2.8. Modes of convergence

Definition 2.23. We say that $X_n \xrightarrow{P} X$ ("X_n converges to X in probability") if for any $\delta > 0$, $\mathbf{P}(|X_n - X| > \delta) \to 0$ as $n \to \infty$. Recall that we say that $X_n \xrightarrow{a.s.} X$ if $\mathbf{P}(\omega : \lim X_n(\omega) = X(\omega)) = 1$.

2.8.1. Almost sure and in probability.

Are they really different? Usually looking at Bernoulli random variables elucidates the matter.

Example 2.24. Suppose A_n are events in a probability space. Then one can see that

$$\text{(a) } \mathbf{1}_{A_n} \xrightarrow{P} 0 \iff \lim_{n\to\infty} \mathbf{P}(A_n) = 0. \qquad \text{(b) } \mathbf{1}_{A_n} \xrightarrow{a.s.} 0 \iff \mathbf{P}(\limsup_{n\to\infty} A_n) = 0.$$

By Fatou's lemma, $\mathbf{P}(\limsup_{n\to\infty} A_n) \ge \limsup \mathbf{P}(A_n)$, and hence we see that a.s convergence of $\mathbf{1}_{A_n}$ to zero implies convergence in probability. The converse is clearly false. For instance, if A_n are independent events with $\mathbf{P}(A_n) = n^{-1}$, then by the second Borel-Cantelli, $\mathbf{P}(A_n)$ goes to zero but $\mathbf{P}(\limsup A_n) = 1$. This example has all the ingredients for the following two implications.

Lemma 2.25. *Suppose X_n, X are r.v. on the same probability space. Then,*

(1) *If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{P} X$.*
(2) *If $X_n \xrightarrow{P} X$ "fast enough" so that $\sum_n \mathbf{P}(|X_n - X| > \delta) < \infty$ for every $\delta > 0$, then $X_n \xrightarrow{a.s.} X$.*

PROOF. Note that analogous to the example,

$$\text{(a) } X_n \xrightarrow{P} X \iff \forall \delta > 0,\ \lim_{n\to\infty} \mathbf{P}(|X_n - X| > \delta) = 0.$$
$$\text{(b) } X_n \xrightarrow{a.s.} X \iff \forall \delta > 0,\ \mathbf{P}(\limsup_{n\to\infty} |X_n - X| > \delta) = 0.$$

Thus, applying Fatou's we see that a.s convergence implies convergence in probability. By the first Borel Cantelli lemma, if $\sum_n \mathbf{P}(|X_n - X| > \delta) < \infty$, then $\mathbf{P}(|X_n - X| > \delta \text{ i.o}) = 0$ and hence $\limsup |X_n - X| < \delta$. Apply this to all rational δ to get $\limsup |X_n - X| = 0$ and thus we get a.s. convergence. ∎

Exercise 2.26. (1) If $X_n \xrightarrow{P} X$, show that $X_{n_k} \xrightarrow{a.s.} X$ for some subsequence.
(2) Show that $X_n \xrightarrow{a.s.} X$ if and only if every subsequence of $\{X_n\}$ has a further subsequence that converges a.s.
(3) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ (all r.v.s on the same probability space), show that $aX_n + bY_n \xrightarrow{P} aX + bY$ and $X_n Y_n \xrightarrow{P} XY$.

2.8.2. In distribution and in probability.

We say that $X_n \xrightarrow{d} X$ if the distributions of X_n converges to the distribution of X. This is a matter of language, but note that X_n and X need not be on the same probability space for this to make sense. In comparing it to convergence in probability, however, we must take them to be defined on a common probability space.

Lemma 2.27. *Suppose X_n, X are r.v. on the same probability space. Then,*

(1) *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.*
(2) *If $X_n \xrightarrow{d} X$ and X is a constant a.s, then $X_n \xrightarrow{P} X$.*

PROOF. (1) Suppose $X_n \xrightarrow{P} X$. Since for any $\delta > 0$

$\mathbf{P}(X_n \le t) \le \mathbf{P}(X \le t+\delta) + \mathbf{P}(X - X_n > \delta)$, and $\mathbf{P}(X \le t-\delta) \le \mathbf{P}(X_n \le t) + \mathbf{P}(X_n - X > \delta)$,

we see that $\limsup \mathbf{P}(X_n \le t) \le \mathbf{P}(X \le t+\delta)$ and $\liminf \mathbf{P}(X_n \le t) \ge \mathbf{P}(X \le t-\delta)$ for any $\delta > 0$. Taking $\delta \downarrow 0$ and letting t be a continuity point of the cdf of$X$, we immediately get $\lim \mathbf{P}(X_n \le t) = \mathbf{P}(X \le t)$. Thus, $X_n \xrightarrow{d} X$.

(2) If $X = a$ a.s (a is a constant), then the cdf of X is $F_X(t) = \mathbf{1}_{t \ge a}$. Hence, $\mathbf{P}(X_n \le t-\delta) \to 0$ and $\mathbf{P}(X_n \le t+\delta) \to 1$ for any $\delta > 0$ as $t \pm \delta$ are continuity points of F_X. Therefore $\mathbf{P}(|X_n - a| > \delta) \to 0$ and we see that $X_n \xrightarrow{P} a$. ■

Exercise 2.28. (1) Give an example to show that convergence in distribution does not imply convergence in probability.

(2) Suppose that X_n is independent of Y_n for each n (no assumptions about independence across n). If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, then $(X_n, Y_n) \xrightarrow{d} (U, V)$ where $U \overset{d}{=} X$, $V \overset{d}{=} Y$ and U, V are independent. Further, $aX_n + bY_n \xrightarrow{d} aU + bV$.

(3) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{d} Y$ (all on the same probability space), then show that $X_n Y_n \xrightarrow{d} XY$.

2.8.3. In probability and in L^p. How do convergence in L^p and convergence in probability compare? Suppose $X_n \xrightarrow{L^p} X$ (actually we don't need $p \ge 1$ here, but only $p > 0$ and $\mathbf{E}[|X_n - X|^p] \to 0$). Then, for any $\delta > 0$,

$$\mathbf{P}(|X_n - X| > \delta) \le \delta^{-p} \mathbf{E}[|X_n - X|^p] \to 0$$

and thus $X_n \xrightarrow{P} X$. The converse is not true as the following example shows.

Example 2.29. Let $X_n = 2^n$ w.p $1/n$ and $X_n = 0$ w.p $1 - 1/n$. Then, $X_n \xrightarrow{P} 0$ but $\mathbf{E}[X_n^p] = n^{-1} 2^{np}$ for all n, and hence X_n does not go to zero in L^p (for any $p > 0$).

As always, the fruitful question is to ask for additional conditions to convergence in probability that would ensure convergence in L^p. Let us stick to $p = 1$. Is there a reason to expect a (weaker) converse? Indeed, suppose $X_n \xrightarrow{P} X$. Then write $\mathbf{E}[|X_n - X|] = \int_0^\infty \mathbf{P}(|X_n - X| > t)dt$. For each t the integrand goes to zero. Will the integral go to zero? Surely, if $|X_n| \le 10$ a.s. for all n, (then the same holds for $|X|$) the integral reduces to the interval $[0, 20]$ and then by DCT (since the integrand is bounded by 1 which is integrable over the interval [0,20]), we get $\mathbf{E}[|X_n - X|] \to 0$.

As example 2.29 shows, the converse is not true in full generality either. What goes wrong in this example is that with a small probability X_n can take a very very large value and hence the expected value stays away from zero. This observation makes the next definition more palatable. We put the new concept in a separate section to give it the due respect that it deserves.

2.9. Uniform integrability

Definition 2.30. A family $\{X_i\}_{i \in I}$ of random variables is said to be *uniformly integrable* if given any $\epsilon > 0$, there exists A large enough so that $\mathbf{E}[|X_i| \mathbf{1}_{|X_i| > A}] < \epsilon$ for all $i \in I$.

Example 2.31. A finite set of integrable r.v.s is uniformly integrable. More interestingly, an L^p-bounded family with $p > 1$ is u.i. For, if $\mathbf{E}[|X_i|^p] \le M$ for all $i \in I$ for some $M > 0$, then $\mathbf{E}[|X_i|\mathbf{1}_{|X_i|>t}] \le t^{-(p-1)}M$ which goes to zero as $t \to \infty$. Thus, given $\epsilon > 0$, one can choose t large so that $\sup_{i\in I}\mathbf{E}[|X_i|\mathbf{1}_{|X_i|>t}] < \epsilon$.

This fails for $p = 1$ as the example 2.29 shows a family of L^1 bounded random variables that are not u.i. However, a u.i family must be bounded in L^1. To see this find $A > 0$ so that $\mathbf{E}[|X_i|\mathbf{1}_{|X_i|>A}] < 1$ for all i. Then, for any $i \in I$, we get $\mathbf{E}[|X_i|] = \mathbf{E}[|X_i|\mathbf{1}_{|X_i|<A}] + \mathbf{E}[|X_i|\mathbf{1}_{|X_i|\ge A}] \le A + 1$.

Exercise 2.32. If $\{X_i\}_{i\in I}$ and $\{Y_j\}_{j\in J}$ are both u.i, then $\{X_i + Y_j\}_{(i,j)\in I\times J}$ is u.i. What about the family of products, $\{X_iY_j\}_{(i,j)\in I\times J}$?

Lemma 2.33. *Suppose X_n, X are r.v. on the same probability space. Then, the following are equivalent.*

(1) $X_n \stackrel{L^1}{\to} X$.
(2) $X_n \stackrel{P}{\to} X$ *and* $\{X_n\}$ *is u.i.*

PROOF. If $Y_n = X_n - X$, then $X_n \stackrel{L^1}{\to} X$ iff $Y_n \stackrel{L^1}{\to} 0$, while $X_n \stackrel{P}{\to} X$ iff $Y_n \stackrel{P}{\to} 0$ and by the first part of exercise 2.32, $\{X_n\}$ is u.i if and only if $\{Y_n\}$ is. Hence we may work with Y_n instead (i.e., we may assume that the limiting r.v. is 0 a.s).

First suppose $Y_n \stackrel{L^1}{\to} 0$. Then we showed that $Y_n \stackrel{P}{\to} 0$. To show that $\{Y_n\}$ is u.i, let $\epsilon > 0$ and fix N_ϵ so that $\mathbf{E}[|Y_n|] < \epsilon$ for all $n \ge N_\epsilon$. Then, pick $A > 1$ so large that $\mathbf{E}[|Y_k|\mathbf{1}_{|Y_k|>A}] \le \epsilon$ for all $k \le N$. With the same A and any $k \ge N_\epsilon$, we get $\mathbf{E}[|Y_k|\mathbf{1}_{|Y_k|>A}] \le A^{-1}\mathbf{E}[|Y_k|] < \epsilon$ since $A > 1$ and $\mathbf{E}[|Y_k|] < \epsilon$. Thus we have found one A which works for all Y_k. Hence $\{Y_k\}$ is u.i.

Next suppose $Y_n \stackrel{P}{\to} 0$ and that $\{Y_n\}$ is u.i. Then, fix $\epsilon > 0$ and find $A > 0$ so that $\mathbf{E}[|Y_k|\mathbf{1}_{|Y_k|>A}] \le \epsilon$ for all k. Then,

$$\mathbf{E}[|Y_k|] \le \mathbf{E}[|Y_k|\mathbf{1}_{|Y_k|\le A}] + \mathbf{E}[|Y_k|\mathbf{1}_{|Y_k|>A}] \le \int_0^A \mathbf{P}(|Y_k| > t)dt + \epsilon.$$

For all $t \in [0, A]$, by assumption $\mathbf{P}(|Y_k| > t) \to 0$, while we also have $\mathbf{P}(|Y_k| > t) \le 1$ for all k and 1 is integrable on $[0, A]$. Hence, by DCT the first term goes to 0 as $k \to \infty$. Thus $\limsup \mathbf{E}[|Y_k|] \le \epsilon$ for any ϵ and it follows that $Y_k \stackrel{L^1}{\to} 0$. ■

Corollary 2.34. *If $X_n \stackrel{a.s.}{\to} X$, then $X_n \stackrel{L^1}{\to} X$ if and only if $\{X_n\}$ is u.i.*

To deduce convergence in mean from a.s convergence, we have so far always invoked DCT. As shown by Lemma 2.33 and corollary 2.34, uniform integrability is the sharp condition, so it must be weaker than the assumption in DCT. Indeed, if $\{X_n\}$ are dominated by an integrable Y, then whatever A works for Y in the u.i condition will work for the whole family $\{X_n\}$. Thus a dominated family is u.i., while the converse is false.

Remark 2.35. Like tightness of measures, uniform integrability is also related to a compactness question. On the space $L^1(\mu)$, apart from the usual topology coming from the norm, there is another one called *weak topology* (where $f_n \to f$ if and only if $\int f_n g d\mu \to \int f g d\mu$ for all $g \in L^\infty(\mu)$). The *Dunford-Pettis theorem* asserts that pre-compact subsets of $L^1(\mu)$ in this weak topology are precisely uniformly integrable subsets of $L^1(\mu)$! A similar question can be asked in L^p for $p > 1$ where weak topology means that $f_n \to f$ if and only if $\int f_n g d\mu \to \int f g d\mu$ for all $g \in L^q(\mu)$ where $q^{-1} + p^{-1} = 1$. Another part of Dunford-Pettis theorem asserts that pre-compact subsets of $L^p(\mu)$ in this weak topology are precisely those that are bounded in the $L^p(\mu)$ norm.