Convergence of random variables

(telegram style notes)

P.J.C. Spreij

this version: September 6, 2005

1 Introduction

As we know, *random variables* are by definition measurable functions on some underlying measurable space $(\Omega, \mathcal{F})$. If we have a sequence $X_1, X_2, \dots$ of them and we ask for limit behaviour of this sequence, then we have to specify the type of convergence. And since we are dealing with functions, there are many useful types available. In these notes we will treat the best known ones. They are called *a.s. convergence, convergence in probability and convergence in p-th mean.* Another important concept, *weak convergence*, with the Central Limit Theorem as its best known example is treated somewhere else. In this notes we finally arrive at the *Strong law of large numbers* for an *iid* sequence of random variables. This law states that averages of an *iid* sequence converge almost surely to their common expectation.

2 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, In the sequel, unless stated otherwise, we assume that all random variables are defined on this space. The σ-algebra $\mathcal{F}$ on Ω is the collection of *events*. One says that an event F takes place *almost surely*, if $\mathbb{P}(F) = 0$.

For events $E_1, E_2, \dots$ we define

$$\limsup E_n = \bigcap_{n \geq 1} \bigcup_{m \geq n} E_m,$$

and

$$\liminf E_n = \bigcup_{n \geq 1} \bigcap_{m \geq n} E_m.$$

Notice that $(\limsup E_n)^c = \liminf E_n^c$. For the event $\limsup E_n$ one often writes E_n *i.o.* (i.o. means infinitely often) and for the event $\liminf E_n$ one also writes E_n *eventually*.

We will mostly consider *real valued* random variables X, functions $X : \Omega \to \mathbb{R}$ that are *measurable*. Recall that a map $X : \Omega \to \mathbb{R}$ is called measurable if $X^{-1}[B] \in \mathcal{F}$ for all $B \in \mathcal{B}$, the Borel sets of $\mathbb{R}$. Measurability thus depends on the choice of the σ-algebra $\mathcal{F}$ on Ω. There is always a σ-algebra on Ω that makes a given function X measurable, the power set. More interesting is the *smallest* σ-algebra that turns X into a measurable function. This σ-algebra is denoted by $\sigma(X)$ and it is given by $\sigma(X) = \{X^{-1}[B] : B \in \mathcal{B}\}$.

3 Independence

We are used to call two events E and F independent if $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$. Below we extend this to independence of an arbitrary sequence of events, which comes at the end of a sequence of definitions.

Definition 3.1 (i) Let $\mathcal{G}_1, \mathcal{G}_2, \ldots$ be a sequence of sub-σ-algebras of $\mathcal{F}$. One says that this sequence is *independent* if for all $n \in \mathbb{N}$ and $G_{i_k} \in \mathcal{G}_{i_k}$ it holds that

$$\mathbb{P}(G_{i_1} \cap \cdots \cap G_{i_n}) = \prod_{i=1}^{n} \mathbb{P}(G_{i_k}).$$

For finite sequences $\mathcal{G}_1, \ldots, \mathcal{G}_n$ of σ-algebras, independence is defined as independence of the sequence $\mathcal{G}_1, \mathcal{G}_2, \ldots$, where $\mathcal{G}_k = \{\emptyset, \Omega\}$, for $k > n$.
(ii) A sequence $X_1, X_2, \ldots$ of random variables is said to be an *independent* sequence, if the σ-algebras $\mathcal{F}_k = \sigma(X_k)$ $(k \in \mathbb{N})$ are independent.
(iii) A sequence $E_1, E_2, \ldots$ of events is called *independent*, if the random variables $X_k := 1_{E_k}$ are independent.

Remark 3.2 Notice that an independent sequence of events remain independent, if in any subsequence of it, the events E_n are replaced with their complements.

Lemma 3.3 (Borel-Cantelli) *Let $E_1, E_2, \ldots$ be a sequence of events.*
(i) If it has the property that $\sum_{n \geq 1} \mathbb{P}(E_n) < \infty$, then $\mathbb{P}(\limsup E_n) = 0$.
(ii) If $\sum_{n \geq 1} \mathbb{P}(E_n) = \infty$ and if, moreover, the sequence is independent, then $\mathbb{P}(\limsup E_n) = 1$.

Proof (i) Let $U_n = \bigcup_{m \geq n} E_m$. Notice that the sequence (U_n) decreases to $U = \limsup E_n$. Hence we have $\mathbb{P}(U) \leq \mathbb{P}(U_n) \leq \sum_{m \geq n} \mathbb{P}(E_m)$, which converges to zero by assumption.
(ii) We prove that $\mathbb{P}(\liminf E_n^c) = 0$. Let $D_n^N = \bigcap_{m=n}^N E_m^c$ $(N \geq n)$. Notice that for fixed n the sequence $(D_n^N)_{N \geq n}$ decreases to $D_n^\infty := \bigcap_{m=n}^\infty E_m^c$. By independence we obtain $\mathbb{P}(D_n^N) = \prod_{m=n}^N (1 - \mathbb{P}(E_m))$, which is less than $\exp(-\sum_{m=n}^N \mathbb{P}(E_m))$. Hence by taking limits for $N \to \infty$, we obtain for every n that $\mathbb{P}(D_n^\infty) \leq \exp(-\sum_{m=n}^\infty \mathbb{P}(E_m)) = 0$. Finally, we observe that $\liminf E_n^c = \bigcup_{n=1}^\infty D_n^\infty$ and hence $\mathbb{P}(\liminf E_n^c) \leq \sum_{n=1}^\infty \mathbb{P}(D_n^\infty) = 0$. □

Next to (ordinary) independence, we also have the notion of conditional independence. As in definition 3.1 this concept can be defined for infinite sequences of σ-algebras. We only need a special case.

Definition 3.4 Two σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ are called conditionally independent given a third σ-algebra $\mathcal{G}$ if

$$\mathbb{P}(F_1 \cap F_2 | G) = \mathbb{P}(F_1 | G)\mathbb{P}(F_2 | G), \tag{3.1}$$

for all $F_1 \in \mathcal{F}_1$, $F_2 \in \mathcal{F}_2$ and $G \in \mathcal{G}$ with $\mathbb{P}(G) > 0$.

The equality in (3.1) is easily seen to be equivalent to

$$\mathbb{P}(F_1 \cap F_2 \cap G) = \mathbb{P}(F_1 \cap G)\mathbb{P}(F_2 | G). \tag{3.2}$$

Furthermore, (3.2) is obviously also equivalent to

$$\mathbb{P}(F_2 | F_1 \cap G) = \mathbb{P}(F_2 | G),$$

provided that $\mathbb{P}(F_2 \cap G) > 0$.

4 Convergence concepts

Let $X, X_1, X_2, \dots$ be random variables. We have the following definitions of different modes of convergence. We will always assume that the parameter n tends to infinity, unless stated otherwise.

Definition 4.1 (i) If $\mathbb{P}(\omega : X_n(\omega) \to X(\omega)) = 1$, then we say that X_n converges to X almost surely (a.s.).
(ii) If $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ for all $\varepsilon > 0$, then we say that X_n converges to X in probability.
(iii) If $\mathbb{E}\,|X_n - X|^p \to 0$ for some $p > 0$, then we say that X_n converges to X in p-th mean, or in $\mathcal{L}^p$.
For these types of convergence we use the following notations: $X_n \stackrel{a.s.}{\to} X$, $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathcal{L}^p}{\to} X$ respectively.

First we study a bit more in detail almost sure convergence of X_n to X. If this type of convergence takes place we have

$$\mathbb{P}(\omega : \forall \varepsilon > 0 : \exists N : \forall n \geq N : |X_n(\omega) - X(\omega)| < \varepsilon) = 1.$$

But then also (dropping the ω in the notation)

$$\text{for all } \varepsilon > 0\text{: } \mathbb{P}(\exists N : \forall n \geq N : |X_n - X| < \varepsilon) = 1. \tag{4.3}$$

Conversely, if (4.3) holds, we have almost sure convergence. Notice that we can rewrite the probability in (4.3) as $\mathbb{P}(\liminf E_n^\varepsilon) = 1$, with $E_n^\varepsilon = \{|X_n - X| < \varepsilon\}$.

Limits are often required to be unique in an appropriate sense. The natural concept of uniqueness here is that of almost sure uniqueness.

Proposition 4.2 *In each of convergence concepts in definition 4.1 the limit, when it exists, is almost surely unique. This means that if there are two candidate limits X and X', one must have $\mathbb{P}(X = X') = 1$.*

Proof Suppose that $X_n \stackrel{a.s.}{\to} X$ and $X_n \stackrel{a.s.}{\to} X'$. Let Ω_0 be the set of probability one on which $X_n(\omega) \to X(\omega)$ and Ω_0' be the set of probability one on which $X_n(\omega) \to X'(\omega)$. Then also $\mathbb{P}(\Omega_0 \cap \Omega_0') = 1$ and by uniqueness of limits of real numbers we must have that $X(\omega) = X'(\omega)$ for all $\omega \in \Omega_0 \cap \Omega_0'$. Hence $\mathbb{P}(X = X') \geq \mathbb{P}(\Omega_0 \cap \Omega_0') = 1$.
If $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathbb{P}}{\to} X'$, then we have by the triangle inequality for any $\varepsilon > 0$

$$\mathbb{P}(|X - X'| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - X'| > \varepsilon/2),$$

and the right hand side converges to zero by assumption.
Finally we consider the third convergence concept. We need the basic inequality $|a + b|^p \leq c_p(|a|^p + |b|^p)$ (exercise 6.3), where $c_p = \max\{2^{p-1}, 1\}$. This allows us to write $\mathbb{E}\,|X - X'|^p \leq c_p(\mathbb{E}\,|X_n - X|^p + \mathbb{E}\,|X_n - X'|^p)$. It follows that $\mathbb{E}\,|X - X'|^p = 0$ and hence that $\mathbb{P}(X = X') = 1$. □

The following relations hold between the types of convergence introduced in definition 4.1.

Proposition 4.3 *(i) If* $X_n \stackrel{a.s.}{\to} X$*, then* $X_n \stackrel{\mathbb{P}}{\to} X$*.*
(ii) If $X_n \stackrel{\mathcal{L}^p}{\to} X$*, then* $X_n \stackrel{\mathbb{P}}{\to} X$*.*
(iii) If $p > q > 0$ *and* $X_n \stackrel{\mathcal{L}^p}{\to} X$*, then* $X_n \stackrel{\mathcal{L}^q}{\to} X$*.*

Proof (i) Fix $\varepsilon > 0$ and let $A_n = \{|X_n - X| \geq \varepsilon\}$. From (4.3) we know that $\mathbb{P}(\liminf A_n^c) = 1$, or that $\mathbb{P}(\limsup A_n) = 0$. But $A_n \subset U_n := \bigcup_{m \geq n} A_m$ and the U_n form a decreasing sequence with $\limsup A_n$ as its limit. Hence we have $\limsup \mathbb{P}(A_n) \leq \lim \mathbb{P}(U_n) = 0$.
(ii) By Markov's inequality we have

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leq \frac{1}{\varepsilon^p} \mathbb{E}\,|X_n - X|^p,$$

and the result follows.
(iii) Recall that the function $x \mapsto |x|^r$ is convex for $r \geq 1$. Hence, Jensen's inequality ($|\mathbb{E}\,Z|^r \leq \mathbb{E}\,|Z|^r$) yields for $r = p/q$ the inequality $(\mathbb{E}\,|X_n - X|^q)^r \leq \mathbb{E}\,|X_n - X|^p$. □

We close this section with criterions that can be used to decide whether convergence almost surely or in probability takes place.

Proposition 4.4 *(i) If for all* $\varepsilon > 0$ *the series* $\sum_n \mathbb{P}(|X_n - X| > \varepsilon)$ *is convergent, then* $X_n \stackrel{a.s.}{\to} X$*.*
(ii) There is equivalence between
- *(a)* $X_n \stackrel{\mathbb{P}}{\to} X$ *and*
- *(b) every subsequence of* (X_n) *contains a further subsequence that is almost surely convergent to* X*.*

Proof (i) Fix $\varepsilon > 0$ and let $E_n = \{|X_n - X| > \varepsilon\}$. The first part of the Borel-Cantelli lemma (lemma 3.3) gives that $\mathbb{P}(\limsup E_n) = 0$, equivalently $\mathbb{P}(\liminf E_n^c) = 1$, but this is just (4.3).
(ii) Assume that (a) holds, then for any $\varepsilon > 0$ and any subsequence we also have $\mathbb{P}(|X_{n_k} - X| > \varepsilon) \to 0$. Hence for every $p \in \mathbb{N}$, there is $k_p \in \mathbb{N}$ such that $\mathbb{P}(|X_{n_{k_p}} - X| > \varepsilon) \leq 2^{-p}$. Now we apply part (i) of this proposition, which gives us (b). Conversely, assume that (b) holds. We reason by contradiction. Suppose that (a) doesn't hold. Then there exist an $\varepsilon > 0$ and a level $\delta > 0$ such that along some subsequence (n_k) one has

$$\mathbb{P}(|X_{n_k} - X| > \varepsilon) > \delta, \text{ for all } k. \tag{4.4}$$

But the sequence X_{n_k} by assumption has an almost surely convergent subsequence $(X_{n_{k_p}})$, which, by proposition 4.3 (i), also converges in probability. But this contradicts (4.4). □

5 The strong law

The main result of this section is the *strong law of large numbers* for an *iid* sequence of random variables who have a finite expectation. Readers should be familiar with the *weak* law of large numbers for a sequence of random variables that have a finite variance (otherwise, make exercise 6.4 now!). The proof of the theorem 5.2 uses approximations with bounded random variables. The following lemma (sometimes called the *truncation lemma*) prepares for that.

Lemma 5.1 *Let $X_1, X_2, \ldots$ be an iid sequence with $\mathbb{E}\,|X_1| < \infty$ and $\mathbb{E}\,X_1 = \mu$. Put $Y_n = X_n 1_{\{|X_n| \le n\}}$. Then the following assertions hold true.*
(i) $\mathbb{P}(X_n = Y_n \text{ eventually}) = 1$.
(ii) $\mathbb{E}\,Y_n \to \mu$.
(iii) $\sum_{n\ge1} \frac{1}{n^2} \mathrm{Var}\,Y_n < \infty$.

Proof (i) Let $E_n = \{X_n \ne Y_n\} = \{|X_n| > n\}$. We will use the first part of the Borel-Cantelli lemma to conclude that the assertion holds, i.e. we will show that $\mathbb{P}(\limsup E_n) = 0$. We therefore look at $\sum_n \mathbb{P}(E_n)$. Since all X_n have the same distribution as X_1, we also have $\mathbb{P}(|X_n| > n) = \mathbb{P}(X_1| > n)$. Recall the familiar inequality $\mathbb{E}\,|X_1| \ge \sum_{n\ge1} \mathbb{P}(|X_1| > n)$ (exercise 6.18). Using these ingredients we get $\sum_{n\ge1} \mathbb{P}(E_n) \le \mathbb{E}\,|X_1| < \infty$. Indeed, the Borel-Cantelli now give us the result.
(ii) Since X_n has the same distribution as X_1, we also have that Y_n has the same distribution as $X_1 1_{\{|X_1| \le n\}}$. In particular, they have the same expectation. Hence $\mathbb{E}\,Y_n = \mathbb{E}\,X_1 1_{\{|X_1| \le n\}}$, which tends to $\mathbb{E}\,X_1$ in view of theorem A.1 in the appendix (see also exercise 6.17).
(iii) This proof is a little tricky. It is sufficient to show that $\sum_{n\ge1} \frac{1}{n^2} \mathbb{E}\,Y_n^2 < \infty$. The sum is equal to $\sum_{n\ge1} \frac{1}{n^2} \mathbb{E}\,X_1^2 1_{\{|X_1| \le n\}}$. Interchanging expectation and summation gives $\mathbb{E}\,X_1^2 \sum_{n\ge1} \frac{1}{n^2} 1_{\{|X_1| \le n\}}$ and we study the summation. Split it as follows:

$$\sum_{n\ge1} \frac{1}{n^2} 1_{\{|X_1| \le n\}} 1_{\{|X_1| \le 1\}} + \sum_{n\ge1} \frac{1}{n^2} 1_{\{|X_1| \le n\}} 1_{\{|X_1| > 1\}}.$$

The first summation is less than $2 \cdot 1_{\{|X_1| \le 1\}}$. For the second summation we have

$$\begin{aligned}
\sum_{n\ge1} \frac{1}{n^2} 1_{\{|X_1| \le n\}} 1_{\{|X_1| > 1\}} &= \sum_{n \ge |X_1|} \frac{1}{n^2} 1_{\{|X_1| > 1\}} \\
&\le 2 \sum_{n \ge |X_1|} \int_n^{n+1} \frac{1}{x^2}\,dx\, 1_{\{|X_1| > 1\}} \\
&\le 2 \int_{|X_1|}^{\infty} \frac{1}{x^2}\,dx\, 1_{\{|X_1| > 1\}} \\
&= 2 \frac{1}{|X_1|}\, 1_{\{|X_1| > 1\}}.
\end{aligned}$$

Hence

$$\mathbb{E}\, X_1^2 \sum_{n\geq 1} \frac{1}{n^2} 1_{\{|X_1|\leq n\}} \leq 2\mathbb{E}\, X_1^2 (1_{\{|X_1|\leq 1\}} + \frac{1}{|X_1|} 1_{\{|X_1|>1\}})$$
$$\leq 2(\mathbb{E}\, |X_1| 1_{\{|X_1|\leq 1\}} + \mathbb{E}\, |X_1| 1_{\{|X_1|>1\}}) = 2\mathbb{E}\, |X_1|,$$

which is finite by assumption. □

Here is the announced strong law of large numbers.

Theorem 5.2 *Let $X_1, X_2, \ldots$ be a sequence of* iid *random variables and assume that $\mathbb{E}\, |X_1| < \infty$. Let $\mu = \mathbb{E}\, X_1$, then*

$$\overline{X}_n \to \mu \text{ a.s.} \tag{5.5}$$

Proof Let us first assume that the X_n are nonnegative. Put $Y_n = X_n 1_{\{X_n \leq n\}}$. Then

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} Y_k + \frac{1}{n} \sum_{k=1}^{n} (X_k - Y_k).$$

Notice that on a set of probability one the sum $\sum_{k=1}^{n} (X_k(\omega) - Y_k(\omega))$ contains only finitely many nonzero terms (this follows from lemma 5.1 (i)), so that it is sufficient to show that

$$\overline{Y}_n \stackrel{a.s.}{\to} \mu. \tag{5.6}$$

Fix $\alpha > 1$, $\beta_n = [\alpha^n]$ and put

$$\eta_n = \frac{1}{\beta_n} \sum_{k=1}^{\beta_n} Y_k.$$

We first show that

$$\eta_n - \mathbb{E}\, \eta_n \stackrel{a.s.}{\to} 0 \tag{5.7}$$

by applying proposition 4.4(i). Below we need the following technical result. There exists a constant C_α such that for all $i \geq 1$ it holds that $\sum_{n:\beta_n \geq i} \frac{1}{\beta_n^2} \leq \frac{C_\alpha}{i^2}$ (exercise 6.10). Consider for any $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}(|\eta_n - \mathbb{E}\, \eta_n| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \operatorname{Var} \eta_n$$
$$= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{\beta_n^2} \sum_{i=1}^{\beta_n} \operatorname{Var} Y_i$$
$$= \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} (\sum_{n:\beta_n \geq i} \frac{1}{\beta_n^2}) \operatorname{Var} Y_i$$
$$\leq C_\alpha \sum_{i=1}^{\infty} \frac{1}{i^2} \operatorname{Var} Y_i,$$

which is finite by lemma 5.1(iii). Hence proposition 4.4(i) yields the result. It is easy to show that $\mathbb{E}\,\eta_n$ converges to μ (exercise 6.11), and we then conclude from (5.7) that

$$\eta_n \stackrel{a.s.}{\to} \mu. \tag{5.8}$$

Recall that η_n depends on $\alpha > 1$. We proceed by showing that, by a limit argument, the result (5.8) is also valid for $\alpha = 1$.
For every n, let $m = m(n) = \inf\{k : \beta_k > n\}$. Then $\beta_m > n \geq \beta_{m-1}$ and therefore

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \leq \frac{1}{\beta_{m-1}}\sum_{i=1}^{\beta_m} Y_i = \frac{\beta_m}{\beta_{m-1}}\eta_m.$$

Using (5.8) and $\frac{\beta_m}{\beta_{m-1}} \to \alpha$ as $m \to \infty$, we conclude that

$$\limsup \overline{Y}_n \leq \alpha\mu\, a.s.,$$

and, since this is true for every $\alpha > 1$, we must also have

$$\limsup \overline{Y}_n \leq \mu\, a.s. \tag{5.9}$$

By a similar argument we have

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \geq \frac{1}{\beta_{m-1}}\sum_{i=1}^{\beta_{m-1}} Y_i = \frac{\beta_{m-1}}{\beta_m}\eta_{m-1}.$$

Using (5.8) again, we conclude that

$$\liminf \overline{Y}_n \geq \frac{1}{\alpha}\mu\, a.s.,$$

and then we must also have

$$\liminf \overline{Y}_n \leq \mu\, a.s. \tag{5.10}$$

Combining (5.9) and (5.10) yields (5.6) for nonnegative X_i. Finally, for arbitrary X_i we proceed as follows. For every real number x we define $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$. Then $x = x^+ - x^-$. Similar notation applies to random variables. We apply the above results to the averages of the X_i^+ and X_i^- to get

$$\overline{X}_n \stackrel{a.s.}{\to} \mathbb{E}\,X_1^+ - \mathbb{E}\,X_1^- = \mathbb{E}\,X_1 = \mu.$$

□

The next proposition shows that if a.s. convergence of the averages to a constant takes place, this constant must be the (common) expectation.

Proposition 5.3 *Let $X_1, X_2, \dots$ be a sequence of* iid *random variables, and assume that*

$$\overline{X}_n \to \mu \; a.s. \tag{5.11}$$

for some constant $\mu \in \mathbb{R}$. Then $\mathbb{E}\,|X_1|$ is finite and $\mu = \mathbb{E}\,X_1$.

Proof Write $\overline{X}_n = \frac{n-1}{n}\overline{X}_{n-1} + \frac{X_n}{n}$ to conclude that $\frac{X_n}{n} \overset{a.s.}{\to} 0$. In particular we have $\frac{|X_n|}{n} \leq 1$ *eventually*. I.e. $\mathbb{P}(\liminf\{\frac{|X_n|}{n} \leq 1\}) = 1$, or $\mathbb{P}(\limsup\{\frac{|X_n|}{n} > 1\}) = 0$. Using the second half of the Borel-Cantelli lemma (lemma 3.3), we conclude that $\sum_n \mathbb{P}(\frac{|X_n|}{n} \leq 1) = \infty$, and thus $\sum_n \mathbb{P}(|X_1| \leq n) = \infty$. Since $E|X_1| \leq \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n)$ (exercise 6.18), we thus have $\mathbb{E}\,|X_1| < \infty$ and theorem 5.2 then yields that $\mu = \mathbb{E}\,X_1$. □

The assertion of theorem 5.2 is stated under the assumption of the existence of a finite expectation. In the case where one deals with nonnegative random variables, this assumption can be dropped.

Theorem 5.4 *Let $X_1, X_2, \dots$ be a sequence of* nonnegative iid *random variables, defined on a common probability space. Let $\mu = \mathbb{E}\,X_1 \leq \infty$, then*

$$\overline{X}_n \to \mu \; a.s. \tag{5.12}$$

Proof We only need to consider the case where $\mu = \infty$. Fix $N \in \mathbb{N}$ and let $X_n^N = X_n 1_{\{X_n \leq N\}}$, $n \in \mathbb{N}$. Then theorem 5.2 applies and we have that $\frac{1}{n}\sum_{k=1}^n X_k^N \overset{a.s.}{\to} \mathbb{E}\,X_1^N$. But $\overline{X}_n \geq \frac{1}{n}\sum_{k=1}^n X_k^N$ and hence $\ell := \liminf \overline{X}_n \geq \mathbb{E}\,X_1^N$ *a.s.* for all N, and thus also $\ell \geq \lim_{N\to\infty} EX_1^N$ *a.s.* But, by theorem A.1, the latter limit is equal to $\mathbb{E}\,X_1$. Hence $\ell = \infty$ *a.s.* □

6 Exercises

6.1 Let E_1 and E_2 be two events such that $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$. Show that these events are independent in the sense of definition 3.1.

6.2 Let $E_1, E_2, \dots$ be events and let $X_n = 1_{E_n}$, $n \geq 1$. Show that $\liminf X_n = 1_{\liminf E_n}$ and that $\limsup X_n = 1_{\limsup E_n}$.

6.3 Show that for any two real number a and b and for any $p > 0$ it holds that $(|a| + |b|)^p \leq \max\{2^{p-1}, 1\}(|a|^p + |b|^p)$.

6.4 Let $X_1, X_2, \dots$ be an *iid* sequence of random variables with common finite variance σ^2 and expectation μ and put $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Show that $\mathrm{Var}\,\overline{X}_n = \frac{\sigma^2}{n}$ and deduce from Chebychev's inequality that $\overline{X}_n \overset{\mathbb{P}}{\to} \mu$.

6.5 Let $X, X_1, X_2, \dots$ be random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Show that the set $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ is an *event*, i.e. it is measurable.

6.6 Suppose that the real random variables $X_1, X_2, \dots$ are defined on a common probability space and that they are *iid* with a uniform distribution on $[0,1]$. Let $M_n = \max\{X_1, \dots, X_n\}$. Show that $M_n \xrightarrow{\mathbb{P}} 1$ and that even $M_n \xrightarrow{a.s.} 1$.

6.7 Give an example of a sequence of random variables that converge in probability, but not almost surely.

6.8 Let $X_1, X_2, \dots$ be an a.s. bounded sequence of random variables $\mathbb{P}(|X_n| \leq M) = 1$, for some real number M. Assume that for some random variable X one has $X_n \xrightarrow{\mathbb{P}} X$. Show that also $\mathbb{P}(|X| \leq M) = 1$ and that for all $p > 0$ one has $X_n \xrightarrow{\mathcal{L}^p} X$.

6.9 Let $Z_1, Z_2, \dots$ be an *iid* sequence of standard normals and let $S_n = \sum_{k=1}^n Z_k$. Let $a, b \in \mathbb{R}$ and define

$$X_n = \exp(aS_n - bn).$$

(i) Express for every $\varepsilon > 0$ the probability $\mathbb{P}(X_n > \varepsilon)$ in the cumulative distribution function of Z_1 and deduce that $X_n \xrightarrow{\mathbb{P}} 0$ iff $b > 0$.
(ii) Show that $\mathbb{E}\exp(\lambda Z_1) = \exp(\frac{1}{2}\lambda^2)$ for $\lambda \in \mathbb{R}$ and compute $\mathbb{E}\, X_n^p$, $(p > 0)$.
(iii) Show that $X_n \xrightarrow{\mathcal{L}^p} 0$ iff $p < 2b/a^2$.
(iv) Show that $X_n \xrightarrow{a.s.} 0$ iff $b > 0$. *Hint:* Use Markov's inequality for X_n^p.

6.10 Let $\alpha > 1$ and $\beta_k = [\alpha^k]$. Show that there exists a constant C_α such that for all integers i on has

$$\sum_{n:\beta_n \geq i} \leq \frac{C_\alpha}{i^2}.$$

Show also that $\frac{\beta_{k+1}}{\beta_k} \to \alpha$.

6.11 Let x_n be real numbers with $x_n \to x$. Let $y_n = \frac{1}{n}\sum_{i=1}^n x_i$. Show that $y_n \to x$. Take the η_n from the proof of theorem 5.2. Show that $\mathbb{E}\,\eta_n \to \mu$.

6.12 Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables with $\mathbb{E}\, X_1^2 < \infty$. The aim is to show is that both $\overline{X}_n \xrightarrow{\mathcal{L}^2} \mu$ where $\mu = \mathbb{E}\, X_1$ and $\overline{X}_n \xrightarrow{a.s.} \mu$.
(i) Show the $\mathcal{L}^2$ convergence.
(ii) Use Chebychev's inequality to show that $\sum_n \mathbb{P}(|\overline{X}_{n^2} - \mu| > \varepsilon) < \infty$ and deduce form a well known lemma that $\overline{X}_{n^2} \xrightarrow{a.s.} \mu$.
(iii) Show the almost sure convergence of $\overline{X}_n$ by "filling the gaps".

6.13 Let $X_1, X_2, \dots$ be real random variables and $g : \mathbb{R} \to \mathbb{R}$ a uniformly continuous function. Show that $g(X_n) \xrightarrow{\mathbb{P}} g(X)$ if $X_n \xrightarrow{\mathbb{P}} X$. What can be said of the $g(X_n)$ if $X_n \xrightarrow{a.s.} X$?

6.14 Let $X_1, Y_1, X_2, Y_2, \dots$ be an i.i.d. sequence whose members have a uniform distribution on $[0,1]$ and let $f : [0,1] \to [0,1]$ be continuous. Define $Z_i = 1_{\{f(X_i) > Y_i\}}$.
(i) Show that $\frac{1}{n} \sum_{i=1}^n Z_i \to \int_0^1 f(x)\,dx$ a.s.
(ii) Show that $\mathbb{E}\,(\frac{1}{n} \sum_{i=1}^n Z_i - \int_0^1 f(x)\,dx)^2 \leq \frac{1}{4n}$.
(iii) Explain why these two results are useful.

6.15 If $X_n \xrightarrow{\mathbb{P}} X$ and g is a continuous function, then also $g(X_n) \xrightarrow{\mathbb{P}} g(X)$. Show this.

6.16 Assume that $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$. Show that $X_n + Y_n \xrightarrow{\mathbb{P}} X$ and similar statements for products and ratios. What about convergence of the pairs (X_n, Y_n)? How do the results look for almost sure convergence instead of convergence in probability.

6.17 Let $X_1, X_2, \dots$ be nonnegative random variables such that $X_{n+1} \geq X_n$ and let $X = \lim X_n$ a.s. Then $\lim \mathbb{E}\, X_n \leq \mathbb{E}\, X$. Let X be a nonnegative random variable and $X_n = X\, 1_{\{X \leq n\}}$, $n \in \mathbb{N}$. Show also that $\mathbb{E}\, X_n \to \mathbb{E}\, X$, when X has density, or when X is discrete. (This is a special case of theorem A.1).

6.18 Prove for every random variable X the double inequalities

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| > n) \leq \mathbb{E}\,|X| \leq \sum_{n=0}^{\infty} \mathbb{P}(|X| > n)$$

and

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n) \leq \mathbb{E}\,|X| \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n).$$

A The monotone convergence theorem

Although familiarity with measure theory is not assumed in this course, occasionally we need one of the main theorems that will be provided in any introductory course in measure theory. One that deals with interchanging limits and expectation. For a proof we refer to a course in measure theory, although a part of the proof of theorem A.1 is elementary (see exercise 6.17).

Theorem A.1 (Monotone convergence theorem) *Let $X_1, X_2, \dots$ be nonnegative random variables with property that $X_{n+1} \geq X_n$ and let $X = \lim X_n$. Then $\lim \mathbb{E}\, X_n = \mathbb{E}\, X$.*