Math 525: Lecture 13

February 20, 2018

For the next few lectures, we will work towards proving the central limit theorem (CLT). Recall that for a "sufficiently nice" sequence of i.i.d. random variables $(X_n)_n$, the law of large numbers told us

$$\frac{S_n}{n} \equiv \frac{X_1 + \cdots + X_n}{n} \to \mathbb{E}X_1 \text{ a.s.}$$

The CLT will tell us about the distribution of S_n/n. Namely,

$$\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}X_1\right) \xrightarrow{\mathcal{D}} Y$$

where $Y \sim \mathcal{N}(0, \mathrm{Var}(X_1))$. To prove the CLT, we will need to build up a series of results culminating in *Levy's continuity theorem.*

1 Convergence in distribution

We begin with an example:

Example 1.1. Let $X = 0$ be a "deterministic" random variable with distribution function $F = [0, \infty)$.

1. Let $X_n = -\frac{1}{n}$. Then, its distribution function is $F_n = I_{[1/n,\infty)}$. Note, in particular, that

$$F_n(x) \to F(x)$$

at all points x.

2. Now consider $X_n = \frac{1}{n}$. Then, its distribution function is $F_n = I_{[\frac{1}{n},\infty)}$. Note, in particular, that

$$F_n(x) \to F(x)$$

for all x except $x = 0$ since $F_n(0) = 0$ for all n and $F(0) = 1$!

In the second example above, while we expect the distribution function of X_n to converge to that of X (after all, $X_n \to X$ pointwise), it does not! The problem here is the point of discontinuity of F at $x = 0$. This motivates the definition of convergence in distribution (which we actually saw in a previous lecture).

Definition 1.2. Let $(X_n)_n$ be a sequence of random variables and X be a random variable. 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 $X_n \xrightarrow{\mathcal{D}} X$ or $F_n \Rightarrow F$ in this case.

The next proposition tells us that convergence in probability is stronger than convergence in distribution (the converse is not true). This in turn implies that any kind of convergence that we have covered (a.s. or L^p) are stronger than convergence in distribution.

Proposition 1.3. *If $X_n \to X$ in probability, then $X_n \xrightarrow{\mathcal{D}} X$.*

Proof. Let $\epsilon > 0$. Trivially,

$$\begin{aligned} \mathbb{P}\{X_n \le x\} &= \mathbb{P}\{X_n \le x, X \le x+\epsilon\} + \mathbb{P}\{X_n \le x, X > x+\epsilon\} \\ &\le \mathbb{P}\{X \le x+\epsilon\} + \mathbb{P}\{|X_n - X| > \epsilon\}. \end{aligned}$$

In other words,

$$F_n(x) \le F(x+\epsilon) + \mathbb{P}\{|X_n - X| > \epsilon\}.$$

Similarly, we can show

$$F(x-\epsilon) \le F_n(x) + \mathbb{P}\{|X_n - X| > \epsilon\}.$$

Combining these inequalities,

$$F(x-\epsilon) - \mathbb{P}\{|X_n - X| > \epsilon\} \le F_n(x) \le F(x+\epsilon) + \mathbb{P}\{|X_n - X| > \epsilon\}.$$

Taking $n \to \infty$,

$$F(x-\epsilon) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x+\epsilon).$$

If x is a point of continuity, taking $\epsilon \downarrow 0$ yields $F(x-\epsilon) \to F(x)$ and $F(x+\epsilon) \to F(x)$, and the result is proved. $\square$

We now visit an alternate characterization of convergence in distribution. Recall that $f : \mathbb{R} \to \mathbb{R}$ is said to be bounded if we can find $M > 0$ such that $|f(x)| \le M$ for all $x \in \mathbb{R}$.

Proposition 1.4. *$X_n \xrightarrow{\mathcal{D}} X$ if and only if for all bounded functions $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]. \tag{1}$$

Proof ($\Rightarrow$). Suppose $X_n \xrightarrow{\mathcal{D}} X$. We start by establishing (1) for all continuous functions f with compact support (i.e., $f(x) = 0$ if $|x| \ge r$ for some $r > 0$). In this case, f is uniformly continuous.

Now, let $\epsilon > 0$. Consider a partition $\{(x_{i-1}, x_i]\}_{i=1}^n$ of $(-r, r]$ satisfying

$$|f(x) - f(x_i)| < \epsilon/2 \qquad \text{for all } x \in (x_{i-1}, x_i]$$

and chosen such that each x_i is a point of continuity of F, the distribution function of X. Let ψ be the function defined by

$$\psi(x) = \sum_{i=1}^{n} f(x_i) I_{(x_{i-1},x_i]}(x).$$

Note, in particular, that

$$|f(x) - \psi(x)| \leq \sum_{i=1}^{n} |f(x) - f(x_i)|\, I_{(x_{i-1},x_i]}(x) < \epsilon/2.$$

Recalling that any bounded random variable is trivially integrable, we have

$$|\mathbb{E}f(X) - \mathbb{E}f(X_n)| \leq |\mathbb{E}f(X) - \mathbb{E}\psi(X)| + |\mathbb{E}\psi(X) - \mathbb{E}\psi(X_n)| + |\mathbb{E}\psi(X_n) - \mathbb{E}f(X_n)|\,.$$

The first and last expectations are bounded by $\epsilon/2$. As for the middle expectation,

$$\begin{aligned}
|\mathbb{E}\psi(X) - \mathbb{E}\psi(X_n)| &= |\mathbb{E}\left[\psi(X) - \psi(X_n)\right]| \\
&= \left| \sum_{i=1}^{n} f(x_i) \mathbb{E}\left[I_{(x_{i-1},x_i]}(X) - I_{(x_{i-1},x_i]}(X_n) \right] \right| \\
&= \left| \sum_{i=1}^{n} f(x_i) \left[(F(x_i) - F(x_{i-1})) - (F_n(x_i) - F_n(x_{i-1})) \right] \right|
\end{aligned}$$

Now, since each x_i is a point of continuity of F, we have $F_n(x_i) \to F(x_i)$ for all i. Therefore,

$$\limsup_n |\mathbb{E}f(X) - \mathbb{E}f(X_n)| < \epsilon.$$

Now take $\epsilon \downarrow 0$ to get

$$\lim_n |\mathbb{E}f(X) - \mathbb{E}f(X_n)| = 0.$$

Now, let's see how to modify the proof if f is just bounded and continuous but not necessarily of compact support. Let $\epsilon > 0$ and choose K such that $\pm K$ are continuity points of F, $F(-K) < \epsilon/4$ and $F(K) > 1 - \epsilon/4$ (remember, F is a distribution function, so that $F(-x) \to 0$ and $F(x) \to 1$ as $x \to \infty$). Let g be a continuous function such that $g(x) = 1$ if $|x| \leq K$ and $g(x) = 0$ if $|x| \geq K + 1$. Then, fg is continuous, equals f on $[-K, K]$, and has compact support. Therefore,

$$\begin{aligned}
|\mathbb{E}f(X) - \mathbb{E}f(X_n)| &\leq \|f\|_\infty \left(\mathbb{P}\{X \leq -K\} + \mathbb{P}\{X_n \leq -K\}\right) \\
&\quad + \|f\|_\infty \left(\mathbb{P}\{X > K\} + \mathbb{P}\{X_n > K\}\right) + |\mathbb{E}\left[f(X)g(X)\right] - \mathbb{E}\left[f(X_n)g(X_n)\right]|\,. \qquad (2)
\end{aligned}$$

Note that

$$\mathbb{P}\{X \leq -K\} + \mathbb{P}\{X_n \leq -K\} = F(-K) + F_n(-K) \to 2F(-K) < \epsilon/2$$

Similarly,

$$\mathbb{P}\{X > K\} + \mathbb{P}\{X_n > K\} = 1 - F(K) + 1 - F_n(K) \to 2\left(1 - F(K)\right) < \epsilon/2.$$

Therefore, the sum of the first two terms on the right hand side of (2) is bounded by $\epsilon\|f\|_\infty$. The last term is handled as in the previous paragraph, since fg has compact support. $\square$

Proof (⇐). Suppose (1) holds for all bounded and continuous functions f. Let a be a point of continuity of F. Let f_m and g_m be continuous and bounded functions such that

$$f_m(x) = \begin{cases} 1 & \text{if } x \leq a - 1/m \\ 0 & \text{if } x \geq a \end{cases}$$

and

$$g_m(x) = \begin{cases} 1 & \text{if } x \leq a \\ 0 & \text{if } x \geq a + 1/m. \end{cases}$$

Moreover, suppose f_m and g_m are linear on $[a - 1/m, a]$ and $[a, a + 1/m]$, respectively.

Note that $f_m \leq I_{(-\infty,a]} \leq g_m$. Therefore,

$$\mathbb{E}\left[f_m(X_n)\right] \leq F_n(a) \leq \mathbb{E}\left[g_m(X_n)\right]$$

since

$$\mathbb{E}\left[I_{(-\infty,a]}(X_n)\right] = \mathbb{P}\left\{X_n \leq a\right\} = F_n(a).$$

Now, by (1), we can take $n \to \infty$ to get $\mathbb{E}[f_m(X_n)] \to \mathbb{E}[f_m(X)]$ and $\mathbb{E}[g_m(X_n)] \to \mathbb{E}[g_m(X)]$. Therefore,

$$\mathbb{E}\left[f_m(X)\right] \leq \liminf_n F_n(a) \leq \limsup_n F_n(a) \leq \mathbb{E}\left[g_m(X)\right].$$

Moreover, as $m \to \infty$, $f_m \to I_{(-\infty,a)}$ and $g_m \to I_{(-\infty,a]}$. By the DCT,

$$\mathbb{E}\left[f_m(X)\right] \to \mathbb{E}\left[I_{(-\infty,a)}(X)\right] = \mathbb{P}\left\{X < a\right\} = F(a-)$$

and

$$\mathbb{E}\left[g_m(X)\right] \to \mathbb{E}\left[I_{(-\infty,a]}(X)\right] = \mathbb{P}\left\{X \leq a\right\} = F(a).$$

But a is a continuity point of F, and hence $F(a-) = F(a)$, from which we get the desired result. □

2 Helly's theorem

In the context of probability, Helly's theorem is a compactness result for distribution functions.

Proposition 2.1 (Helley's theorem)**.** *Let $(F_n)_n$ be a sequence of distribution functions. Then, there exists a subsequence $(n_k)_k$ and a right continuous nondecreasing function F such that*

$$F_{n_k}(x) \to F(x) \textit{ for all continuity points } x \textit{ of } F.$$

Note that F is not necessarily a distribution, since for this to be true, we must have $F(-\infty) = 0$ and $F(\infty) = 1$.

Proof. Let x be arbitrary. Note that $(F_n(x))_n$ is a bounded sequence of real numbers, and thereby admits a convergent subsequence.

Pick a dense countable subset of $\mathbb{R}$, call it $D = \{x_1, x_2, \ldots\}$ (e.g., $D = \mathbb{Q}$). By a diagonalization argument, we can find a subsequence $(F_{n_k})_k$ such that F_{n_k} converges on D. Define

$$\tilde{F}(x) = \lim_{k\to\infty} F_{n_k}(x) \text{ for } x \in D$$

and

$$\tilde{F}(x) = \sup_{D \ni x_k \leq x} F_\infty(x_k) \text{ for } x \notin D.$$

Note that $\tilde{F}$ takes values in $[0,1]$ and is nondecreasing.

Let $\mathcal{C}$ be the set of continuity points of $\tilde{F}$. Let $x \in \mathcal{C}$ be arbitrary. Choose $y, \xi \in D$ with $y < x < \xi$. Therefore, $F_{n_k}(y) \leq F_{n_k}(x) \leq F_{n_k}(\xi)$. Letting $k \to \infty$,

$$\tilde{F}(y) \leq \liminf_k F_{n_k}(x) \leq \limsup_k F_{n_k}(x) \leq \tilde{F}(\xi).$$

Since $\tilde{F}$ is continuous at x, we get

$$\lim_k F_{n_k}(x) = \tilde{F}(y).$$

The only problem now is that $\tilde{F}$ is not necessarily right continuous. Define the function F by

$$F(x) = \begin{cases} \tilde{F}(x) & \text{if } x \in \mathcal{C} \\ \lim_{y\downarrow x} \tilde{F}(y) & \text{otherwise.} \end{cases}$$

Then, F satisfies our requirements. $\square$