Convergence in Distribution Does Not Imply Convergence in Probability

Let X be a random variable such that

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}.$$

Define the sequence $(X_n)_n$ by

$$X_{2k+1} = X, \qquad X_{2k} = -X,$$

i.e.,

$$X_1 = X, \qquad X_2 = -X, \qquad X_3 = X, \qquad X_4 = -X, \qquad \dots$$

Example (2/3)

Let us show that X_n converges in distribution to X.

- X and $-X$ have the same distribution:

$$\mathbb{P}(-X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}, \quad \text{and} \quad \mathbb{P}(-X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

- Hence, $X_n \sim X$ for all n.
- Therefore, for every $x \in \mathbb{R}$,

$$F_{X_n}(x) = F_X(x).$$

- Hence $F_{X_n}(x) \to F_X(x)$ for every continuity point of F_X, so

$$X_n \xrightarrow{d} X.$$

Example (3/3)

However, (X_n) does not converge to X in probability.

- For every $k \geq 1$, $X_{2k} = -X$.
- Thus
$$|X_{2k} - X| = |-X - X| = 2|X| = 2.$$
- Therefore, if $0 < \varepsilon < 2$,
$$\mathbb{P}(|X_{2k} - X| > \varepsilon) = 1 \neq 0.$$
- So the sequence
$$\mathbb{P}(|X_n - X| > \varepsilon)$$
does not tend to 0.
- Hence
$$X_n \xrightarrow{P} X.$$

Two Useful Special Cases

Proposition 6.

Let $a \in \mathbb{R}$, and let $(X_n)_n$ and $(Y_n)_n$ be sequences of random variables.

1. *If $X_n \xrightarrow{d} a$, then*
$$X_n \xrightarrow{P} a.$$
2. *If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} a$, then*
$$X_n + Y_n \xrightarrow{d} X + a \qquad \text{and} \qquad X_n Y_n \xrightarrow{d} aX.$$

Two Useful Special Cases: Interpretation

1. **First point.**
   - A constant a has a degenerate distribution: all its mass is at one point.
   - If X_n converges in distribution to a, the distribution of X_n must become increasingly concentrated around a. Hence large deviations become unlikely, which explains convergence in probability.
2. **Second point.**
   - If Y_n converges in distribution to a constant a, then for large n the random variable Y_n behaves approximately like the constant a.
   - Therefore
   $$X_n + Y_n \approx X_n + a, \qquad X_n Y_n \approx a X_n,$$
   which explains why the limiting distributions are $X + a$ and aX.

Proof of Point 1 (1/2)

- Since $X_n \xrightarrow{d} a$, we have $F_{X_n}(x) \to F_a(x)$ for every continuity point of F_a.
- The CDF of the degenerate random variable equal to a is

$$F_a(x) = \begin{cases} 0 & x < a, \\ 1 & x \geq a. \end{cases}$$

- Let $\varepsilon > 0$. Then $a - \varepsilon < a < a + \varepsilon$, and the points $a - \varepsilon$ and $a + \varepsilon$ are continuity points of F_a.
- We have

$$\mathbb{P}(|X_n - a| > \varepsilon) = \mathbb{P}(X_n < a - \varepsilon) + \mathbb{P}(X_n > a + \varepsilon).$$

- Since

$$\mathbb{P}(X_n < a - \varepsilon) = F_{X_n}(a - \varepsilon) \quad \text{and} \quad \mathbb{P}(X_n > a + \varepsilon) = 1 - F_{X_n}(a + \varepsilon),$$

it follows that

$$\mathbb{P}(|X_n - a| > \varepsilon) = F_{X_n}(a - \varepsilon) + 1 - F_{X_n}(a + \varepsilon).$$

Proof of Point 1 (2/2)

- By convergence in distribution,

$$F_{X_n}(a-\varepsilon) \to F_a(a-\varepsilon) = 0$$

and

$$F_{X_n}(a+\varepsilon) \to F_a(a+\varepsilon) = 1.$$

- Therefore

$$\mathbb{P}(|X_n - a| > \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

Since this holds for every $\varepsilon > 0$, we conclude that

$$X_n \xrightarrow{P} a.$$

□

Continuous Mapping Theorem

Theorem 2. Continuous Mapping Theorem

Let $(X_n)_n$ be a sequence of random variables such that $X_n \xrightarrow{d} X$. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then

$$f(X_n) \xrightarrow{d} f(X).$$

- Convergence in distribution means that the distributions of X_n become closer and closer to the distribution of X.
- If we apply a **continuous transformation** f to these random variables, the transformation does not create sudden jumps or distortions.
- Therefore, if X_n behaves approximately like X for large n, then $f(X_n)$ should behave approximately like $f(X)$.
- Hence, this theorem says that **continuous functions preserve convergence in distribution**.

Convergence in Distribution in the Discrete Case

Proposition 7.

Let $(X_n)_n$ be a sequence of random variables taking values in a discrete set E, and let X be a random variable with support in E. Then,

$$\forall x \in E, \qquad \lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x) \quad \Longleftrightarrow \quad X_n \xrightarrow{d} X.$$

- In the discrete case, convergence in distribution simply means that the probability masses converge at each point of the support.

Convergence in Distribution in the Continuous Case

Proposition 8.

Let $(X_n)_n$ be a sequence of continuous random variables with densities f_{X_n} on $\mathbb{R}$, and let X be a random variable with density f_X. Then,

$$\forall x \in \mathbb{R}, \qquad \lim_{n\to\infty} f_{X_n}(x) = f_X(x) \quad \Longrightarrow \quad X_n \xrightarrow{d} X.$$

- For continuous random variables, the distribution is completely determined by the density function.
- If the densities f_{X_n} converge pointwise to the density f_X, then the probability mass around every point x behaves more and more like that of X. As a result, the cumulative distribution functions F_{X_n} approach F_X, which implies

$$X_n \xrightarrow{d} X.$$

Counterexample (1/3)

Be careful, a discrete sequence that converges in distribution does not necessarily converge to a discrete random variable.

Example: A Discrete Sequence May Converge to a Continuous Limit

Let $\theta > 0$, and for each $n \geq 1$, let X_n be a geometric random variable with parameter θ/n, i.e.

$$\mathbb{P}(X_n = k) = \left(1 - \frac{\theta}{n}\right)^{k-1} \frac{\theta}{n}, \qquad k \in \mathbb{N}^*.$$

Define

$$Y_n := \frac{X_n}{n}.$$

- Each Y_n is still a **discrete** random variable, since it takes values in $\left\{\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots\right\}$.
- We will show that $Y_n \xrightarrow{d} Y$, where Y has an exponential distribution with parameter θ.

Counterexample (2/3)

Let $x \in \mathbb{R}$. Then

$$F_{Y_n}(x) = \mathbb{P}\left(\frac{X_n}{n} \leq x\right) = \mathbb{P}(X_n \leq nx).$$

Since X_n is geometric,

$$\mathbb{P}(X_n \leq m) = 1 - \left(1 - \frac{\theta}{n}\right)^m, \qquad m \in \mathbb{N}^*.$$

Hence, for $x \geq 0$,

$$F_{Y_n}(x) = 1 - \left(1 - \frac{\theta}{n}\right)^{\lfloor nx \rfloor},$$

where $\lfloor \cdot \rfloor$ denotes the integer part.

Counterexample (3/3)

- If $x < 0$, then $F_{Y_n}(x) = 0$.
- If $x \geq 0$, then

$$\left(1 - \frac{\theta}{n}\right)^{\lfloor nx \rfloor} \longrightarrow e^{-\theta x}.$$

Therefore

$$F_{Y_n}(x) \longrightarrow \begin{cases} 0 & x < 0, \\ 1 - e^{-\theta x} & x \geq 0. \end{cases}$$

We recognize the CDF of an exponential random variable with parameter θ. Thus

$$Y_n \xrightarrow{d} \mathrm{Exp}(\theta).$$

Hence, a sequence of discrete random variables may converge in distribution to a continuous random variable.

2.5. Summary

Summary of Convergence Notions

Convergence	Definition	Intuition
Almost sure	$\mathbb{P}\big(\lim_{n\to\infty} X_n = X\big) = 1$	The whole sequence converges pointwise, except on a negligible set.
In probability	$\forall \varepsilon > 0,\ \mathbb{P}(\lvert X_n - X\rvert > \varepsilon) \to 0$	The probability of being far from X goes to zero.
In L^2	$\mathbb{E}\big[(X_n - X)^2\big] \to 0$	The mean squared error goes to zero (average distance shrinks).
In distribution	$F_{X_n}(x) \to F_X(x)$ (for all continuity points of F_X)	The laws converge: shapes of distributions become similar.

Hierarchy of Convergences

$$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$$

- Almost sure convergence is the strongest notion.
- Convergence in distribution is the weakest.
- None of the reverse implications hold in general.

Additional fact

If $X_n \xrightarrow{L^r} X$, then $X_n \xrightarrow{P} X$.

3. Weak Law of Large Numbers

Intuition

Let $X_1, \ldots, X_n$ be independent random variables with common mean m, and let

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- Each X_i is random, so a single observation can be 'far' from m.
- But $\overline{X}_n$ averages many observations, so positive and negative fluctuations tend to compensate.
- As n grows, the empirical mean becomes more and more stable.
- The **weak law of large numbers** states that $\overline{X}_n$ **gets arbitrarily close to** m **in probability**.

$$\overline{X}_n \xrightarrow{P} m$$

- Said differently, for every $\varepsilon > 0$,

$$\mathbb{P}\big(|\overline{X}_n - m| > \varepsilon\big) \longrightarrow 0.$$

Weak Law of Large Numbers

Theorem 3. Weak Law of Large Numbers

Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables such that

$$\mathbb{E}(X_n) = m, \qquad \mathrm{Var}(X_n) = \sigma^2, \qquad \forall n \geq 1.$$

Define

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then

$$\overline{X}_n \xrightarrow{P} m.$$

Equivalently, for every $\varepsilon > 0$, $\mathbb{P}\big(|\overline{X}_n - m| > \varepsilon\big) \longrightarrow 0.$

The empirical average converges in probability to the common expectation.

Proof

- We compute the expectation and the variance of $\overline{X}_n$. Since expectation is linear,

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n} m = m.$$

- Because the X_i's are independent,

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

- Then, we apply the Bienaymé–Tchebychev inequality: for every $\varepsilon > 0$,

$$\mathbb{P}(|\overline{X}_n - m| > \varepsilon) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

- Since $\frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$ as $n \to \infty$, we obtain $\mathbb{P}(|\overline{X}_n - m| > \varepsilon) \longrightarrow 0$.
- Therefore, $\overline{X}_n \xrightarrow{P} m$.

Intuition (again)

For one observation X_i, randomness can be large.

For the empirical mean,

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

the random fluctuations are averaged out:

- the mean stays equal to m;
- the variance becomes

$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n},$$

which tends to 0 as n tends to ∞.

So when n is large, $\overline{X}_n$ is very unlikely to be far from m ($\overline{X}_n \approx m$ for large n).

Remark

Remark

For the weak law of large numbers does not require the random variables to have the same distribution.

The weak law of large numbers remains valid for independent random variables with possibly different distributions, as long as their variability does not grow too fast.

Example: Empirical Frequency of Success

Empirical Frequency of Success

- *Assume we perform n independent experiments, each with probability $p \in [0,1]$ of success.*
- *For each experiment i, define the random variable X_i that takes the value 1 if experiment i is a success, 0 otherwise.*
- *Then $X_i \sim \mathcal{B}(p)$, and the variables $(X_i)_{1 \leq i \leq n}$ are independent.*
- *We consider the empirical mean (i.e., frequency of success):*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- *$\overline{X}_n$ is the proportion of successes among the n experiments. It is a random quantity, since the outcomes are random.*

Interpretation

- We have:
$$\mathbb{E}(X_i) = p, \qquad \mathrm{Var}(X_i) = p(1-p).$$
- By the weak law of large numbers, we have
$$\overline{X}_n \xrightarrow{P} p,$$
That is, the empirical frequency of success converges to the true probability p.
- We can verify the result easily:
$$\mathbb{E}(\overline{X}_n) = p, \quad \mathrm{Var}(\overline{X}_n) = \frac{1}{n}p(1-p).$$
- Apply the Bienaymé–Tchebychev inequality:
$$\mathbb{P}\left(\mid \overline{X}_n - p \mid > \varepsilon\right) \leq \frac{p(1-p)}{n} \xrightarrow[n\to\infty]{} 0.$$

Exercise

Exercise

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with the same law as X, where X is absolutely continuous with density

$$f(x) = \frac{\theta}{2} e^{-\theta|x|}, \qquad \theta > 0,\ x \in \mathbb{R}.$$

Study the convergence in probability of the sequence

$$S_n = \frac{1}{n} \sum_{i=1}^{n} |X_i|.$$

Solution: Expectation (1/5)

- Let
$$Y_i = |X_i|, \qquad i \geq 1.$$
- Then $(Y_i)_{i\geq 1}$ is still an i.i.d. sequence, and $S_n = \frac{1}{n}\sum_{i=1}^n Y_i$.
- We compute the expectation of $Y_1 = |X|$:
$$\mathbb{E}(|X|) = \int_{\mathbb{R}} |x|\, f(x)\, dx = \int_{\mathbb{R}} |x|\, \frac{\theta}{2} e^{-\theta|x|}\, dx.$$
- Since the density is symmetric, indeed
$$f(-x) = \frac{\theta}{2} e^{-\theta|-x|} = \frac{\theta}{2} e^{-\theta|x|} = f(x),$$
we obtain
$$\mathbb{E}(|X|) = \int_{\mathbb{R}} |x| f(x)\, dx = 2\int_0^{+\infty} x f(x)\, dx = \int_0^{+\infty} x\theta e^{-\theta x}\, dx.$$

Solution: Expectation (2/5)

- Integrating by parts with

$$u(x) = x, \quad u'(x) = 1, \quad v'(x) = \theta e^{-\theta x}, \quad v(x) = -e^{-\theta x},$$

we have

$$\int_0^{+\infty} x\theta e^{-\theta x}\, dx = \Big[-xe^{-\theta x}\Big]_0^{+\infty} + \int_0^{+\infty} e^{-\theta x}\, dx = \left[-\frac{1}{\theta} e^{-\theta x}\right]_0^{+\infty} = \frac{1}{\theta},$$

- hence

$$\mathbb{E}(|X|) = \frac{1}{\theta}.$$

Solution: Variance (3/5)

- Now, to check that the second moment exists, we compute
  $\mathbb{E}(|X|^2) = \mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f(x)\,dx = \int_{\mathbb{R}} x^2 \frac{\theta}{2} e^{-\theta|x|}\,dx.$
- Since the density is symmetric,
  $$\mathbb{E}(X^2) = 2\int_0^{+\infty} x^2 f(x)\,dx = \theta \int_0^{+\infty} x^2 e^{-\theta x}\,dx.$$
- Integrating by parts with
  $$u(x) = x^2, \quad u'(x) = 2x, \quad v'(x) = \theta e^{-\theta x}, \quad v(x) = -e^{-\theta x},$$
  we get
  $$\theta \int_0^{+\infty} x^2 e^{-\theta x}\,dx = \int_0^{+\infty} x^2 \theta e^{-\theta x}\,dx = \left[-x^2 e^{-\theta x}\right]_0^{+\infty} + 2\int_0^{+\infty} x e^{-\theta x}\,dx.$$

Solution: Variance (4/5)

- We previously obtained

$$\int_0^{+\infty} x\theta e^{-\theta x}\, dx = \frac{1}{\theta}$$

  Hence,

$$\mathbb{E}(X^2) = 2\int_0^{+\infty} xe^{-\theta x}\, dx = \frac{2}{\theta^2}.$$

- The variance is then

$$\mathrm{Var}(\mid X \mid) = \mathbb{E}(\mid X \mid^2) - \mathbb{E}(\mid X \mid)^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{2}{\theta^2}.$$

- Thus, the variance is finite.

Solution: Conclusion (5/5)

- Since the random variables $Y_i = |X_i|$ are independent and identically distributed with finite expectation
$$\mathbb{E}(Y_i) = \mathbb{E}(|X_i|) = \frac{1}{\theta},$$
and finite variance
$$\mathrm{Var}(Y_i) = \mathrm{Var}(|X_i|) = \frac{1}{\theta^2},$$
the weak law of large numbers applies to the empirical mean
$$S_n = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$
- Therefore,
$$S_n \xrightarrow{\mathbb{P}} \frac{1}{\theta}.$$
- The empirical average of the absolute values converges in probability to $\frac{1}{\theta}$.

4. Central Limit Theorem

Central Limit Theorem: Some Intuition

Let $(X_i)_{i\geq 1}$ be i.i.d. random variables with mean m and variance σ^2.

- The **weak law of large numbers** tells us that:
  - for a sequence of independent random variables $(X_n)_{n\geq 1}$
  - with mean $\mathbb{E}(X_n) = m$ and $\mathrm{Var}(X_n) = \sigma^2$,
  - the sample mean $\overline{X}$ converges in probability to m: $\overline{X}_n \xrightarrow{P} m$ (i.e., for large n, $\overline{X}_n \approx m$).
- **But how does $\overline{X}_n$ fluctuate around m?**
- The **Central Limit Theorem** answers this question: it describes the **distribution of the fluctuations**.

Normalization

- Consider the **sum** and the **sample mean**:

$$S_n = \sum_{i=1}^{n} X_i, \qquad \overline{X}_n = \frac{S_n}{n}.$$

- We have:

$$\mathbb{E}(\overline{X}_n) = m, \qquad \mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

- To obtain a non-degenerate limit, we rescale:

$$\sqrt{n}(\overline{X}_n - m).$$

- Then, with this transformation, we have::

$$\mathbb{E}\big[\sqrt{n}(\overline{X}_n - m)\big] = 0, \qquad \mathrm{Var}\big[\sqrt{n}(\overline{X}_n - m)\big] = \sigma^2.$$

Central Limit Theorem

Theorem 4. Central Limit Theorem

Let $(X_i)_{i\geq 1}$ *be i.i.d. random variables such that*

$$\mathbb{E}(X_i) = m, \qquad \mathrm{Var}(X_i) = \sigma^2 < +\infty.$$

Define

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then

$$\sqrt{n}(\overline{X}_n - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Remark

Equivalently,

$$\frac{\sqrt{n}(\overline{X}_n - m)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

Interpretation of the CLT

- For large n,
$$\sqrt{n}(\overline{X}_n - m) \approx \mathcal{N}(0, \sigma^2).$$
- Equivalently,
$$\overline{X}_n \approx \mathcal{N}\left(m, \frac{\sigma^2}{n}\right).$$
- Hence, the CLT describes the asymptotic distribution of the error $\overline{X}_n - m$.
  - the average is approximately normal,
  - the variance decreases as $1/n$,
  - fluctuations are of order $1/\sqrt{n}$.
- $\overline{X}_n = m + \frac{\sigma}{\sqrt{n}} \times$ (Gaussian noise).

Law of Large Numbers vs Central Limit Theorem

Weak Law of Large Numbers	**Central Limit Theorem**
$\overline{X}_n \xrightarrow{P} m$	$\sqrt{n}(\overline{X}_n - m) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

- Weak LLN: convergence to the mean,
- CLT: distribution of the error $\overline{X}_n - m$.
  - After standardization:
  $$\frac{\overline{X}_n - m}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$
  - Therefore, for large n,
  $$\overline{X}_n \approx \mathcal{N}\left(m, \frac{\sigma^2}{n}\right).$$
  - In other words,
  $$\overline{X}_n = m + \frac{\sigma}{\sqrt{n}} \times \text{(Gaussian noise)}.$$

Example: Bernoulli Case

Normal Approximation of Proportions

Let $X_i \sim \mathcal{B}(p)$, i.i.d.

- Then
$$\mathbb{E}(X_i) = p, \qquad \mathrm{Var}(X_i) = p(1-p).$$
- By the Central Limit Theorem,
$$\sqrt{n}(\overline{X}_n - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)).$$
- Hence, for large n,
$$\overline{X}_n \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right).$$
- In particular, the empirical proportion is approximately normal.

Poisson to Normal Convergence

Proposition 9. Poisson to Normal Convergence

Let $(X_n)_n$ be a sequence of random variables such that

$$X_n \sim \mathcal{P}(n).$$

Then

$$\frac{X_n - n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

- In other words, for large n,
$$X_n \approx \mathcal{N}(n,n).$$
- A Poisson distribution with large parameter behaves approximately like a normal distribution.

Poisson to Normal: Sketch of Proof (1/2)

Let $(X_n)_n$ be a sequence of random variables such that

$$X_n \sim \mathcal{P}(n).$$

We want to show that

$$\frac{X_n - n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

- Recall the additive property of the Poisson distribution: if $X_i \sim \mathcal{P}(\lambda_i)$ are independent, then

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right).$$

- Using this property, X_n has the same law as

$$\sum_{i=1}^{n} Y_i,$$

where $Y_1, \ldots, Y_n$ are independent random variables with

$$Y_i \sim \mathcal{P}(1).$$

Poisson to Normal: Sketch of Proof (2/2)

- Therefore, ($\overset{\mathcal{L}}{=}$ means **equality in distribution**: the two random variables have the same law.)
$$\frac{X_n - n}{\sqrt{n}} \overset{\mathcal{L}}{=} \frac{\sum_{i=1}^n Y_i - n}{\sqrt{n}}.$$
- Since $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$, we have
$$\frac{\sum_{i=1}^n Y_i - n}{\sqrt{n}} = \sqrt{n}\,(\overline{Y}_n - 1).$$
- In addition,
$$\mathbb{E}(Y_i) = 1, \qquad \mathrm{Var}(Y_i) = 1,$$
so the Central Limit Theorem implies
$$\sqrt{n}\,(\overline{Y}_n - 1) \xrightarrow{d} \mathcal{N}(0,1).$$
- Hence,
$$\frac{X_n - n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Binomial to Poisson Approximation

Proposition 10.

Let $(X_n)_n$ be a sequence of random variables such that

$$X_n \sim \mathcal{B}(n, p_n), \qquad p_n \to 0, \qquad np_n \to \lambda > 0.$$

Then

$$X_n \xrightarrow{d} \mathcal{P}(\lambda).$$

- In other words, for large n and small p_n,

$$X_n \approx \mathcal{P}(\lambda).$$

Intuition

- We consider n independent trials with success probability p_n.
- **Each event is rare** ($p_n \to 0$), but there are **many trials** ($n \to \infty$).
- The expected number of successes is

$$\mathbb{E}(X_n) = np_n \to \lambda.$$

- So, we are counting the number of rare independent events.
- In this regime, only the total number of successes matters, not which trials succeed.

Binomial to Poisson Approximation: Proof Sketch (1/2)

- For $k \in \mathbb{N}$,

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}.$$

- Rewrite (using $p_n^k = \frac{(np_n)^k}{n^k}$):

$$\mathbb{P}(X_n = k) = \frac{n(n-1)\cdots(n-k+1)}{k!}\, p_n^k\, (1-p_n)^{n-k}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{(np_n)^k}{k!} \cdot (1-p_n)^n \cdot (1-p_n)^{-k}.$$

- As $n \to \infty$,

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \to 1, \qquad (np_n)^k \to \lambda^k,$$

and

$$(1-p_n)^n \to e^{-\lambda}, \qquad (1-p_n)^{-k} \to 1.$$

Binomial to Poisson Approximation: Proof Sketch (2/2)

- Since $np_n \to \lambda$, we have
$$p_n^k n^k \to \lambda^k.$$
- Hence
$$\lim_{n \to +\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$
- We recognize the law $\mathcal{P}(\lambda)$.

References I

Garnier, J., Méléard, S. and Touzi, N. (2021). *Aléatoire*. Département de Mathématiques Appliquées, Ecole Polytechnique.

Hansen, B. (2022). *Probability and statistics for economists*. Princeton University Press.

Miller, S. J. (2017). *The probability lifesaver: all the tools you need to understand chance*. Princeton University Press.