# Advanced Probability and Applications: WEEK 5

## 1 Laws of large numbers

We state below the law of large numbers. This law justifies the notion of theoretical expectation, as it shows that the average of a large number of independent and identically distributed (i.i.d.) random variables converges to this expectation (in probability and almost surely).

**Theorem 1.1.** Let \((X_n, n \geq 1)\) be a sequence of i.i.d. random variables such that \(E(X_1^2) < +\infty\), and let \(S_n = X_1 + \ldots + X_n\). Then:

a) \(\frac{S_n}{n} \xrightarrow{p} E(X_1)\) (weak law).

b) \(\frac{S_n}{n} \to E(X_1)\) almost surely (strong law).

**Remark 1.2.** Both laws above hold under the weaker assumption that \(E(|X_1|) < +\infty\), but in this case, the proof of the theorem becomes significantly longer!

**Remark 1.3.** One may wonder why should one state both a weak and a strong law, as the latter is obviously a stronger result than the former. There is a good reason for this: both the weak and the strong law can be generalized to different sets of assumptions on the random variables \(X_i\)'s. But more generalizations are possible for the weak law than for the strong law. We consider here a slightly restrictive set of assumptions that allows for a relatively short and self-contained proof of the theorem.

**Proof.** a) For all \(\varepsilon > 0\), we have

\[
P \left( \left| \frac{S_n}{n} - E(X_1) \right| > \varepsilon \right) = P \left( \left| S_n - nE(X_1) \right| > n\varepsilon \right) = P \left( \left| S_n - nE(S_n) \right| > n\varepsilon \right) \leq \frac{E((S_n - nE(S_n))^2)}{n^2\varepsilon^2} = \frac{\text{Var}(S_n)}{n^2\varepsilon^2} = \frac{\text{Var}(X_1)}{n\varepsilon^2} \to 0,\]

where we have used Chebychev’s inequality and the fact that the variance of a sum of independent variables is the sum of the variances of these random variables. This implies that \(\frac{S_n}{n} \xrightarrow{p} E(X_1)\) and therefore proves the weak law of large numbers.

b) Note that the former proof does not allow to conclude here, because we only showed that

\[
P \left( \left| \frac{S_n}{n} - E(X_1) \right| > \varepsilon \right) = O \left( \frac{1}{n} \right),
\]

so we cannot apply Borel-Cantelli’s lemma in this case, as mentioned already in the last lecture. There is nevertheless an elegant solution to this problem, as described in the sequel.

- Observe first that we may simply replace \(n\) by \(n^2\) in the previous equality, so as to obtain:

\[
P \left( \left| \frac{S_n^2}{n^2} - E(X_1) \right| > \varepsilon \right) = O \left( \frac{1}{n^2} \right).
\]

Using the Borel-Cantelli lemma and the fact that \(\sum_{n \geq 1} \frac{1}{n^2} < \infty\), we obtain that \(\frac{S_n^2}{n^2} \to E(X_1)\) almost surely. This alone of course does not prove the result, but...

- Assume for now that \(X_n \geq 0\) for all \(n \geq 1\) and consider an integer \(m\) such that \(n^2 \leq m \leq (n + 1)^2\). Because the \(X_i\)'s are positive, the sequence \(S_n\) is increasing, so we obtain in this case

\[
\frac{S_n^2}{(n + 1)^2} \leq \frac{S_m}{m} \leq \frac{S_{(n+1)^2}}{n^2}.
\]
Note that by what was just shown above and by the fact that \(\frac{(n+1)^2}{n^2} \to 1\), both the left-most and the right-most terms converge almost surely to \(E(X_1)\) as \(n \to \infty\). As \(\frac{S_m}{m}\) is lower and upper bounded by these two terms, respectively, we deduce that \(\frac{S_m}{m}\) also converges almost surely to \(E(X_1)\) as \(m \to \infty\).

- Finally, we need to address the case where the \(X\)'s are not necessarily positive. Let us define in this case \(X_n^+ = \max(X_n, 0)\) and \(X_n^- = \max(-X_n, 0)\), so that \(X_n = X_n^+ - X_n^-\). Similarly, let

\[
S_n^+ = \sum_{j=1}^{n} X_j^+ \quad \text{and} \quad S_n^- = \sum_{j=1}^{n} X_j^-,
\]

so that \(S_n = S_n^+ - S_n^-\).

Note that it is not necessarily the case that \(S_n^+ = \max(S_n, 0)\) and \(S_n^- = \max(-S_n, 0)\), but this does not matter here.

Applying the above argument, we see that

\[
\frac{S_n^+}{n} \to_{n \to \infty} E(X_1^+) \quad \text{a.s. and} \quad \frac{S_n^-}{n} \to_{n \to \infty} E(X_1^-) \quad \text{a.s.}
\]

which in turn implies that

\[
\frac{S_n}{n} = \frac{S_n^+}{n} - \frac{S_n^-}{n} \to_{n \to \infty} E(X_1^+) - E(X_1^-) = E(X_1) \quad \text{a.s.}
\]

and therefore completes the proof.

### 1.1 Application: convergence of the empirical distribution

Let again \((X_n, n \geq 1)\) be a sequence of i.i.d. random variables (but without any assumption on their integrability), and let \(F\) denote their common cdf \((F(t) = P(\{X_1 \leq t\}), t \in \mathbb{R})\).

Let now \(F_n(t) = \frac{1}{n} \sum_{1 \leq j \leq n} \mathbb{1}_{\{X_j \leq t\}}\) for \(t \in \mathbb{R}; F_n\) is the empirical distribution (or cdf) of the first \(n\) random variables \(X_1, \ldots, X_n\). Note that for fixed \(n\), it is a discrete distribution (i.e., the cdf is a staircase function). As well as the law of large numbers provides a justification for the notion of expectation, the statement below provides a justification for the notion of distribution.

**Theorem 1.4.** For every \(t \in \mathbb{R},\)

\[
F_n(t) \to_{n \to \infty} F(t) \quad \text{almost surely.}
\]

**Proof.** Fix \(t \in \mathbb{R}\) and let \(Y_j = \mathbb{1}_{\{X_j \leq t\}}\). As the \(X\)'s are i.i.d., so are the \(Y\)'s. On top of that, the \(Y\)'s are square-integrable, as they are Bernoulli random variables, taking values in \(\{0, 1\}\) only. Also, \(F_n(t)\) may be rewritten as

\[
F_n(t) = \frac{1}{n} \sum_{j=1}^{n} Y_j \to_{n \to \infty} E(Y_1) \quad \text{almost surely,}
\]

by the strong law of large numbers. Noting finally that \(E(Y_1) = P(\{X_1 \leq t\}) = F(t)\) completes the proof.

### 1.2 Kolmogorov’s 0-1 law

The strong law cannot be extended beyond the \(E(|X_1|) < +\infty\) assumption. One can show actually the following more precise statement:

- If \(E(|X_1|) < +\infty\), then \(\lim_{n \to \infty} \frac{S_n}{n} = E(X_1)\) a.s.
• If \( \mathbb{E}(|X_1|) = +\infty \), then \( \limsup_{n \to \infty} \left| \frac{S_n}{n} \right| = +\infty \) a.s., i.e., \( \frac{S_n}{n} \) diverges a.s.

The aim here is not to give a full proof of the above statements, but rather to explain why \( \frac{S_n}{n} \) can only converge or diverge a.s.

Let \( (X_n, n \geq 1) \) be a sequence of random variables, all defined on the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\).

For \( n \geq 1 \), define
\[
G_n = \sigma(X_n, X_{n+1}, X_{n+2}, \ldots) \quad \text{and} \quad T = \cap_{n \geq 1} G_n.
\]

\( T \) is called the tail \( \sigma \)-field of the sequence \( (X_n, n \geq 1) \). In words, it is the information related to the asymptotic behavior of the sequence when \( n \to \infty \).

Here is an example of event in \( T \):
\[
A_1 = \left\{ \omega \in \Omega : \sum_{n \geq 1} X_n(\omega) \text{ converges} \right\}
\]

Indeed, notice that for every \( N \geq 1 \), we have
\[
A_1 = \left\{ \omega \in \Omega : \sum_{n \geq N} X_n(\omega) \text{ converges} \right\},
\]

so \( A_1 \in G_N \) for every \( N \geq 1 \). It therefore also belongs to \( T = \cap_{N \geq 1} G_N \).

Along the same lines, here is another example of event in \( T \):
\[
A_2 = \left\{ \omega \in \Omega : \lim_{n \to \infty} \frac{S_n(\omega)}{n} \text{ exists} \right\},
\]

which is of direct interest to us in the sequel.

**Theorem 1.5.** (Kolmogorov’s 0-1 law)

If the sequence \( (X_n, n \geq 1) \) is independent and \( A \in T \), then \( \mathbb{P}(A) \in \{0, 1\} \).

**Remark.** Please note that \( \mathbb{P}(A) \in \{0, 1\} \) means that either \( \mathbb{P}(A) = 0 \) or \( \mathbb{P}(A) = 1 \), not that \( 0 \leq \mathbb{P}(A) \leq 1 \), which is always true.

**Consequence.** Because the event \( A_2 \) above belongs to \( T \), we can therefore conclude that either the sequence \( \frac{S_n}{n} \) converges a.s., or it diverges a.s., but it cannot be that convergence takes place with a probability which is strictly between 0 and 1. It finally turns out that a.s. convergence takes place if and only if \( \mathbb{E}(|X_1|) < +\infty \), and correspondingly that a.s. divergence takes place if and only if \( \mathbb{E}(|X_1|) = +\infty \).

**Proof.** The strategy for the proof of the above theorem is to show that when the \( X_n \)’s are independent, any event \( A \in T \) is independent of itself! So that \( \mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2 \), implying \( \mathbb{P}(A) \in \{0, 1\} \).

First notice that because of the independence of the \( X \)’s, for every \( n \geq 1 \), the \( \sigma \)-fields
\[
\mathcal{F}_n = \sigma(X_1, \ldots, X_n) \quad \text{and} \quad \mathcal{G}_{n+1} = \sigma(X_{n+1}, X_{n+2}, \ldots)
\]

are independent. As \( T \subset \mathcal{G}_{n+1} \) for every \( n \geq 1 \), this also implies that \( T \) is independent of \( \mathcal{F}_n \) for every \( n \geq 1 \). But this implies that \( T \) is also independent of \( \sigma(X_1, X_2, \ldots, X_n, \ldots) \), which is the \( \sigma \)-field generated by all the \( \mathcal{F}_n \)’s. Finally, observe that \( T \subset \sigma(X_1, X_2, \ldots, X_n, \ldots) \), which implies that \( T \) is independent of itself! So that any event \( A \in T \) has probability 0 or 1, as mentioned above. \( \square \)
2 Convergence in distribution

Convergence in distribution is a key tool in probability, allowing notably to state the central limit theorem.

**Definition 2.1.** Let \((X_n, n \geq 1)\) be a sequence of random variables, not necessarily defined on the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\). The sequence \((X_n, n \geq 1)\) is said to converge in distribution to a limiting random variable \(X\) (and this is denoted as \(X_n \xrightarrow{d} X\)) if

\[
F_{X_n}(t) = \mathbb{P}(\{X_n \leq t\}) \xrightarrow{n \to \infty} F_X(t) = \mathbb{P}(\{X \leq t\}),
\]

for every \(t \in \mathbb{R}\) continuity point of the limiting cdf \(F_X\).

**Remark 2.2.** Why asking only for convergence in continuity points of \(F_X\) and not in all \(t \in \mathbb{R}\)? There are two main reasons for this:

- The fact is, it may happen that the limit cdf \(F_X\) is itself discontinuous (if it is e.g. the cdf of a discrete random variable, in which case we recall that \(F_X\) is a staircase function). In this case, it would be asking for too much to have a sequence of functions converging to \(F_X\) in every \(t \in \mathbb{R}\), including in points where the function \(F_X\) makes a jump; one can at least imagine easily examples of sequences of functions that converge everywhere except in these points.

- Besides, as we know that the limit \(F_X\) is a right-continuous function, by definition, this implies that even if there is no convergence in a point where \(F_X\) makes a jump, it is always possible to reconstruct \(F_X\) in this point by taking the limit from the right.

The proposition below shows that convergence in distribution is the weakest of the four notions of convergence we have seen so far.

**Proposition 2.3.** Let \((X_n, n \geq 1)\) be a sequence of random variables defined on the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and \(X\) be another random variable defined on \((\Omega, \mathcal{F}, \mathbb{P})\). If \(X_n \xrightarrow{d} X\) and \(X_n \xrightarrow{p} X\), then \(X_n \xrightarrow{d} X\).

**Proof.** Recall that \(X_n \xrightarrow{p} X\) means that for all \(\varepsilon > 0\), \(\lim_{n \to \infty} \mathbb{P}(\{|X_n - X| > \varepsilon\}) = 0\). We show below that this implies that \(\lim_{n \to \infty} F_{X_n}(t) = F_X(t)\) for every \(t \in \mathbb{R}\) continuity point of \(F_X\).

- Let us first compute, for a given fixed \(\varepsilon > 0\):

\[
F_{X_n}(t) = \mathbb{P}(\{X_n \leq t\}) = \mathbb{P}(\{X_n \leq t, X \leq t + \varepsilon\}) + \mathbb{P}(\{X_n \leq t, X > t + \varepsilon\})
\]

\[
\leq \mathbb{P}(\{X \leq t + \varepsilon\}) + \mathbb{P}(\{|X_n - X| > \varepsilon\}).
\]

Because of the assumption made, this implies that \(\lim_{n \to \infty} F_{X_n}(t) \leq F_X(t + \varepsilon)\) + 0.

- Let us then compute, again for a given fixed \(\varepsilon > 0\):

\[
F_X(t - \varepsilon) = \mathbb{P}(\{X \leq t - \varepsilon\}) = \mathbb{P}(\{X \leq t - \varepsilon, X_n \leq t\}) + \mathbb{P}(\{X \leq t - \varepsilon, X_n > t\})
\]

\[
\leq \mathbb{P}(\{X_n \leq t\}) + \mathbb{P}(\{|X_n - X| > \varepsilon\}).
\]

Again, because of the assumption made, this implies that \(F_X(t - \varepsilon) \leq \lim_{n \to \infty} F_{X_n}(t) + 0\).

- In conclusion, we obtain for any given \(\varepsilon > 0\):

\[
F_X(t - \varepsilon) \leq \lim_{n \to \infty} F_{X_n}(t) \leq \lim_{n \to \infty} \sup F_{X_n}(t) \leq F_X(t + \varepsilon).
\]

Assuming \(t \in \mathbb{R}\) is a continuity point of \(F_X\), we have \(\lim_{\varepsilon \downarrow 0} F_X(t - \varepsilon) = \lim_{\varepsilon \downarrow 0} F_X(t + \varepsilon) = F_X(t)\), so by the above inequalities,

\[
\lim_{n \to \infty} F_{X_n}(t) = F_X(t),
\]

which proves the claim. □
2.1 Equivalent criterion for convergence in distribution

The following theorem, also known as (part of) the “Portmanteau theorem”, gives an equivalent criterion for convergence in distribution.

**Theorem 2.4.** Let $(X_n, n \geq 1)$ be a sequence of random variables and $X$ be another random variable. Then $X_n \xrightarrow{d} X$ if and only if

$$E(g(X_n)) \xrightarrow{n \to \infty} E(g(X)),$$

for every continuous and bounded function $g : \mathbb{R} \to \mathbb{R}$.

*Proof sketch of the “if” part.* Observe first that Definition 2.1 is equivalent to saying that

$$E(h(X_n)) \xrightarrow{n \to \infty} E(h(X)),$$

for every function $h : \mathbb{R} \to \mathbb{R}$ of the form $h(x) = 1_{\{x \leq t\}}$ where $t$ is a continuity point of $F$. Our aim is to show that for fixed $t \in \mathbb{R}$, there is a way to approximate from above and from below the step function $h$ with continuous and bounded functions. To this end, let us define for $m \geq 1$:

$$g_m(x) = \begin{cases} 
1 & \text{if } x \leq t \\
 m \left(1 - \frac{1}{m}\right) & \text{if } t \left(1 - \frac{1}{m}\right) < x \leq t \\
0 & \text{if } x > t 
\end{cases}$$

and

$$G_m(x) = \begin{cases} 
1 & \text{if } x \leq t \\
 m \left(1 + \frac{1}{m} - \frac{1}{m}\right) & \text{if } t < x \leq t \left(1 + \frac{1}{m}\right) \\
0 & \text{if } x > t \left(1 + \frac{1}{m}\right)
\end{cases}$$

As one easily sees from the figure above, the functions $g_m$ and $G_m$ are continuous and bounded and $g_m(x) \leq h_t(x) \leq G_m(x)$ for every $x \in \mathbb{R}$ and every $m \geq 1$. Also, $\lim_{m \to \infty} g_m(x) = \lim_{m \to \infty} G_m(x) = h_t(x)$ for every $x \neq t$. From the assumption made, we obtain that for every $m \geq 1$:

$$E(g_m(X)) = \lim_{n \to \infty} E(g_m(X_n)) \leq \lim_{n \to \infty} E(h_t(X_n)) \leq \lim_{n \to \infty} E(G_m(X_n)) = E(G_m(X))$$

A last technical but important detail (not shown here) allows to conclude that for $t$ a continuity point of $F_X$, both $E(g_m(X)) \xrightarrow{m \to \infty} E(h_t(X))$ and $E(G_m(X)) \xrightarrow{m \to \infty} E(h_t(X))$, implying the result. □

**Remark 2.5.** For $k \in \mathbb{N}$, let $C^k_b(\mathbb{R})$ denote the space of $k$ times continuously differentiable functions $g : \mathbb{R} \to \mathbb{R}$, which are bounded and whose all $k$ derivatives are also bounded. Replacing the above functions $g_m$ and $G_m$ by regular cubic splines, one can show the following improvement of the above theorem:

If $E(g(X_n)) \xrightarrow{n \to \infty} E(g(X))$ for every function $g \in C^3_b(\mathbb{R})$, then $X_n \xrightarrow{d} X$.

This remark will be useful for next week.
Appendix: extensions of the weak law

The weak law can be extended in multiple directions, but in particular:

- to dependent random variables: see exercises.
- to non-integrable random variables: see below.

Let us consider a sequence \((X_n, n \geq 1)\) of i.i.d. random variables such that

\[
M \cdot \mathbb{P}(\{|X_1| \geq M\}) \to M \to \infty 0. \tag{1}
\]

Then it holds that

\[
\frac{S_n}{n} - \mathbb{E} \left( X_1 \mathbf{1}_{\{|X_1| \geq n\}} \right) \xrightarrow{p} n \to \infty 0.
\]

**Remarks.** - Condition (1) is weaker than asking \(\mathbb{E}(|X_1|) < +\infty\). Indeed, one can show that

\[
\mathbb{E}(|X_1|) \simeq \sum_{M \geq 1} \mathbb{P}(\{|X_1| \geq M\})
\]

is not necessarily finite under the above condition. Take e.g. \(\mathbb{P}(\{|X_1| \geq M\}) = O \left( \frac{1}{M \log M} \right)\). This satifies (1), but \(\sum_{M \geq 1} \mathbb{P}(\{|X_1| \geq M\}) = +\infty\).

- If \(\mathbb{E}(|X_1|) < \infty\), then the sequence of numbers \(a_n = \mathbb{E} (X_1 \mathbf{1}_{\{|X_1| \geq n\}})\) converges towards \(\mathbb{E}(X_1)\).

- One can show that condition (1) is optimal, that is: if it is not satisfied, then there does not exist a sequence of numbers \(a_n\) such that

\[
\frac{S_n}{n} - a_n \xrightarrow{p} n \to \infty 0.
\]