Recitation Note: Modes of Convergence Review

Tyler Maunu
maunut@mit.edu

September 3, 2019

This note will review some convergence concepts that are assumed for this course.

Convergence of Random Variables

There are three modes of convergence that we will use in this course. These are *convergence in distribution*, *convergence in probability*, and *convergence almost surely*. First, we will give the mathematical definition for each of these modes of convergence. Then, since the definitions themselves may not be that illuminating, we will give some examples that will help to understand what these modes of convergence mean.

In each of the following, we let $X_1, X_2, \ldots$ be a sequence of random variables.

Definition 1 (Convergence in Distribution)**.** *We say that $X_1, X_2, \ldots$ converges in distribution to a random variable X if*

$$F_{X_n}(x) \to F_X(x),\ n \to \infty. \tag{1}$$

This is pointwise convergence of the distribution function.

Definition 2 (Convergence in Probability)**.** *The sequence $X_1, X_2, \ldots$ converges in probability to a random variable X if, for all $\epsilon > 0$,*

$$P(|X_n - X| > \epsilon) \to 0,\ n \to \infty. \tag{2}$$

Definition 3 (Convergence Almost Surely)**.** *The sequence $X_1, X_2, \ldots$ converges to X almost surely if*

$$P(X_n \to X,\ n \to \infty) = 1. \tag{3}$$

We have the following implications:

$$\text{a.s. convergence} \implies \text{convergence in probability} \implies \text{convergence in distribution}. \tag{4}$$

It is not hard to prove this chain, and the curious reader can look up proofs online.[1] We will not prove these here, and instead will present some examples.

[1] See, for example, https://en.wikipedia.org/wiki/Proofs_of_convergence_of_random_variables

Examples

Example 4 (Convergence Almost Surely)**.** *Let $U_1, U_2, \ldots$ be uniform random variables on $[0,1]$ and let $X_1, X_2, \ldots$ be given by $X_n = U + U^n$ and $X = U$. For all $U \in [0,1)$, notice that $U^n \to 0$ as $n \to \infty$. On the other hand, if $U = 1$, then $X_n \to 2$ while $X = 1$. This is a probability 0 event though. Therefore, we have, for any $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = P(\lim_{n\to\infty} |X_n - X| < \epsilon | U \in [0,1)) = P(0 < \epsilon) = 1. \tag{5}$$

Example 5 (Convergence in Probability: LLN)**.** *A classic example of convergence in probability is the weak law of large numbers. Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with mean μ and variance σ^2. Let $\overline{X_n}$ be the sample mean of the first n X's in the sequence. Then, for any $\epsilon > 0$,*

$$P(|\overline{X_n} - \mu| < \epsilon) = P((\overline{X_n} - \mu)^2 < \epsilon^2) \le \frac{\mathbb{E}(\overline{X_n} - \mu)^2}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}. \tag{6}$$

Therefore, $P(|\overline{X_n} - \mu| < \epsilon) \to 0$ as $n \to \infty$.

Actually, the strong law of large numbers gives convergence almost surely, but for the purposes of this class convergence in probability is usually sufficient.

Example 6 (Convergence in Probability)**.** *Suppose that $X_1, \ldots, X_n \sim \mathsf{Unif}([0,1])$, and let $X_{(1)} = \min_i X_i$. Then, it is easy to see that $X_{(1)} \xrightarrow{p} 0$. Indeed, for any $\epsilon > 0$,*

$$P(|X_{(1)} - 0| > \epsilon) = P(X_{(1)} > \epsilon) = P(X_i > \epsilon,\ \forall i) = \left[\frac{1-\epsilon}{1}\right]^n \to 0 (n \to \infty). \tag{7}$$

Example 7 (Convergence in Probability but not Almost Surely)**.** *This example is taken from Casella and Berger [1].*

Let $U \sim \mathsf{Unif}([0,1])$. Define the sequence

$$\begin{aligned} X_1 &= U + \mathbb{1}(U \in [0,1]), \\ X_2 &= U + \mathbb{1}(U \in [0,1/2]), \\ X_3 &= U + \mathbb{1}(U \in [1/2,1]), \\ X_4 &= U + \mathbb{1}(U \in [0,1/3]), \\ X_5 &= U + \mathbb{1}(U \in [1/3,2/3]), \\ X_6 &= U + \mathbb{1}(U \in [2/3,1]), \\ &\cdots \end{aligned} \tag{8}$$

Notice that X_n takes values either U or $U+1$. It is not hard to see that $X_n \xrightarrow{p} U$: for any $\epsilon > 0$,

$$P(|X_n - U| > \epsilon) = P(U \in [a_n, b_n]) = b_n - a_n, \tag{9}$$

where $[a_n, b_n]$ is the interval in the indicator of the nth random variable. The length of these intervals goes to zero, and thus we have convergence in probability.

However, this is not convergence almost surely. Indeed, $X_n - U$ alternates between 0 and 1 infinitely often, and thus $|X_n - U|$ is not a convergence sequence.

Example 8 (Convergence in Distribution: CLT). *A typical example of convergence in distribution is the central limit theorem. This states that, for $X_1, X_2, \ldots$ i.i.d. with mean μ and variance σ^2, then*

$$\frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1). \tag{10}$$

Example 9 (Convergence in Distribution: Maximum of Uniforms). *Here we will present another example of convergence in distribution. Suppose that $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Unif}[0, \theta]$, and let $X_{(n)}$ denote their maximum. What is the distribution and asymptotic distribution of $X_{(n)}$?*

Well, we cannot just read off the distribution, and it would be wrong to use the central limit theorem here (the maximum will not give an asymptotic normal). On the other hand, we can directly calculate it. Writing down the cdf, we have

$$F_{X_{(n)}}(x) = P(X_{(n)} < x) = P(X_i < x,\ i = 1, \ldots, n) = [P(X_1 < x)]^n = \left[\frac{x}{\theta}\right]^n. \tag{11}$$

This doesn't appear to be too helpful: for any $x < \theta$, we have $F_{X_{(n)}}(x) \to 0$. Here we will use a little trick. Notice that

$$[P(X_1 < \theta(1 - t/n))]^n = \left[1 - \frac{t}{n}\right]^n \to e^{-t}. \tag{12}$$

It is not hard to see that this is equivalent to

$$P\left(n\left(1 - \frac{X_{(n)}}{\theta}\right) < t\right) \to 1 - e^{-t}. \tag{13}$$

Thus, we have shown that $n\left(1 - \frac{X_{(n)}}{\theta}\right) \xrightarrow{d} \mathsf{Exp}(1)$.

We finish by remembering Slutsky's Theorem, since it is a very useful result for combining modes of convergence for random variables.

Theorem 10. *Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$ (where a is a constant). Then,*

1. $Y_n X_n \xrightarrow{d} aX$
2. $X_n + Y_n \xrightarrow{d} X + a$

References

[1] George Casella and Roger L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.