17. Convergence of Random Variables

In elementary mathematics courses (such as Calculus) one speaks of the convergence of functions: $f_n : \mathbf{R} \to \mathbf{R}$, then $\lim_{n\to\infty} f_n = f$ if $\lim_{n\to\infty} f_n(x) = f(x)$ for all x in $\mathbf{R}$. This is called *pointwise convergence of functions.* A random variable is of course a function ($X : \Omega \to \mathbf{R}$ for an abstract space Ω), and thus we have the same notion: a sequence $X_n : \Omega \to \mathbf{R}$ *converges pointwise to* X *if* $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, for all $\omega \in \Omega$. This natural definition is surprisingly useless in probability. The next example gives an indication why.

Example 1: Let X_n be an i.i.d. sequence of random variables with $P(X_n = 1) = p$ and $P(X_n = 0) = 1 - p$. For example we can imagine tossing a slightly unbalanced coin (so that $p > \frac{1}{2}$) repeatedly, and $\{X_n = 1\}$ corresponds to heads on the n^{th} toss and $\{X_n = 0\}$ corresponds to tails on the n^{th} toss. In the "long run", we would expect the proportion of heads to be p; this would justify our model that claims the probability of heads is p. Mathematically we would want

$$\lim_{n\to\infty} \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} = p \quad \text{for all } \omega \in \Omega.$$

This simply does not happen! For example let $\omega_0 = \{T, T, T, \ldots\}$, the sequence of all tails. For this ω_0,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j(\omega_0) = 0.$$

More generally we have the event

$$A = \{\omega : \text{ only a finite number of heads occur}\}.$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j(\omega) = 0 \text{ for all } \omega \in A.$$

We readily admit that the event A is very unlikely to occur. Indeed, we can show (Exercise 17.13) that $P(A) = 0$. In fact, what we will eventually show (see the Strong Law of Large Numbers [Chapter 20]) is that

$$P\left(\left\{\omega : \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j(\omega) = p\right\}\right) = 1.$$

This type of convergence of random variables, where we do not have convergence for *all* ω but do have convergence for *almost all* ω (i.e., the set of ω where we do have convergence has probability one), is what typically arises.

Caveat: In this chapter we will assume that all random variables are defined on a given, fixed probability space $(\Omega, \mathcal{A}, P)$ and takes values in $\mathbf{R}$ or $\mathbf{R}^n$. We also denote by $|x|$ the Euclidean norm of $x \in \mathbf{R}^n$.

Definition 17.1. *We say that a sequence of random variables* $(X_n)_{n\geq 1}$ converges almost surely *to a random variable* X *if*

$$N = \left\{\omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\right\} \text{ has } P(N) = 0.$$

Recall that the set N *is called a* null set*, or a negligible set.*

Note that

$$N^c = \Lambda = \left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} \text{ and then } P(\Lambda) = 1.$$

We usually abbreviate almost sure convergence by writing

$$\lim_{n\to\infty} X_n = X \text{ a.s.}$$

We have given an example of almost sure convergence from coin tossing preceding this definition.

Just as we defined almost sure convergence because it naturally occurs when "pointwise convergence" (for all "points") fails, we need to introduce two more types of convergence. These next two types of convergence also arise naturally when a.s. convergence fails, and they are also useful as tools to help to show that a.s. convergence holds.

Definition 17.2. *A sequence of random variables* $(X_n)_{n\geq 1}$ converges in L^p *to* X *(where*$1 \leq p < \infty$*) if* $|X_n|$, $|X|$ *are in* L^p *and:*

$$\lim_{n\to\infty} E\{|X_n - X|^p\} = 0.$$

Alternatively one says X_n converges to X in p^{th} mean*, and one writes*

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

The most important cases for convergence in p^{th} mean are when $p = 1$ and when $p = 2$. When $p = 1$ and all r.v.'s are one-dimensional, we have

$|E\{X_n - X\}| \leq E\{|X_n - X|\}$ and $|E\{|X_n|\} - E\{|X|\}| \leq E\{|X_n - X|\}$ because $||x| - |y|| \leq |x - y|$. Hence

$$X_n \xrightarrow{L^1} X \quad \text{implies} \quad E\{X_n\} \to E\{X\} \quad \text{and} \quad E\{|X_n|\} \to E\{|X|\}. \qquad (17.1)$$

Similarly, when $X_n \xrightarrow{L^p} X$ for $p \in (1, \infty)$, we have that $E\{|X_n|^p\}$ converges to $E\{|X|^p\}$: see Exercise 17.14 for the case $p = 2$.

Definition 17.3. *A sequence of random variables* $(X_n)_{n\geq 1}$ converges in probability *to* X *if for any* $\varepsilon > 0$ *we have*

$$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

This is also written

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

and denoted

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

Using the epsilon-delta definition of a limit, one could alternatively say that X_n tends to X in probability if for any $\varepsilon > 0$, any $\delta > 0$, there exists $N = N(\delta)$ such that

$$P(|X_n - X| > \varepsilon) < \delta$$

for all $n \geq N$.

Before we establish the relationships between the different types of convergence, we give a surprisingly useful small result which characterizes convergence in probability.

Theorem 17.1. $X_n \xrightarrow{P} X$ *if and only if*

$$\lim_{n\to\infty} E\left\{\frac{|X_n - X|}{1 + |X_n - X|}\right\} = 0.$$

Proof. There is no loss of generality by taking $X = 0$. Thus we want to show $X_n \xrightarrow{P} 0$ if and only if $\lim_{n\to\infty} E\{\frac{|X_n|}{1+|X_n|}\} = 0$. First suppose that $X_n \xrightarrow{P} 0$. Then for any $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0$. Note that

$$\frac{|X_n|}{1 + |X_n|} \leq \frac{|X_n|}{1 + |X_n|} 1_{\{|X_n| > \varepsilon\}} + \varepsilon 1_{\{|X_n| \leq \varepsilon\}} \leq 1_{\{|X_n| > \varepsilon\}} + \varepsilon.$$

Therefore

$$E\left\{\frac{|X_n|}{1 + |X_n|}\right\} \leq E\left\{1_{\{|X_n| > \varepsilon\}}\right\} + \varepsilon = P(|X_n| > \varepsilon) + \varepsilon.$$

Taking limits yields

$$\lim_{n\to\infty} E\left\{\frac{|X_n|}{1+|X_n|}\right\} \leq \varepsilon;$$

since ε was arbitrary we have $\lim_{n\to\infty} E\{\frac{|X_n|}{1+|X_n|}\} = 0$.

Next suppose $\lim_{n\to\infty} E\{\frac{|X_n|}{1+|X_n|}\} = 0$. The function $f(x) = \frac{x}{1+x}$ is strictly increasing. Therefore

$$\frac{\varepsilon}{1+\varepsilon} 1_{\{|X_n|>\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|} 1_{\{|X_n|>\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|}.$$

Taking expectations and then limits yields

$$\frac{\varepsilon}{1+\varepsilon} \lim_{n\to\infty} P(|X_n| > \varepsilon) \leq \lim_{n\to\infty} E\left\{\frac{|X_n|}{1+|X_n|}\right\} = 0.$$

Since $\varepsilon > 0$ is fixed, we conclude $\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0$. □

Remark: What this theorem says is that $X_n \xrightarrow{P} X$ iff[1] $E\{f(|X_n - X|)\} \to 0$ for the function $f(x) = \frac{|x|}{1+|x|}$. A careful examination of the proof shows that the same equivalence holds for any function f on $\mathbf{R}_+$ which is bounded, strictly increasing on $[0,\infty)$, continuous, and with $f(0) = 0$. For example we have $X_n \xrightarrow{P} X$ iff $E\{|X_n - X| \wedge 1\} \to 0$ and also iff $E\{\arctan(|X_n - X|)\} \to 0$.

The next theorem shows that convergence in probability is the weakest of the three types of convergence (a.s., L^p, and probability).

Theorem 17.2. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables.*

a) *If $X_n \xrightarrow{L^p} X$, then $X_n \xrightarrow{P} X$.*
b) *If $X_n \xrightarrow{\text{a.s.}} X$, then $X_n \xrightarrow{P} X$.*

Proof. (a) Recall that for an event A, $P(A) = E\{1_A\}$, where 1_A is the indicator function of the event A. Therefore,

$$P\{|X_n - X| > \varepsilon\} = E\left\{1_{\{|X_n - X|>\varepsilon\}}\right\}.$$

Note that $\frac{|X_n - X|^p}{\varepsilon^p} > 1$ on the event $\{|X_n - X| > \varepsilon\}$, hence

$$\begin{aligned} &\leq E\left\{\frac{|X_n - X|^p}{\varepsilon^p} 1_{\{|X_n - X|>\varepsilon\}}\right\} \\ &= \frac{1}{\varepsilon^p} E\left\{|X_n - X|^p 1_{\{|X_n - X|>\varepsilon\}}\right\}, \end{aligned}$$

and since $|X_n - X|^p \geq 0$ always, we can simply drop the indicator function to get:

[1] The notation *iff* is a standard notation shorthand for "if and only if"

$$\leq \frac{1}{\varepsilon^p} E\{|X_n - X|^p\}.$$

The last expression tends to 0 as n tends to ∞ (for fixed $\varepsilon > 0$), which gives the result.

(b) Since $\frac{|X_n - X|}{1+|X_n - X|} \leq 1$ always, we have

$$\lim_{n\to\infty} E\left\{\frac{|X_n - X|}{1 + |X_n - X|}\right\} = E\left\{\lim_{n\to\infty} \frac{|X_n - X|}{1 + |X_n - X|}\right\} = E\{0\} = 0$$

by Lebegue's Dominated Convergence Theorem (9.1(f)). We then apply Theorem 17.1. □

The converse to Theorem 17.2 is not true; nevertheless we have two partial converses. The most delicate one concerns the relation with a.s. convergence, and goes as follows:

Theorem 17.3. *Suppose* $X_n \xrightarrow{P} X$. *Then there exists a* subsequence n_k *such that* $\lim_{k\to\infty} X_{n_k} = X$ *almost surely.*

Proof. Since $X_n \xrightarrow{P} X$ we have that $\lim_{n\to\infty} E\{\frac{|X_n - X|}{1+|X_n - X|}\} = 0$ by Theorem 17.1. Choose a subsequence n_k such that $E\{\frac{|X_{n_k} - X|}{1+|X_{n_k} - X|}\} < \frac{1}{2^k}$. Then $\sum_{k=1}^{\infty} E\{\frac{|X_{n_k} - X|}{1+|X_{n_k} - X|}\} < \infty$ and by Theorem 9.2 we have that $\sum_{k=1}^{\infty} \frac{|X_{n_k} - X|}{1+|X_{n_k} - X|} < \infty$ a.s.; since the general term of a convergent series must tend to zero, we conclude

$$\lim_{n\to n} |X_{n_k} - X| = 0 \text{ a.s.}$$

□

Remark 17.1. Theorem 17.3 can also be proved fairly simply using the Borel–Cantelli Theorem (Theorem 10.5).

Example 2: $X_n \xrightarrow{P} X$ does not necessarily imply that X_n converges to X almost surely. For example take $\Omega = [0,1]$, $\mathcal{A}$ the Borel sets on $[0,1]$, and P the uniform probability measure on $[0,1]$. (That is, P is just Lebesgue measure restricted to the interval $[0,1]$.) Let A_n be any interval in $[0,1]$ of length a_n, and take $X_n = 1_{A_n}$. Then $P(|X_n| > \varepsilon) = a_n$, and as soon as $a_n \to 0$ we deduce that $X_n \xrightarrow{P} 0$ (that is, X_n tends to 0 in probability). More precisely, let $X_{n,j}$ be the indicator of the interval $[\frac{j-1}{n}, \frac{j}{n}]$, $1 \leq j \leq n$, $n \geq 1$. We can make one sequence of the $X_{n,j}$ by ordering them first by increasing n, and then for each fixed n by increasing j. Call the new sequence Y_m. Thus the sequence would be:

$$\begin{array}{llllllll} X_{1,1}, & X_{2,1}, & X_{2,2}, & X_{3,1}, & X_{3,2}, & X_{3,3}, & X_{4,1}, & \ldots \\ Y_1, & Y_2, & Y_3, & Y_4, & Y_5, & Y_6, & Y_7, & \ldots \end{array}$$

Note that for each ω and every n, there exists a j such that $X_{n,j}(\omega) = 1$. Therefore $\limsup_{m\to\infty} Y_m = 1$ a.s., while $\liminf_{m\to\infty} Y_m = 0$ a.s. Clearly then the sequence Y_m does not converge a.s. However Y_n is the indicator of an interval whose length a_n goes to 0 as $n \to \infty$, so the sequence Y_n does converge to 0 in probability.

The second partial converse of Theorem 17.2 is as follows:

Theorem 17.4. *Suppose $X_n \xrightarrow{P} X$ and also that $|X_n| \leq Y$, all n, and $Y \in L^p$. Then $|X|$ is in L^p and $X_n \xrightarrow{L^p} X$.*

Proof. Since $E\{|X_n|^p\} \leq E\{Y^p\} < \infty$, we have $X_n \in L^p$. For $\varepsilon > 0$ we have

$$\begin{aligned}\{|X| > Y + \varepsilon\} &\subset \{|X| > |X_n| + \varepsilon\}\\ &\subset \{|X| - |X_n| > \varepsilon\}\\ &\subset \{|X - X_n| > \varepsilon\},\end{aligned}$$

hence

$$P(|X| > Y + \varepsilon) \leq P(|X - X_n| > \varepsilon),$$

and since this is true for each n, we have

$$P(|X| > Y + \varepsilon) \leq \lim_{n\to\infty} P(|X - X_n| > \varepsilon) = 0,$$

by hypothesis. This is true for each $\varepsilon > 0$, hence

$$P(|X| > Y) \leq \lim_{m\to\infty} P(|X| > Y + \frac{1}{m}) = 0,$$

from which we get $|X| \leq Y$ a.s. Therefore $X \in L^p$ too.

Suppose now that X_n does not converge to X in L^p. There is a subsequence (n_k) such that $E\{|X_{n_k} - X|^p\} \geq \varepsilon$ for all k, and for some $\varepsilon > 0$. The subsequence X_{n_k} trivially converges to X in probability, so by Theorem 17.3 it admits a further subsequence $X_{n_{k_j}}$ which converges a.s. to X. Now, the r.v.'s $X_{n_{k_j}} - X$ tend a.s. to 0 as $j \to \infty$, while staying smaller than $2Y$, so by Lebesgue's Dominated Convergence we get that $E\{|X_{n_{k_j}} - X|^p\} \to 0$, which contradicts the property that $E\{|X_{n_k} - X|^p\} \geq \varepsilon$ for all k: hence we are done. □

The next theorem is elementary but also quite useful to keep in mind.

Theorem 17.5. *Let f be a continuous function.*

a) *If $\lim_{n\to\infty} X_n = X$ a.s., then $\lim_{n\to\infty} f(X_n) = f(X)$ a.s.*
b) *If $X_n \xrightarrow{P} X$, then $f(X_n) \xrightarrow{P} f(X)$.*

Proof. (a) Let $N = \{\omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}$. Then $P(N) = 0$ by hypothesis. If $\omega \notin N$, then

$$\lim_{n\to\infty} f(X_n(\omega)) = f\left(\lim_{n\to\infty} X_n(\omega)\right) = f(X(\omega)),$$

where the first equality is by the continuity of f. Since this is true for any $\omega \notin N$, and $P(N) = 0$, we have the almost sure convergence.

(b) For each $k > 0$, let us set:

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|f(X_n) - f(X)| > \varepsilon, |X| \leq k\} \cup \{|X| > k\}. \quad (17.2)$$

Since f is continuous, it is uniformly continuous on any bounded interval. Therefore for our ε given, there exists a $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ if $|x - y| \leq \delta$ for x and y in $[-k, k]$. This means that

$$\{|f(X_n) - f(X)| > \varepsilon, |X| \leq k\} \subset \{|X_n - X| > \delta, |X| \leq k\} \subset \{|X_n - X| > \delta\}.$$

Combining this with (17.2) gives

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|X_n - X| > \delta\} \cup \{|X| > k\}. \quad (17.3)$$

Using simple subadditivity ($P(A \cup B) \leq P(A) + P(B)$) we obtain from (17.3):

$$P\{|f(X_n) - f(X)| > \varepsilon\} \leq P(|X_n - X| > \delta) + P(|X| > k).$$

However $\{|X| > k\}$ tends to the empty set as k increases to ∞ so $\lim_{k\to\infty} P(|X| > k) = 0$. Therefore for $\gamma > 0$ we choose k so large that $P(|X| > k) < \gamma$. Once k is fixed, we obtain the δ of (17.3), and therefore

$$\lim_{n\to\infty} P(|f(X_n) - f(X)| > \varepsilon) \leq \lim_{n\to\infty} P(|X_n - X| > \delta) + \gamma = \gamma.$$

Since $\gamma > 0$ was arbitrary, we deduce the result. $\square$

Exercises for Chapter 17

17.1 Let $X_{n,j}$ be as given in Example 2. Let $Z_{n,j} = n^{\frac{1}{p}} X_{n,j}$. Let Y_m be the sequence obtained by ordering the $Z_{n,j}$ as was done in Example 2. Show that Y_m tends to 0 in probability but that $(Y_m)_{m\geq 1}$ does not tend to 0 in L^p, although each Y_n belongs to L^p.

17.2 Show that Theorem 17.5(b) is false in general if f is not assumed to be continuous. (*Hint:* Take $f(x) = 1_{\{0\}}(x)$ and the X_n's tending to 0 in probability.)

17.3 Let X_n be i.i.d. random variables with $P(X_n = 1) = \frac{1}{2}$ and $P(X_n = -1) = \frac{1}{2}$. Show that

$$\frac{1}{n}\sum_{j=1}^{n} X_j$$

converges to 0 in probability. (*Hint:* Let $S_n = \sum_{j=1}^{n} X_j$, and use Chebyshev's inequality on $P\{|S_n| > n\varepsilon\}$.)

17.4 Let X_n and S_n be as in Exercise 17.3. Show that $\frac{1}{n^2} S_{n^2}$ converges to zero a.s. (*Hint:* Show that $\sum_{n=1}^{\infty} P\{\frac{1}{n^2}|S_{n^2}| > \varepsilon\} < \infty$ and use the Borel-Cantelli Theorem.)

17.5 * Suppose $|X_n| \leq Y$ a.s., each n, $n = 1, 2, 3 \ldots$ Show that $\sup_n |X_n| \leq Y$ a.s. also.

17.6 Let $X_n \xrightarrow{P} X$. Show that the characteristic functions φ_{X_n} converge pointwise to φ_X (*Hint:* Use Theorem 17.4.)

17.7 Let $X_1, \ldots, X_n$ be i.i.d. Cauchy random variables with parameters $\alpha = 0$ and $\beta = 1$. (That is, their density is $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.) Show that $\frac{1}{n}\sum_{j=1}^{n} X_j$ also has a Cauchy distribution. (*Hint:* Use Characteristic functions: See Exercise 14.1.)

17.8 Let $X_1, \ldots, X_n$ be i.i.d. Cauchy random variables with parameters $\alpha = 0$ and $\beta = 1$. Show that there is no constant γ such that $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{P} \gamma$. (*Hint:* Use Exercise 17.7.) Deduce that there is no constant γ such that $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j = \gamma$ a.s. as well.

17.9 Let $(X_n)_{n\geq 1}$ have finite variances and zero means (i.e., $\mathrm{Var}(X_n) = \sigma^2_{X_n} < \infty$ and $E\{X_n\} = 0$, all n). Suppose $\lim_{n\to\infty} \sigma^2_{X_n} = 0$. Show X_n converges to 0 in L^2 and in probability.

17.10 Let X_j be i.i.d. with finite variances and zero means. Let $S_n = \sum_{j=1}^{n} X_j$. Show that $\frac{1}{n} S_n$ tends to 0 in both L^2 and in probability.

17.11 * Suppose $\lim_{n\to\infty} X_n = X$ a.s. and $|X| < \infty$ a.s. Let $Y = \sup_n |X_n|$. Show that $Y < \infty$ a.s.

17.12 * Suppose $\lim_{n\to\infty} X_n = X$ a.s. Let $Y = \sup_n |X_n - X|$. Show $Y < \infty$ a.s. (see Exercise 17.11), and define a new probability measure Q by

$$Q(A) = \frac{1}{c} E\left\{1_A \frac{1}{1+Y}\right\}, \text{ where } c = E\left\{\frac{1}{1+Y}\right\}.$$

Show that X_n tends to X in L^1 under the probability measure Q.

17.13 Let A be the event described in Example 1. Show that $P(A) = 0$. (*Hint:* Let

$$A_n = \{ \text{ Heads on } n^{\text{th}} \text{ toss } \}.$$

Show that $\sum_{n=1}^{\infty} P(A_n) = \infty$ and use the Borel-Cantelli Theorem (Theorem 10.5.))

17.14 Let X_n and X be real-valued r.v.'s in L^2, and suppose that X_n tends to X in L^2. Show that $E\{X_n^2\}$ tends to $E\{X^2\}$ (*Hint:* use that $|x^2 - y^2| = (x-y)^2 + 2|y||x-y|$ and the Cauchy-Schwarz inequality).

17.15 * (Another *Dominated Convergence Theorem.*) Let $(X_n)_{n\geq 1}$ be random variables with $X_n \xrightarrow{P} X$ ($\lim_{n\to\infty} X_n = X$ in probability). Suppose $|X_n(\omega)| \leq C$ for a constant $C > 0$ and all ω. Show that $\lim_{n\to\infty} E\{|X_n - X|\} = 0$. (*Hint:* First show that $P(|X| \leq C) = 1$.)