Advanced Probability: WEEK 4

1 Convergence of sequences of random variables

1.1 Preliminary: convergence of sequences of numbers

Let us recall that a sequence of real numbers \((a_n, n \geq 1)\) converges to a limit \(a \in \mathbb{R}\) (and this is denoted as \(a_n \to_{n \to \infty} a\)) if and only if
\[
\forall \varepsilon > 0, \exists N \geq 1 \text{ such that } \forall n \geq N, |a_n - a| \leq \varepsilon.
\]
Reciprocally, the sequence \((a_n, n \geq 1)\) does not converge to \(a \in \mathbb{R}\) if and only if
\[
\exists \varepsilon > 0 \text{ such that } \forall N \geq 1, \exists n \geq N \text{ such that } |a_n - a| > \varepsilon.
\]
This is still equivalent to saying that
\[
\exists \varepsilon > 0 \text{ such that } |a_n - a| > \varepsilon \text{ for an infinite number of values of } n.
\]
In the previous sentence, “for an infinite number of values of \(n\)” may be abbreviated as “infinitely often”.

1.2 Definitions

In order to extend the notion of convergence from sequences of numbers to sequences of random variables, there are quite a few possibilities. We have indeed seen in the previous lectures that random variables are functions. In a first year analysis course, one hears about various notions of convergence for sequences of functions, among which pointwise and uniform convergence. There are actually many others. We will see four of them that are most useful in the context of probability, and three of them in today’s lecture.

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

1) Quadratic convergence. Assume that all \(X_n, X\) are square-integrable. The sequence \((X_n, n \geq 1)\) is said to converge in quadratic mean to \(X\) (and this is denoted as \(X_n \overset{L^2}{\to}_{n \to \infty} X\)) if
\[
E(|X_n - X|^2) \to_{n \to \infty} 0.
\]

2) Convergence in probability. The sequence \((X_n, n \geq 1)\) is said to converge in probability to \(X\) (and this is denoted as \(X_n \overset{P}{\to}_{n \to \infty} X\)) if
\[
\forall \varepsilon > 0, \quad \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \to_{n \to \infty} 0.
\]

3) Almost sure convergence. The sequence \((X_n, n \geq 1)\) is said to converge almost surely to \(X\) (and this is denoted as \(X_n \to_{n \to \infty} X\) a.s.) if
\[
\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.
\]

\(^1\)Note that aside from exceptional cases, a set \(\Omega\) on which one can define an infinite sequence of random variables has to be quite large, in order to contain all possible outcomes or realizations of this infinite sequence.
1.3 Relations between these three notions of convergence

**Quadratic convergence implies convergence in probability.** This is a direct consequence of Chebychev’s inequality. Assume indeed that \( X_n \xrightarrow{L^2}{\infty} X \). Then we have, for any \( \varepsilon > 0 \),

\[
P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \leq \frac{E(|X_n - X|^2)}{\varepsilon^2} \xrightarrow{n \to \infty} 0 \text{ by assumption, so } X_n \xrightarrow{P}{\infty} X.
\]

**Convergence in probability does not imply quadratic convergence.** This is left for homework. Note that counter-examples exist even in the case where all \( X_n, X \) are square-integrable.

**Almost sure convergence implies convergence in probability.** In order to show this, we will need two facts that can be easily deduced from the basic axioms on probability measures, as well as a lemma that provides an alternate characterization of almost sure convergence.

**Fact 1.** \( P(A_M) = 0 \) for all \( M \geq 1 \) if and only if \( P(\bigcup_{M \geq 1} A_M) = 0 \).

**Fact 2.** \( P(\bigcap_{N \geq 1} A_N) = \lim_{N \to \infty} P(A_N) \) if \( A_N \supset A_{N+1} \) for all \( N \geq 1 \).

**Lemma 1.1.** (characterization of almost sure convergence)

\( X_n \xrightarrow{n \to \infty} X \text{ a.s. if and only if } \forall \varepsilon > 0, P(\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \text{ infinitely often} \}) = 0 \).

**Proof.** We drop here the full notation with \( \omega \)'s and use the abbreviation “i.o.” for “infinitely often” in order to lighten the writing. Based on what was said on the coverage of sequences of numbers, we obtain the following series of equivalences:

\[
X_n \xrightarrow{n \to \infty} X \text{ a.s. } \iff P(\left\{ \lim_{n \to \infty} X_n = X \right\}) = 1 \iff P(\left\{ \lim_{n \to \infty} X_n \neq X \right\}) = 0
\]

\[
\iff P(\{ \exists \varepsilon > 0 \text{ such that } |X_n - X| > \varepsilon \text{ i.o.} \}) = 0
\]

\[
\iff P(\left\{ \exists M \geq 1 \text{ such that } |X_n - X| > \frac{1}{M} \text{ i.o.} \right\}) = 0
\]

\[
\iff P\left( \bigcup_{M \geq 1} \left\{ |X_n - X| > \frac{1}{M} \text{ i.o.} \right\} \right) = 0.
\]

By Fact 1, this last assertion is equivalent to saying that

\[
\forall M \geq 1, P(\left\{ |X_n - X| > \frac{1}{M} \text{ i.o.} \right\}) = 0,
\]

which is in turn equivalent to saying that

\[
\forall \varepsilon > 0, P(\{ |X_n - X| > \varepsilon \text{ i.o.} \}) = 0,
\]

and completes the proof. Note the “hat trick” used here in order to replace an uncountable union over \( \varepsilon \)'s by a countable union over \( M \)'s. \( \square \)
Proof that almost sure convergence implies convergence in probability. We have the following series of equivalences. By Lemma 1.1,

\[ X_n \xrightarrow{\text{a.s.}} X \quad \iff \quad \forall \varepsilon > 0, \quad \mathbb{P} \left( \{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \text{ infinitely often} \} \right) = 0 \]

\[ \iff \quad \forall \varepsilon > 0, \quad \mathbb{P} \left( \{ \forall N \geq 1, \exists n \geq N \text{ such that } |X_n - X| > \varepsilon \} \right) = 0 \]

\[ \iff \quad \forall \varepsilon > 0, \quad \mathbb{P} \left( \bigcap_{N \geq 1} \bigcup_{n \geq N} \{ |X_n - X| > \varepsilon \} \right) = 0. \]

By Fact 2, this is equivalent to saying that \( \forall \varepsilon > 0, \lim_{N \to \infty} \mathbb{P} \left( \bigcup_{n \geq N} \{ |X_n - X| > \varepsilon \} \right) = 0 \), and this last assertion implies that \( \forall \varepsilon > 0, \lim_{N \to \infty} \mathbb{P} \left( \{ |X_N - X| > \varepsilon \} \right) = 0 \). Said otherwise: \( X_N \xrightarrow{P} X \). \( \square \)

Convergence in probability does not imply almost sure convergence, as surprising as this may sound! Here is a counter-example: let us consider a sequence of independent and identically distributed (i.i.d.) heads (H) and tails (T). Out of this sequence, we construct another sequence of random variables:

\[ X_1 = H_1, \quad X_2 = T_2, \quad X_3 = H_1 H_2, \quad X_4 = H_1 T_2, \quad X_5 = T_2 H_1, \quad X_6 = T_2 T_1, \quad X_7 = H_1 H_2 H_1, \quad X_8 = H_1 H_2 T_1 \ldots \]

meaning “\( X_1 = 1 \) iff the first coin falls on heads”, “\( X_2 = 1 \) iff the first coin falls on tails”, “\( X_3 = 1 \) iff the first two coins fall on heads”, etc. Note that this new sequence of random variables is everything but independent: there is indeed a strong dependency e.g. between \( X_3, X_4, X_5 \) and \( X_6 \), as only one of these random variables can take the value 1 for a given sequence of heads and tails.

On the one hand, \( X_n \xrightarrow{P} 0 \). Indeed, one can check that for any \( \varepsilon > 0 \), the probability

\[ \mathbb{P} \left( \{ |X_n - 0| > \varepsilon \} \right) = O \left( \frac{1}{n} \right). \]

It therefore converges to 0 as \( n \to \infty \).

On the other hand, \( X_n \not\xrightarrow{\text{a.s.}} 0 \). Indeed, for a given realization \( \omega \) of heads and tails, such as e.g. HHTHTT..., the sequence \( (X_n(\omega), n \geq 1) \) is equal to 10100001000000... Note that as you explore further the sequence, you encounter less and less 1’s; nevertheless, you always encounter a 1 after a sufficiently large number of steps. So the sequence \( (X_n(\omega), n \geq 1) \) is an alternating sequence of 0’s and 1’s, that therefore does not converge to 0 (according the the first definition of today’s lecture), and this is true for any realization \( \omega \).

In conclusion,

\[ \mathbb{P} \left( \{ \omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = 0 \} \right) = 0, \]

which is the complete opposite of the definition of almost sure convergence.

Remark 1.2. As we know that \( X_n \xrightarrow{P} 0 \), the sequence cannot converge to anything else almost surely. Indeed, as almost sure convergence implies convergence in probability, if \( X_n \) were to converge a.s. to another limit \( X \neq 0 \), this would imply that \( X_n \) should also converge in probability towards this same limit \( X \). But we know already that \( X_n \) converges in probability to 0, so \( X \) cannot be different from 0.

1.4 The Borel-Cantelli lemma

As one might guess from the previous pages, proving convergence in probability (using e.g. Chebychev’s inequality) is in general much easier than proving almost sure convergence. Still, these two notions are not equivalent, as the previous counter-example shows. So it would be convenient to have a criterion saying that if both convergence in probability and another easy-to-check condition hold, then almost sure convergence holds. This criterion is Borel-Cantelli’s lemma.

Reminder. Let \((a_n, n \geq 1)\) be a sequence of positive numbers. Then writing that \( \sum_{n \geq 1} a_n < \infty \) exactly means that

\[ \lim_{N \to \infty} \sum_{n \geq N} a_n = 0. \]
which is stronger than \(\lim_{n\to\infty} a_n = 0\). This last condition alone does indeed not guarantee that \(\sum_{n\geq 1} a_n < \infty\). A famous counter-example is the harmonic series \(a_n = \frac{1}{n}\).

**Lemma 1.3.** (Borel-Cantelli)

Let \((A_n, n \geq 1)\) be a sequence of events in \(\mathcal{F}\) such that \(\sum_{n\geq 1} \mathbb{P}(A_n) < \infty\). Then

\[
\mathbb{P}(\{\omega \in \Omega : \omega \in A_n \text{ infinitely often}\}) = 0.
\]

Before proving this lemma, let us see how it can be applied to the convergence of sequences of random variables. Let \((X_n, n \geq 1)\) be a sequence of random variables.

a) If for all \(\varepsilon > 0\), \(\mathbb{P}(\{|X_n - X| > \varepsilon\}) \to 0\) as \(n \to \infty\), then \(X_n \xrightarrow{p} X\), by definition.

b) If for all \(\varepsilon > 0\), \(\sum_{n\geq 1} \mathbb{P}(\{|X_n - X| > \varepsilon\}) < \infty\), then \(X_n \xrightarrow{a.s.} X\). Indeed, by the Borel-Cantelli lemma, the condition on the sum implies that for all \(\varepsilon > 0\),

\[
\mathbb{P}(\{|X_n - X| > \varepsilon \text{ infinitely often}\}) = 0,
\]

which is exactly the characterization of almost sure convergence given in Lemma 1.1.

So we see that if one can prove that \(\mathbb{P}(\{|X_n - X| > \varepsilon\}) = O\left(\frac{1}{n}\right)\) for all \(\varepsilon > 0\), this guarantees convergence in probability, but not almost sure convergence, as the condition on the sum is not necessarily satisfied (cf. the example in the previous section).

**Proof of the Borel-Cantelli lemma.** Let us first rewrite

\[
\mathbb{P}(\{\omega \in \Omega : \omega \in A_n \text{ infinitely often}\}) = \mathbb{P}(\{\forall N \geq 1, \exists n \geq N \text{ such that } \omega \in A_n\})
\]

\[
= \mathbb{P}\left(\bigcap_{N \geq 1} \bigcup_{n \geq N} A_n\right) = \lim_{N \to \infty} \mathbb{P}\left(\bigcup_{n \geq N} A_n\right),
\]

by Fact 2. This implies that

\[
\mathbb{P}(\{\omega \in \Omega : \omega \in A_n \text{ infinitely often}\}) \leq \lim_{N \to \infty} \sum_{n \geq N} \mathbb{P}(A_n) = 0,
\]

by the assumption made on the sum. This proves the Borel-Cantelli lemma. \(\square\)