# 2 Convergence of random variables

In probability theory one uses various modes of convergence of random variables, many of which are crucial for applications. In this section we shall consider some of the most important of them: convergence in $L^r$, convergence in probability and convergence with probability one (a.k.a. almost sure convergence).

## 2.1 Weak laws of large numbers

✍ **Definition 2.1.** Let $r > 0$ be fixed. We say that a sequence $X_j$, $j \geq 1$, of random variables *converges* to a random variable $X$ *in* $L^r$ (write $X_n \xrightarrow{\mathsf{L^r}} X$) as $n \to \infty$, if $\mathsf{E}\big|X_n - X\big|^r \to 0$ as $n \to \infty$.

**Example 2.2.** *Let $\big(X_n\big)_{n\geq 1}$ be a sequence of random variables such that for some real numbers $(a_n)_{n\geq 1}$, we have*

$$\mathsf{P}\big(X_n = a_n\big) = p_n\,, \qquad \mathsf{P}\big(X_n = 0\big) = 1 - p_n\,. \tag{2.1}$$

*Then $X_n \xrightarrow{\mathsf{L^r}} 0$ iff $\mathsf{E}\big|X_n\big|^r \equiv |a_n|^r\, p_n \to 0$ as $n \to \infty$.*

The following result is the $L^2$ weak law of large numbers ($L^2$-WLLN)

☞ **Theorem 2.3.** *Let $X_j$, $j \geq 1$, be a sequence of uncorrelated random variables with $\mathsf{E}X_j = \mu$ and $\mathsf{Var}(X_j) \leq C < \infty$. Denote $S_n = X_1 + \cdots + X_n$. Then $\frac{1}{n}S_n \xrightarrow{\mathsf{L^2}} \mu$ as $n \to \infty$.*

Proof. Immediate from

$$\mathsf{E}\Big(\frac{1}{n}S_n - \mu\Big)^2 = \mathsf{E}\frac{(S_n - n\mu)^2}{n^2} = \frac{\mathsf{Var}(S_n)}{n^2} \leq \frac{Cn}{n^2} \to 0 \qquad \text{as } n \to \infty\,. \qquad \square$$

✍ **Definition 2.4.** We say that a sequence $X_j$, $j \geq 1$, of random variables *converges* to a random variable $X$ *in probability* (write $X_n \xrightarrow{\mathsf{P}} X$) as $n \to \infty$, if for *every fixed* $\varepsilon > 0$

$$\mathsf{P}\big(|X_n - X| \geq \varepsilon\big) \to 0 \qquad \text{as } n \to \infty\,.$$

**Example 2.5.** *Let the sequence $\big(X_n\big)_{n\geq 1}$ be as in (2.1). Then for every $\varepsilon > 0$ we have $\mathsf{P}\big(|X_n| \geq \varepsilon\big) \leq \mathsf{P}\big(X_n \neq 0\big) = p_n$, so that $X_n \xrightarrow{\mathsf{P}} 0$ if $p_n \to 0$ as $n \to \infty$.*

The usual (WLLN) is just a convergence in probability result:

☞ **Theorem 2.6.** *Under the conditions of Theorem 2.3, $\frac{1}{n}S_n \xrightarrow{\mathsf{P}} \mu$ as $n \to \infty$.*

**Exercise 2.7.** Derive Theorem 2.6 from the Chebyshev inequality.

We prove Theorem 2.6 using the following simple fact:

☞ **Lemma 2.8.** *Let $X_j$, $j \geq 1$, be a sequence of random variables. If $X_n \xrightarrow{\mathsf{L^r}} X$ for some fixed $r > 0$, then $X_n \xrightarrow{\mathsf{P}} X$ as $n \to \infty$.*

Proof. By the generalized Markov inequality with $g(x) = x^r$ and $Z_n = |X_n - X| \geq 0$, we get: for every fixed $\varepsilon > 0$

$$\mathsf{P}\big(Z_n \geq \varepsilon\big) \equiv \mathsf{P}\big(|X_n - X|^r \geq \varepsilon^r\big) \leq \frac{\mathsf{E}|X_n - X|^r}{\varepsilon^r} \to 0$$

as $n \to \infty$. □

Proof of Theorem 2.6. Follows immediately from Theorem 2.3 and Lemma 2.8. □

As the following example shows, a high dimensional cube is almost a sphere.

**Example 2.9.** *Let $X_j$, $j \geq 1$ be iid with $X_j \sim U(-1,1)$. Then the variables $Y_j = (X_j)^2$ satisfy $\mathsf{E}Y_j = \frac{1}{3}$, $\mathsf{Var}(Y_j) \leq \mathsf{E}[(Y_j^2)] = \mathsf{E}[(X_j)^4] \leq 1$. Fix $\varepsilon > 0$ and consider the set*

$$A_{n,\varepsilon} \stackrel{\mathsf{def}}{=} \Big\{z \in \mathbb{R}^n : (1-\varepsilon)\sqrt{n/3} < |z| < (1+\varepsilon)\sqrt{n/3}\Big\},$$

*where $|z|$ is the usual Euclidean length in $\mathbb{R}^n$, $|z|^2 = \sum_{j=1}^n (z_j)^2$. By the* WLLN,

$$\frac{1}{n}\sum_{j=1}^n Y_j \equiv \frac{1}{n}\sum_{j=1}^n (X_j)^2 \xrightarrow{\mathsf{P}} \frac{1}{3};$$

*in other words, for every fixed $\varepsilon > 0$, a point $\mathbf{X} = (X_1, \dots, X_n)$ chosen uniformly at random in $(-1,1)^n$ satisfies*

$$\mathsf{P}\Big(\Big|\frac{1}{n}\sum_{j=1}^n (X_j)^2 - \frac{1}{3}\Big| \geq \varepsilon\Big) \equiv \mathsf{P}\big(\mathbf{X} \notin A_{n,\varepsilon}\big) \to 0 \qquad \text{as } n \to \infty,$$

*ie., for large $n$, with probability approaching one,* a random point $\mathbf{X} \in (-1,1)^n$ is near the $n$-dimensional sphere of radius $\sqrt{n/3}$ centred at the origin.

☞ **Theorem 2.10.** *Let random variables $S_n$, $n \geq 1$, have two finite moments, $\mu_n \equiv \mathsf{E}S_n$, $\sigma_n^2 \equiv \mathsf{Var}(S_n) < \infty$. If, for some sequence $b_n$, we have $\sigma_n/b_n \to 0$ as $n \to \infty$, then $(S_n - \mu_n)/b_n \to 0$ as $n \to \infty$, both in $L^2$ and in probability.*

Proof. The result follows immediately from the observation

$$\mathsf{E}\Big(\frac{(S_n - \mu_n)^2}{b_n^2}\Big) = \frac{\mathsf{Var}(S_n)}{b_n^2} \to 0 \qquad \text{as } n \to \infty. \quad □$$

**Example 2.11.** *In the "coupon collector's problem"*[12] *let $T_n$ be the time to collect all $n$ coupons. It is easy to show that $\mathsf{E}T_n = n\sum_{m=1}^n \frac{1}{m} \sim n\log n$ and $\mathsf{Var}(T_n) \leq n^2\sum_{m=1}^n \frac{1}{m^2} \leq \frac{\pi^2 n^2}{6}$, so that*

$$\frac{T_n - \mathsf{E}T_n}{n\log n} \to 0 \qquad \text{ie.,} \qquad \frac{T_n}{n\log n} \to 1$$

*as $n \to \infty$ both in $L^2$ and in probability.*

[12] Problem R4

## 2.2 Almost sure convergence

Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables having mean $\mathsf{E}X_1 = \mu$ and finite second moment. Denote $S_n \stackrel{\mathsf{def}}{=} \sum_{k=1}^n X_k$. Then the usual (weak) law of large numbers (WLLN) tells us that for every $\delta > 0$

$$\mathsf{P}\Big(\big|n^{-1}S_n - \mu\big| > \delta\Big) \to 0 \qquad \text{as } n \to \infty. \tag{2.2}$$

In other words, according to WLLN, $n^{-1}S_n$ converges in probability to a constant random variable $X \equiv \mu = \mathsf{E}(X_1)$, as $n \to \infty$ (recall Definition 2.4, Theorem 2.6).

It is important to remember that convergence in probability is not related to the pointwise convergence, ie., convergence $X_n(\omega) \to X(\omega)$ for a fixed $\omega \in \Omega$. The following useful definition can be realised in terms of a $\mathcal{U}[0,1]$ random variable, recall Remark 1.6.2

**Definition 2.12.** The *canonical probability space* is $\big(\Omega, \mathcal{F}, \mathsf{P}\big)$, where $\Omega = [0,1]$, $\mathcal{F}$ is the smallest $\sigma$-field containing all intervals in $[0,1]$, and $\mathsf{P}$ is the 'length measure' on $\Omega$ (ie., for $A = [a,b] \subseteq [0,1]$, $\mathsf{P}(A) = b - a$).

☞ **Example 2.13.** *Let $\big(\Omega, \mathcal{F}, \mathsf{P}\big)$ be the canonical probability space. For every event $A \in \mathcal{F}$ consider the* indicator random variable

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \textit{if } \omega \in A\,, \\ 0, & \textit{if } \omega \notin A\,. \end{cases} \tag{2.3}$$

*For $n \geq 1$ put $m = [\log_2 n]$, ie., $m \geq 0$ is such that $2^m \leq n < 2^{m+1}$, define*

$$A_n = \Big[\frac{n - 2^m}{2^m}, \frac{n + 1 - 2^m}{2^m}\Big] \subseteq [0,1]$$

*and let $X_n \stackrel{\mathsf{def}}{=} \mathbb{1}_{A_n}$. Since $\mathsf{P}\big(\big|\mathbb{1}_{A_n}\big| > 0\big) = \mathsf{P}(A_n) = 2^{-[\log_2 n]} < \frac{2}{n} \to 0$ as $n \to \infty$, the sequence $X_n$ converges in probability to $X \equiv 0$. However,*

$$\big\{\omega \in \Omega : X_n(\omega) \to X(\omega) \equiv 0 \textit{ as } n \to \infty\big\} = \varnothing\,,$$

*ie., there is no point $\omega \in \Omega$ for which the sequence $X_n(\omega) \in \{0,1\}$ converges to*
☞ *$X(\omega) = 0$.* [Try the R script simulating this sequence from the course webpage!]

The following is the key definition of this section.

✍ **Definition 2.14.** A sequence $(X_k)_{k\geq 1}$ of random variables in $(\Omega, \mathcal{F}, \mathsf{P})$ converges, as $n \to \infty$, to a random variable $X$ *with probability one* (or *almost surely*) if

$$\mathsf{P}\Big(\big\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\big\}\Big) = 1\,. \tag{2.4}$$

☞ **Remark 2.14.1.** For $\varepsilon > 0$, let $A_n(\varepsilon) = \big\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\big\}$. Then the property (2.4) is equivalent to saying that for every $\varepsilon > 0$

$$\mathsf{P}\big(\big\{A_n(\varepsilon) \text{ finitely often}\big\}\big) = 1\,. \tag{2.5}$$

This is why the Borel-Cantelli lemma is so useful in studying almost sure limits.

**Example 1.5 (continued)** *Consider a finite random variable $X$, ie., satisfying $\mathsf{P}(|X| < \infty) = 1$. Then the sequence $(X_k)_{k\ge1}$ defined via $X_k \stackrel{\mathsf{def}}{=} \frac{1}{k}X$ converges to zero with probability one.*

Solution. The previous discussion established exactly (2.5). □

In general, to verify convergence with probability one is not immediate. The following lemma gives a sufficient condition of almost sure convergence.

☞ **Lemma 2.15.** *Let $X_1$, $X_2$, … and $X$ be random variables. If, for every $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} \mathsf{P}\big(|X_n - X| > \varepsilon\big) < \infty\,, \tag{2.6}$$

*then $X_n$ converges to $X$ almost surely.*

Proof. Fix $\varepsilon > 0$ and let $A_n(\varepsilon) = \big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\big\}$. By (2.6), $\sum_n \mathsf{P}\big(A_n(\varepsilon)\big) < \infty$, and, by Lemma 1.6a), only a finite number of $A_n(\varepsilon)$ occur with probability one. This means that for every fixed $\varepsilon > 0$ the event

$$A(\varepsilon) \stackrel{\mathsf{def}}{=} \big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \le \varepsilon \text{ for all } n \text{ large enough}\,\big\}$$

has probability one. By monotonicity ($A(\varepsilon_1) \subset A(\varepsilon_2)$ if $\varepsilon_1 < \varepsilon_2$), the event

$$\big\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\,\big\} = \bigcap_{\varepsilon>0} A(\varepsilon) = \bigcap_{m\ge1} A(1/m)$$

has probability one. The claim follows. □

A straightforward application of Lemma 2.15 improves the WLLN (2.2) and gives the following famous (Borel) Strong Law of Large Numbers (SLLN):

☞ **Theorem 2.16** ($L^4$-SLLN)**.** *Let the variables $X_1$, $X_2$, … be i.i.d. with $\mathsf{E}(X_k) = \mu$ and $\mathsf{E}\big((X_k)^4\big) < \infty$. If $S_n \stackrel{\mathsf{def}}{=} X_1 + X_2 + \cdots + X_n$, then $S_n/n \to \mu$ almost surely, as $n \to \infty$.*

*Proof.* We may and shall suppose[13] that $\mu = \mathsf{E}(X_k) = 0$. Now,

$$\mathsf{E}\big((S_n)^4\big) = \mathsf{E}\Big(\Big(\sum_{k=1}^{n} X_k\Big)^4\Big) = \sum_k \mathsf{E}\big((X_k)^4\big) + 6 \sum_{1\le k<m\le n} \mathsf{E}\big((X_k)^2(X_m)^2\big)$$

so that $\mathsf{E}\big((S_n)^4\big) \le Cn^2$ for some $C \in (0,\infty)$. By Chebyshev's inequality,

$$\mathsf{P}\big(|S_n| > n\varepsilon\big) \le \frac{\mathsf{E}\big((S_n)^4\big)}{(n\varepsilon)^4} \le \frac{C}{n^2\varepsilon^4}$$

and the result follows from (2.6). □

[13] otherwise, consider the centred variables $X'_k = X_k - \mu$ and deduce the result from the relation $\frac{1}{n}S'_n = \frac{1}{n}S_n - \mu$ and linearity of almost sure convergence.

With some additional work,[14] one can obtain the following SLLN (which is due to Kolmogorov):

✍ **Theorem 2.17** ($L^1$-SLLN)**.** *Let $X_1$, $X_2$, ... be i.i.d. r.v. with $\mathsf{E}|X_k| < \infty$. If $\mathsf{E}(X_k) = \mu$ and $S_n \stackrel{\mathsf{def}}{=} X_1 + \cdots + X_n$, then $\frac{1}{n}S_n \to \mu$ almost surely, as $n \to \infty$.*

Notice that verifying almost sure convergence through the Borel-Cantelli lemma (or the sufficient condition (2.6)) is easier than using an explicit construction in the spirit of Example 1.5. We shall see more examples below.

## 2.3 Relations between different types of convergence

It is important to remember the relations between different types of convergence. We know that (Lemma 2.8)

☞
$$X_n \xrightarrow{\mathsf{L^r}} X \qquad \Longrightarrow \qquad X_n \xrightarrow{\mathsf{P}} X\,;$$

one can also show[15]

☞
$$X_n \xrightarrow{\mathsf{a.s.}} X \qquad \Longrightarrow \qquad X_n \xrightarrow{\mathsf{P}} X\,.$$

In addition, according to Example 2.13,

☞
$$X_n \xrightarrow{\mathsf{P}} X \qquad \not\Rightarrow \qquad X_n \xrightarrow{\mathsf{a.s.}} X\,,$$

and the same construction shows that

☞
$$X_n \xrightarrow{\mathsf{L^r}} X \qquad \not\Rightarrow \qquad X_n \xrightarrow{\mathsf{a.s.}} X\,.$$

The following examples fill in the remaining gaps:

☞ **Example 2.18** ($X_n \xrightarrow{\mathsf{L^r}} X \not\Rightarrow X_n \xrightarrow{\mathsf{a.s.}} X$)**.** Let $X_n$ be a sequence of independent random variables such that $\mathsf{P}(X_n = 1) = p_n$, $\mathsf{P}(X_n = 0) = 1 - p_n$. Then

$$X_n \xrightarrow{\mathsf{P}} X \quad \Longleftrightarrow \quad p_n \to 0 \quad \Longleftrightarrow \quad X_n \xrightarrow{\mathsf{L^r}} X \qquad \text{as } n \to \infty,$$

whereas

$$X_n \xrightarrow{\mathsf{a.s.}} X \quad \Longleftrightarrow \quad \sum_n p_n < \infty.$$

In particular, taking $p_n = 1/n$ we deduce the claim. Notice that this example also shows that $X_n \xrightarrow{\mathsf{P}} X \not\Rightarrow X_n \xrightarrow{\mathsf{a.s.}} X$.

☞ **Example 2.19** ($X_n \xrightarrow{\mathsf{P}} X \not\Rightarrow X_n \xrightarrow{\mathsf{L^r}} X$)**.** Let $(\Omega, \mathcal{F}, \mathsf{P})$ be the canonical probability space, recall Definition 2.12. For every $n \ge 1$, define

$$X_n(\omega) \stackrel{\mathsf{def}}{=} e^n \cdot \mathbb{1}_{[0,1/n]}(\omega) \equiv \begin{cases} e^n\,, & 0 \le \omega \le 1/n \\ 0\,, & \omega > 1/n\,. \end{cases}$$

We obviously have $X_n \xrightarrow{\mathsf{a.s.}} 0$ and $X_n \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$; however, for every $r > 0$ $\mathsf{E}|X_n|^r = \frac{e^{nr}}{n} \to \infty$ as $n \to \infty$, ie., $X_n \stackrel{\mathsf{L^r}}{\not\to} 0$. Notice that this example also shows that $X_n \xrightarrow{\mathsf{a.s.}} X \not\Rightarrow X_n \xrightarrow{\mathsf{L^r}} X$.

[14] we will not do this here!
[15] although we shall not do it here!