Math 525: Lecture 9

February 1, 2018

1 Convergence of random variables

In a previous lecture, we saw that (roughly speaking) $B(n, \frac{\lambda}{n}) \to \text{Poisson}(\lambda)$ as $n \to \infty$. But what does "$\to$" mean here? More generally, consider a sequence of random variables $X_1, X_2, \ldots$ What does it mean for this sequence to converge? Are there multiple notions of convergence available? Which notions of convergence are useful? This will be the topic of today's lecture.

Definition 1.1. The sequence $(X_n)_n$ of random variables *converges pointwise* to a random variable X if $X_n(\omega) \to X(\omega)$ for each ω in the sample space.

This is the usual definition of pointwise convergence for functions, but it's not the most useful in probability. Why? Consider two random variables X and Y which are equal to a.s., but we can find ω such that $X(\omega) \neq Y(\omega)$. Then, the alternating sequence $X, Y, X, Y, \ldots$ does not have a pointwise limit, even though the two random variables are "essentially" the same!

The above discussion implies that we need a weaker notion of convergence when it comes to probability:

Definition 1.2. Let $(X_n)_n$ be a sequence of random variables and X be a random variable.

1. $(X_n)_n$ converges *in probability* to X if for all $\epsilon > 0$,
$$\mathbb{P}\left\{|X_n - X| > \epsilon\right\} \to 0 \text{ as } n \to \infty.$$
2. $(X_n)_n$ converges to X *with probability one* or *almost everywhere* (a.e.) if $X_n(\omega) \to X(\omega)$ for all $\omega \notin \Lambda$ where $\mathbb{P}(\Lambda) = 0$.
3. $(X_n)_n$ converges to X in $\mathbb{L}^p$ if X^p is integrable and
$$\mathbb{E}\left[|X_n - X|^p\right] \to 0 \text{ as } n \to \infty.$$
We write $X_n \xrightarrow{\mathbb{L}^p} X$ in this case. When $p = 1$, we call this "convergence in mean".
4. Let F_n and F denote the distribution functions of X_n and X, respectively. $(X_n)_n$ converges to X in distribution if $F_n(x) \to F(x)$ for all continuity points of F. We write $X_n \xrightarrow{\mathcal{D}} X$ in this case.

Some notions of convergence are stronger than others:

Proposition 1.3. *If $X_n \xrightarrow{\mathbb{L}^p} X$, then $X_n \to X$ in probability.*

Proof. This is a consequence of Chebyshev's inequality:

$$\mathbb{P}\left\{|X_n - X| > \epsilon\right\} \leq \frac{1}{\epsilon^p}\mathbb{E}\left[|X_n - X|^p\right] \to 0. \qquad \square$$

Proposition 1.4. *If $X_n \to X$ a.e., then $X_n \to X$ in probability.*

Proof. Suppose $X_n \to X$ pointwise for all $\omega \notin \Lambda$ where $\mathbb{P}(\Lambda) = 1$. Let

$$Z_n = \sup_{k \geq n} |X_k - X|$$

and note that $\lim_n Z_n = \limsup_n |X_n - X|$. Therefore,

$$X_n(\omega) \to X(\omega) \iff Z_n(\omega) \to 0.$$

Let $\epsilon > 0$ and

$$\Gamma_n^\epsilon = \left\{Z_n \geq \epsilon\right\}.$$

If $\omega \in \cap_n \Gamma_n^\epsilon$, then $Z(\omega) \nrightarrow 0$, and hence $\cap_n \Gamma_n^\epsilon \subset \Lambda$. Moreover, note that these sets are decreasing in containment:

$$\Gamma_1^\epsilon \supset \Gamma_2^\epsilon \supset \cdots$$

Therefore, $\mathbb{P}(\Gamma_n^\epsilon) \to \mathbb{P}(\cap_n \Gamma_n^\epsilon) \leq \mathbb{P}(\Lambda) = 0$. Since $|X_n - X| \leq Z_n$,

$$\mathbb{P}\left\{|X_n - X| \geq \epsilon\right\} \leq \mathbb{P}(\Gamma_n^\epsilon) \to 0. \qquad \square$$

The converse of the above is not true in general:

Example 1.5. Let $Y \sim U[0,1]$ and define

$$\begin{aligned}
X_1 &= 1 \\
X_2 &= I_{[0,1/2]}(Y) \\
X_3 &= I_{(1/2,1]}(Y) \\
X_4 &= I_{[0,1/4]}(Y) \\
X_5 &= I_{(1/4,1/2]}(Y) \\
&\vdots
\end{aligned}$$

Note that $X_n \to 0$ in probability. However, there is no ω for which $X_n(\omega) \to 0$!

It is not trivial that a limit of a sequence of random variables is a random variable itself, so we prove that next. As a technical point, since limits may introduce values of $-\infty$ and $+\infty$, we need to work with random variables which can take on infinite values:

Definition 1.6. Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ denote the *extended real line*. An *extended real valued (ERV) random variable* is a function $X\colon \Omega \to \overline{\mathbb{R}}$ such that

$$\{X \leq x\} \in \mathcal{F} \qquad \text{for all } x \in \overline{\mathbb{R}}.$$

Proposition 1.7. *Let $X_1, X_2, \ldots$ be ERV random variables. Then, M, m, and X_∞ are also ERV random variables where*

1. $M(\omega) = \sup X_n(\omega)$.
2. $m(\omega) = \limsup_{n\to\infty} X_n(\omega)$.
3. $X_\infty(\omega) = \begin{cases} \lim_n X_n(\omega) & \textit{if the limit exists} \\ 0 & \textit{otherwise.} \end{cases}$

Note that by taking the negation of the first two, we find that $\inf X_n$ and $\liminf_{n\to\infty} X_n$ are also ERV random variables.

Proof.

1. We need to show that $\{\omega\colon M(\omega) \le x\} = \cap_n \{\omega\colon X_n(\omega) \le x\}$ is in $\mathcal{F}$ for any $x \in \overline{\mathbb{R}}$. Since it is a countable intersection of sets in $\mathcal{F}$, the desired result follows.
2. Note that $m = \inf_n Y_n$ where $Y_n = \sup_{k\ge n} X_k$. We know by the previous point that Y_n is an ERV random variable for each n. Therefore, $\sup_n -Y_n = -\inf Y_n$ is an ERV random variable, and so too is m.
3. Let
$$\Lambda_\infty = \left\{\omega\colon \limsup_n X_n(\omega) = \liminf_n X_n(\omega)\right\}.$$
This set is in $\mathcal{F}$, and hence we can define the random variable
$$Y_n = I_{\Lambda_\infty} X_n.$$
The desired result follows because
$$\limsup_n Y_n = \lim_n X_n$$
is an extended real-valued random variable. □

2 Borel-Cantelli lemma

Definition 2.1. Let $(\Lambda_n)_n$ be a sequence of subsets of Ω. Define
$$\limsup_n \Lambda_n = \{\omega \mid \forall N\colon \exists n \ge N\colon \omega \in \Lambda_n\}$$
$$\liminf_n \Lambda_n = \{\omega \mid \exists N\colon \forall n \ge N\colon \omega \in \Lambda_n\}.$$

Intuitively, we can think of $\limsup_n \Lambda_n$ as the set of all ω such that $\omega \in \Lambda_n$ for "infinitely many n". Similarly, $\liminf_n \Lambda_n$ is the set of all ω such that $\omega \in \Lambda_n$ for "all but finitely many n". Expressed in an equivalent way,
$$\limsup_n \Lambda_n = \bigcap_N \bigcup_{n\ge N} \Lambda_m$$
$$\liminf_n \Lambda_n = \bigcup_N \bigcap_{n\ge N} \Lambda_m.$$

Applying De Morgan's law, we see

$$(\limsup \Lambda_n)^c = \left(\bigcap_N \bigcup_{n \geq N} \Lambda_m \right)^c = \bigcup_N \bigcap_{n \geq N} \Lambda_m^c = \liminf \Lambda_n^c$$

and hence $\limsup \Lambda_n^c = (\liminf \Lambda_n)^c$ also.

Proposition 2.2 (Borel-Cantelli lemma)**.** *Let $(\Lambda_n)_n$ be a sequence in $\mathcal{F}$. Suppose $\sum_n \mathbb{P}(\Lambda_n) < \infty$. Then,*

$$\mathbb{P}\left\{\limsup \Lambda_n\right\} = 0.$$

Proof. Recall that

$$\limsup_n \Lambda_n = \bigcap_N A_N \qquad \text{where} \qquad A_N = \bigcup_{n \geq N} \Lambda_n.$$

In particular, $(A_N)_N$ is decreasing in containment: $A_1 \supset A_2 \supset \cdots$ Therefore,

$$\lim_{N \to \infty} \mathbb{P}(A_N) = \mathbb{P}\left(\bigcap_N A_N \right) = \mathbb{P}\left(\limsup_n \Lambda_n \right) \qquad \text{as} \qquad N \to \infty.$$

But note that

$$\lim_{N \to \infty} \mathbb{P}(A_N) = \lim_{N \to \infty} \mathbb{P}\left(\bigcup_{n \geq N} \Lambda_n \right) \leq \lim_{N \to \infty} \sum_{n \geq N} \mathbb{P}(\Lambda_n) \to 0.$$

□

Next, we look at an important application of the Borel-Cantelli lemma:

Proposition 2.3. *Let $(X_n)_n$ be a sequence of ERV random variables and suppose $X_n \to X$ in probability. Then, there exists a subsequence $(n_k)_k$ such that $X_{n_k} \to X$ a.e.*

Proof. Taking $\epsilon = 1/m$ in the definition of convergence in probability, we find that for each m,

$$\mathbb{P}\left\{|X_n - X| > 1/m\right\} \to 0 \qquad \text{as } n \to \infty.$$

Choose an increasing sequence $(n_m)_m$ such that

$$\mathbb{P}(\Lambda_m) < 1/2^m.$$

where

$$\Lambda_m = \left\{ |X_{n_m} - X| > \frac{1}{m} \right\}.$$

Then, $\sum_m \mathbb{P}(\Lambda_m) < \infty$, and therefore $\mathbb{P}(\limsup_m \Lambda_m) = 0$ by the Borel-Cantelli lemma. Therefore,

$$\mathbb{P}\left(\liminf_m \Lambda_m^c \right) = 1 - \mathbb{P}\left(\limsup_m \Lambda_m \right) = 1.$$

Recall now that

$$\liminf_m \Lambda_m^c = \left\{ \omega \mid \exists N \colon \forall m \geq N \colon \omega \notin \Lambda_n \right\}.$$

Therefore, for each $\omega \in \liminf_m \Lambda_m$, we can find an $N(\omega)$ (depending on ω) such that for all $m \geq N(\omega)$, we have $\omega \notin \Lambda_m$. It follows that for $\omega \in \liminf_m \Lambda_m$,

$$|X_{n_m}(\omega) - X(\omega)| \leq \frac{1}{m} \qquad \text{for all } m \geq N(\omega)$$

and hence

$$|X_{n_m}(\omega) - X(\omega)| \to 0.$$

That is, $X_{n_m} \to X$ pointwise on the set $\liminf_m \Lambda_m$, which is exactly the definition of a.e. convergence. $\square$

3 Modes of convergence diagram

The following diagram summarizes convergence relationships in a probability space. AE refers to a.e. convergence, L^p refers to what we've been calling $\mathbb{L}^p$ convergence, and M refers to convergence in probability. AU is a.u. convergence which we will most likely not encounter, though it's definition is below for those who are interested.

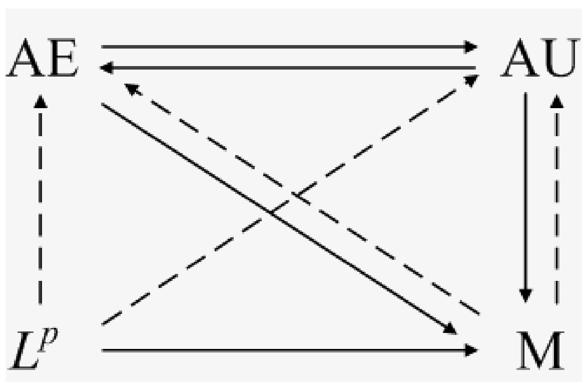

Figure 1: Modes of convergence (diagram by John Cook). A solid line from one mode of convergence to another indicates implication (e.g., a.e. convergence implies convergence in probability). A dashed line means that we can extract a subsequence, as in Proposition 2.3.

3.1 Almost uniform convergence (optional)

Let $(f_n)_n$ be a sequence of real-valued functions and f be a real-valued function such that f_n and f have the same domain. We say $f_n \to f$ *uniformly* if for each $\epsilon > 0$, we can find N such that for all $n \geq N$,

$$\|f_n - f\|_\infty \equiv \sup_x |f_n(x) - f(x)| < \epsilon.$$

Definition 3.1. Let $(X_n)_n$ be a sequence of random variables and X be a random variable. We say $X_n \to X$ almost uniformly (a.u.) if for every $\epsilon > 0$, we can find a set $A \in \mathcal{F}$ with $\mathbb{P}(A) < \epsilon$ such that $X_n|_{A^c} \to X|_{A^c}$ converges uniformly. Here, the symbol $\cdot|_{A^c}$ is the restriction of a function to the set A^c.