Advanced Stochastic Processes.

David Gamarnik

LECTURE 8
Modes of convergence and convergence theorems

Lecture outline

- Modes of convergence.
- Convergence theorems.

Some remarks

The materials of this lecture is a technical toolkit for future lectures. It is a reference type summary of relevant results on different modes of convergence in probability and stochastic processes. You do not need to memorize this, as the content might seem rather chaotic. The goal is to develop enough intuition and be aware of the fact how different modes of convergence are not equivalent to each other.

8.1. Modes of convergence

There are many ways in which a sequence of random variables X_n can be said to converge to a random variable X. They are not all equivalent and in this section we introduce some of them and review their properties. Some of these properties are immediate, some we will prove and some require more complicated proof which we skip.

We already discussed **almost sure convergence**: given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of random variables $X, X_n : \Omega \to \mathbb{R}$, we say that $X_n \to X$ a.s. if

$$\mathbb{P}(\omega \in \Omega : X_n(\omega) \to X(\omega)) = 1.$$

Let us consider some other types of convergences.

- **Convergence in probability**. Given $X, X_n : \Omega \to \mathbb{R}$ we say that $X_n \to X$ in probability if for every $\epsilon > 0$ we have

$$\lim_n \mathbb{P}(|X_n - X| > \epsilon) = 0. \tag{8.1}$$

Proposition 1. Almost sure convergence implies convergence in probability. The converse is not true in general.

Remark. This is why there is a difference in what Weak Law of Large Numbers and Strong Law of Large Numbers are stating.

Proof. Suppose $X_n \to X$ a.s. Let $A_n = \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$ and $B_n = \cup_{m\geq n} A_m$. B_n is a monotone non-increasing sequence of events. Therefore by continuity theorem $\mathbb{P}(\cap_n B_n) = \lim_n \mathbb{P}(B_n)$. Notice that $\cap B_n$ is exactly the set of events ω such that $|X_n(\omega) - X(\omega)| > \epsilon$ for infinitely many n. Since $X_n \to X$ a.s. then $\mathbb{P}(\cap_n B_n) = 0$. Therefore $\lim_n \mathbb{P}(B_n) = 0$. Since $A_n \subset B_n$, then also $\lim_n \mathbb{P}(A_n) = 0$.

Counterexample A. Now we build an example of convergence in probability not implying almost sure convergence. Let $\Omega = [0,1]$. Consider Borel σ-field $\mathcal{B}$ on it and uniform probability measure $\mathbb{P}$. Define a double-index sequence $X_{n,m}$ for all $1 \leq m \leq 2^n$ of random variables as follows $X_{n,m}(\omega) = 1$ for $\omega \in [\frac{m-1}{2^n}, \frac{m}{2^n}]$ and $= 0$ otherwise. We create a single subscript sequence from $X_{n,m}$ by incrementing m for fixed n till $m = 2^n$ and then increment n and start with $m = 1$. Note that $X_{n,m} \to 0$ in probability, since the measure of the set $\{\omega : X_{n,m}(\omega) > \epsilon\} = \{\omega : X_{n,m}(\omega) = 1\}$ is $1/2^n$. However, for no ω the convergence $X_{n,m}(\omega) \to 0$ holds since every $\omega \in [0,1]$ belongs to infinitely many intervals $[\frac{m-1}{2^n}, \frac{m}{2^n}]$. □

Convergence in r-th moment. Fix $r \geq 1$. We say that $X_n \to X$ in r-the moment if $\lim_n \mathbb{E}[|X_n - X|^r] = 0$. When $r = 2$, we also say that X_n converges to X in L_2 norm.

Proposition 2. Convergence in r-th moment implies convergence in probability.

Proof. We have by Markov's inequality

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}[|X_n - X|^r]}{\epsilon^r}.$$

Convergence in probability then follows. □

The three modes of convergence we discussed corresponded to the case when X_n and X where defined on the same probability space. But sometimes the convergence is considered purely with respect to the distribution of X_n and X and/or their expectations.

- **Convergence in distribution.** Consider a sequence of random variables X_n defined on probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ (which might be all different) and a random variable X, defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $F_n(t)$ and $F(t)$ be the corresponding distribution functions. X_n is said to converge to X in distribution (written $X_n \Rightarrow X$ or $F_n \Rightarrow F$) if for every point t at which F is continuous

$$\lim_n F_n(t) = F(t).$$

Remark. Why do we require t be the continuity point of F? Consider the following example. $X_n = 1/n$ with probability one, $X_n = 0$ with probability $1 - 1/n$, and let $X = 0$ with probability one. Then $F_n(t) = 0, t < \frac{1}{n}$ and $F_n(t) = 1, t \geq \frac{1}{n}$; and $F(t) = 0, t < 0, F(t) = 1, t \geq 0$. It makes sense to say that $X_n \Rightarrow X$ as they "almost" have the same distribution. Yet $F(0) = 1 \neq \lim_n F_n(0) = 0$. This is why we exclude the points where continuity does not hold. Later on when we study weak convergence of probability measures we will elaborate on this.

Proposition 3. Convergence in probability implies convergence in distribution.

We skip the proof.

Convergence of expectations. Under the same setting, we define X_n to converge to X in expectation if $\lim_n \mathbb{E}[|X_n|] = \mathbb{E}[|X|]$. We also say that convergence up to r moments holds if $\lim_n \mathbb{E}[|X_n^r|] = \mathbb{E}[|X^r|]$

Clearly convergence in expectation does not imply convergence in distribution. Think about any sequence of variables which have the same expected value but different distributions.

Does the converse hold? In fact we could ask for a weaker question. Does almost sure convergence imply convergence in expectation? Convergence a.s. implies convergence in probability which implies convergence in distribution. So having a no answer would imply no for the previous question as well.

Here is an example when a.s. convergence does not imply convergence in expectation. Let $\Omega = [0,1]$ equipped with Borel σ-field. Consider uniform probability measure $\mathbb{P}$ on $[0,1]$.

Consider a uniform probability measure on $[0,1]$. Let $X(\omega) = 0$ for all $\omega \in [0,1]$, and let $X_n(\omega) = n$ for $\omega \in [0,1/n]$ and $X_n(\omega) = 0$ otherwise. For all non=zero ω we have $X_n(\omega) \to 0$. So $X_n \to 0$ a.s. But $\mathbb{E}[X_n] = 1$ and $\mathbb{E}[X] = 0$.

As we mentioned it does not make sense to expect that convergence in distribution implies almost sure convergence or convergence in probability, since in the first case the underlying random variables may be defined on different probability spaces.

But here is an interesting result.

Theorem 8.2 (Skorokhod's Representation Theorem). *Suppose $X_n \to X$ in distribution. There exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and a sequence of random variables $Y, Y_n : \Omega' \to \mathbb{R}$ such that $X_n \overset{d}{=} Y_n$, $X \overset{d}{=} Y$, and $Y_n \to Y$ a.s. In other words, there is a representation of X_n, X on a single probability space, where the convergence occurs almost surely.*

Why does not this contradict our counterexample showing the convergence in probability does not imply convergence in distribution?

Problem 1. Consider the Counterexample A showing that conv in probability does not imply conv almost surely. Construct a representation $Y_{n,m}, Y$ of $X_{n,m}, X$ on some common probability space such that $Y_{n,m} \to Y$ almost surely. HINT: start by computing probability distribution functions $F_{n,m}$ of $X_{n,m}$.

8.2. Tightness

Tightness is an important concept, which is often used for establishing some limiting properties of a series of probability distributions, or, in general, probability measures. To introduce the intuition behind the concept, consider a sequence of zero mean normal random variables $N(0,\sigma_n)$. If σ_n are uniformly bounded $\sup_n \sigma_n < \infty$, then, by a theorem from real analysis, there exists a subsequence $n(k)$ along which there is a convergence to some limit: $\sigma_{n(k)} \to \sigma^*$. In some sense $N(0,\sigma^*)$ is a limit of distributions $N(0,\sigma_{n(k)})$. Similarly, consider a sequence of uniform distributions on intervals $[a_n, b_n]$ such that $-\infty < \inf a_n \le \sup_n b_n < \infty$. Then again we can find a subsequence $n(k)$ such that $a_{n(k)} \to a^*$ and $b_{n(k)} \to b^*$. Then in some sense the uniform distributions on $[a_{n(k)}, b_{n(k)}]$ converge to the uniform distribution $[a^*, b^*]$. However, in the first example if $\sup \sigma_n = \infty$ we see that there is no meaningful limit of normal distributions $N(0, \sigma_n)$ along any subsequence. Similarly if either $a_n \to -\infty$ or $b_n \to \infty$, then there is no meaningful limit of uniform distributions on $[a_n, b_n]$ along any subsequence. It turns out that the important conditions for having a limit point is not so much the shaper of the distribution, but simply the fact that most of "mass" is concentrated in some bounded interval.

Definition 8.3. A sequence of distribution functions F_n is defined to be *tight* if for every $\epsilon > 0$ there exists a sufficiently large $K > 0$ such that for all n, $F_n(K) - F_n(-K) > 1 - \epsilon$. In other words, if X_n is sequence of random variables with distribution function F_n, then

$$\mathbb{P}(-K \leq X_n \leq K) > 1 - \epsilon.$$

Problem 2. Prove that a sequence $N(\mu_n, \sigma_n)$ is tight iff $\sup_n |\mu_n| < \infty$ and $\sup \sigma_n < \infty$.

The following result is the main reason for introducing the concept of tightness. This theorem is one of the possible paths for proving the existence of Wiener measure.

Theorem 8.4. *Suppose F_n is a tight sequence of distribution functions. Then there exists a distribution function F and a subsequence $n(k)$ such that along this subsequence the convergence in distribution holds: $F_{n(k)} \Rightarrow F$. That, there exists a distribution function F such that for every continuity point t of F, we have* $\lim_k F_{n(k)}(t) = F(t)$.

A converse of this theorem also holds, but we will not prove it.

Proof. Consider an enumeration $q_1, q_2, \ldots, q_r, \ldots$ of all the rational points in $\mathbb{R}$. Since the sequence of values of $F_n(q_1)$ is in $[0, 1]$, then there exists a subsequence $n_1 = n_1(k)$ along which there is a convergence of $F_{n_1(k)}(q_1)$ to some value $G(q_1)$: $\lim_k F_{n_1(k)}(q_1) = G(q_1)$. Similarly, since $F_{n_1(k)}(q_2) \in [0, 1]$, then there exists a subsequence $n_2 \subset n_1$ along which $Fn_2(k)(q_2)$ converges to some value which we denote by $G(q_2)$. We continue this subsequently for all the rational values q_r obtaining a nested sequence of subsequences $n_1 \supset n_2 \supset n_3 \supset \cdots$. No consider the diagonal subsequence $n^*(k) = n_k(k)$. Along this sequence we have convergence for all rational values simultaneously: $\lim_k F_{n^*(k)}(q_r) = G(q_r)$. Now, consider the following function $F : \mathbb{R} \to \mathbb{R}$:

$$F(t) = \inf_{r:q_r > x} G(q_r).$$

We claim that F is indeed a distribution function and $F_{n^*(k)} \Rightarrow F$.

First, we have that $F(t) \in [0, 1]$ since each $G(q_r)$ is a limit of values of $F_{n^*(k)}$. By construction F is non-decreasing: for every pair $t_1 < t_2$ we have $\{r : q_r > t_2\} \subset \{r : q_r > t_2\}$. Now we show that F is right-continuous in every point t. Fix $\epsilon > 0$. By definition, we can find a rational value $q_r > t$ such that $G(q_r) < F(t) + \epsilon$. Select any $\delta < q_r - t$. Then if $t < t' < t + \delta$, then $t' < q_r$, implying $F(t') = \inf\{G(q) : q > t'\} \leq G(q_r) < F(t) + \epsilon$. We showed that any point $t' \in [t, t + \delta)$ satisfies $F(t') \leq F(t) + \epsilon$ and right-continuity is established.

We showed that F is a right-continuous non-decreasing function taking values in $[0, 1]$. It remains to show that $\lim_{t \to -\infty} F(t) = 0, \lim_{t \to \infty} F(t) = 1$. But first we show that in every continuity point t of F we have $\lim_k F_{n^*(k)}(t) = F(t)$, and then we return to proving the limits. For the ease of presentation, let us assume that we have $\lim_n F(q_r) = G(q_r)$ for every rational value q_r. (That is we substitute subsequence n^* for the original sequence n). Thus let t be the continuity point of F. Fix $\epsilon > 0$. Pick rationals r_1, r_2, s with $r_1 < r_2 < t < s$, so that

$$F(t) - \epsilon < F(r_1) \leq F(r_2) \leq F(t) \leq F(s) \leq F(t) + \epsilon.$$

This is possible by continuity and monotonicity of F. Now $F_n(r_2) \to G(r_2) \geq F(r_1)$, and $F_n(s) \to G(s) \leq F(s)$. Therefore, for all sufficiently large n we have

$$F(t) - \epsilon < F_n(r_2) \leq F_n(t) \leq F_n(s) \leq F(t) + \epsilon.$$

We conclude that $\lim_n F_n(t) = F(t)$.

It remains to show that $\lim_{t\to-\infty} F(t) = 0, \lim_{t\to\infty} F(t) = 1$, and this is exactly where we use tightness. Fix $\epsilon > 0$. There exists K such that $F_n(K) \geq F_n(K) - F_n(-K) \geq 1 - \epsilon$. Without loss of generality take K to be the continuity point of F. (The set of points of discontinuity of F has Lebeasgue measure zero, per real analysis, so we can simply find $K' > K$ which is a continuity point). By monotonicity of F_n, we will have $F_n(K') - F_n(-K') \geq 1 - \epsilon$). Then $F_n(K) \to F(K)$ and $F_n(-K) \to F(-K)$. Therefore $1 \geq F(K) \geq F(K) - F(-K) \geq 1 - \epsilon$. This implies that $\lim_{t\to\infty} F(t) = 1$. From the same bound we have $1 - F(-K) \geq F(K) - F(-K) \geq 1 - \epsilon$ or $F(-K) \leq \epsilon$, implying $\lim_{t\to-\infty} F(t) = 0$. This proves that F is indeed a distribution function. □

8.3. Convergence theorems

We now state without proof several useful convergence theorems.

Theorem 8.5 (**Dominated Convergence Theorem**). *Suppose $X_n \to X$ a.s. and suppose there exists a random variable $Y \geq |X_n|$ a.s. such that $\mathbb{E}[Y] < \infty$. Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$. In particular, suppose X_n are a.s. bounded. That is $|X_n| \leq B$ a.s., for some $B > 0$. Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

Note that it is not enough to assume that just $\sup_n \mathbb{E}[|X_n|] < \infty$: consider the example where almost sure convergence does not imply convergence in expectation. There the expected value of each variable was $\mathbb{E}[X_n] = 1$, but $\mathbb{E}[X] = 0$.

Theorem 8.6 (**Monotone Convergence Theorem**). *Suppose $X_n \to X$ a.s. and suppose for each sample $\omega, X_n(\omega), n \in \mathbb{N}$ is non-decreasing sequence. Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$.*

Theorem 8.7 (**Continuous Mapping Theorem**). *Suppose $X_n \to X$ in distribution. Let $g : \mathbb{R} \to \mathbb{R}$ be continuous function. Then $g(X_n) \to g(X)$ in distribution. If, in addition g is bounded, then $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$.*

Proof. We use Skorokhod Representation Theorem. Find a representation Y_n, Y of X_n, X on some probability space where $Y_n \to Y$ a.s. For every sample ω, such that $Y_n(\omega) \to Y(\omega)$, we have by continuity, $g(Y_n(\omega)) \to g(Y(\omega))$. Since the set of such samples ω has measure 1, then $g(Y_n) \to g(Y)$ a.s. This implies that convergence also holds in probability and therefore also in distribution. Therefore $g(Y_n) \Rightarrow g(Y)$. Suppose in addition g is bounded. Then by DCT, we have $\mathbb{E}[g(Y_n)] \to \mathbb{E}[g(Y)]$. This implies $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$. □

8.4. Additional reading materials

- Grimmett and Stirzaker [**2**] Section 7.2.
- Durrett [**1**] Sections 1.3, 2.2.

BIBLIOGRAPHY

1. R. Durrett, *Probability: theory and examples*, Duxbury Press, second edition, 1996.
2. G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*, Oxford Science Publications, 1985.