IEOR 6711: Stochastic Models I

Fall 2013, Professor Whitt

Lecture Notes, Thursday, September 5

Modes of Convergence

1 Overview

We started by stating the two principal laws of large numbers: the strong and weak forms, denoted by SLLN and WLLN. We want to be clear in our understanding of the statements; that led us to a careful definition of a **random variable** and now a careful examination of the basic **modes of convergence for a sequence of random variables**. We also want to focus on the proofs, but in this course (as in the course textbook) we consider only relatively simple **proofs that apply under extra moment conditions**. Even with these extra conditions, important proof techniques appear, which relate to the basic axioms of probability, in particular, to **countable additivity**, which plays a role in understanding and proving the **Borel-Cantelli lemma** (p. 4) . We think that it is helpful to fucus on these more elementary cases before considering the most general conditions.

Key reading for the present lecture: §§1.1-1.3, 1.7-1.8, the Appendix, pp. 56-58.

2 Modes of Convergence

As general background, we need to understand the relation among the basic modes of convergence for a sequence of random variables. The main idea is to realize that there is more than one mode of convergence.

Consider the following limits (as $n \to \infty$):

(i) $X_n \Rightarrow X$ (convergence in distribution)

Definition: That means $F_n(x) \to F(x)$ for all x that are continuity points of the cdf F, where $F_n(x) \equiv P(X_n \le x)$ and $F(x) \equiv P(X \le x)$. A point x is a continuity point of F if (and only if) F does not have a jump at x, i.e., if $F(x) = F(x-)$, where $F(x-)$ is the left limit of F at x. (Recall that F is right continuous by definition; since it is monotone, the left limits necessarily exist.)

(ii) $EX_n \to EX$ (convergence of means)

This is just convergence for a sequence of numbers; the mean is only a partial summary of the full distribution. Obviously convergence of means is a relatively weak property. The meaning of convergence of a sequence of numbers should be clear:

$$\lim_{n\to\infty} x_n = x$$

means that: for all $\epsilon > 0$, there exists an integer $n_0 \equiv n_0(\epsilon)$ such that, for all $n \ge n_0$, $|x_n - x| < \epsilon$.

(iii) $E[|X_n - X|] \to 0$ (convergence in the mean, convergence in L_1)

This is convergence in the mean: for all $\epsilon > 0$, there exists an integer $n_0 \equiv n_0(\epsilon)$ such that, for all $n \ge n_0$, $E[|X_n - X|] < \epsilon$.

We also sometimes consider convergence in the p^{th} mean, especially for $p = 2$.

(iii') $E[|X_n - X|^p] \to 0$ (convergence in the p^{th} mean, convergence in L_p)

(iv) $P(X_n \to X) = 1$ (convergence w.p.1)

This is convergence with probability one (w.p.1). Elaborating, we are saying that

$$P(\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\}) = 1 \ ,$$

where the limit inside is for a sequence of real numbers (for each s), and thus defined just as in (ii) above. (Here we use S for the underlying sample space. I alternate between S and Ω.)

(v) $X_n \to X$ in probability

Here is a definition: For all $\epsilon > 0$ and $\eta > 0$, there exists an integer n_0 such that, for all $n \ge n_0$,

$$P(|X_n - X| > \epsilon) < \eta \ .$$

In general, convergence in probability implies convergence in distribution. However, when the limiting random variable X is a constant, i.e., when $P(X = c) = 1$ for some constant c, the two modes of convergence are equivalent; e.g., see p. 27 of Billingsley, *Convergence of Probability Measures*, second ed., 1999.

References on this topic. limits for sequences of numbers: Chapter 3 of Rudin, Principles of Math. Analysis

limits for sequences of functions: Chapter 7 of Rudin, Principles of Math. Analysis

convergence concepts: Chapter 4 of Chung, A Course in Probability Theory

expectation and convergence of random variables: some of Chapters V and VIII of Feller II

related supplementary reading: Billingsley, Probability and Measure, Chapters 1, 6, 14, 21, 25

3 Relations Among the Modes of Convergence

Consider the following limits (as $n \to \infty$):

(i) $X_n \Rightarrow X$ (convergence in distribution)
(ii) $EX_n \to EX$ (convergence of means)
(iii) $E[|X_n - X|] \to 0$ (convergence in the mean, also known as convergence in L_1, using standard functional analysis terminology)
(iv) $P(X_n \to X) = 1$ (convergence with probability 1 (w.p.1))
(v) $X_n \to X$ in probability (convergence in probability)

What are the implications among these limits? For example, does limit (i) imply limit (ii)?

See Chapter 4 of Chung, *A Course in Probability Theory* for a discussion of convergence concepts. The following implications hold:

$(iv) \to (v) \to (i), \quad (iii) \to (ii) \quad \text{and} \quad (iii) \to (v) \to (i)$

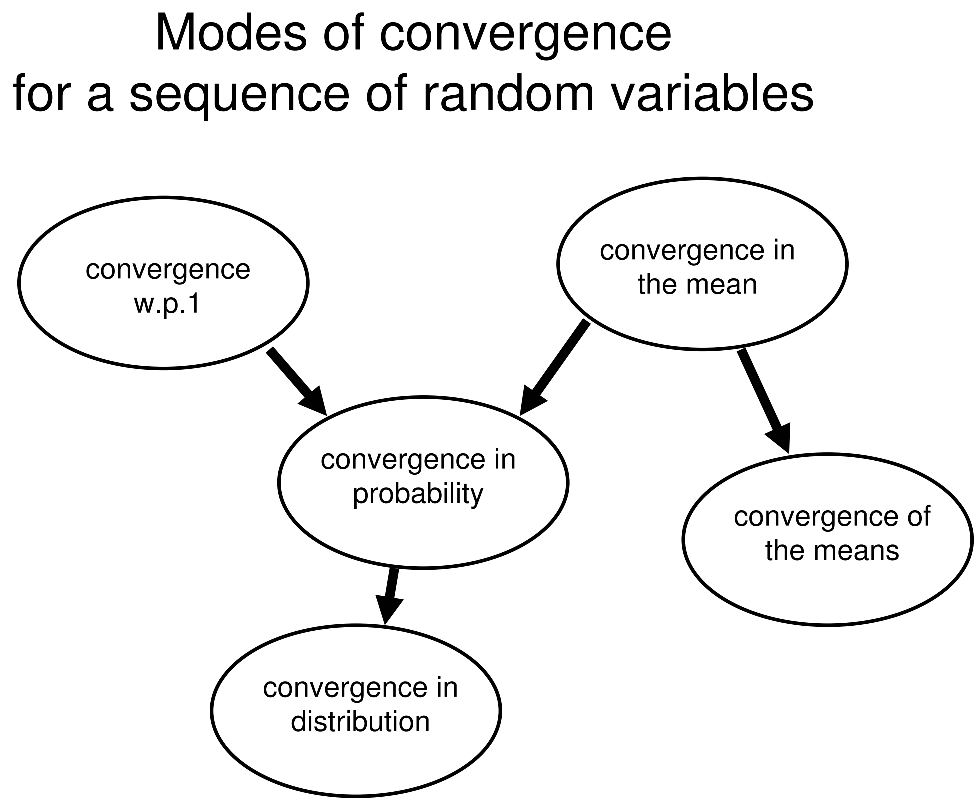

Figure 1: Relations among the five modes of convergence for a sequence of random variables: $X_n \to X$ as $n \to \infty$.

We remark that convergence in probability is actually equivalent to convergence in distribution, i.e., also (i) implies (v), for the special case in which the limiting random variable X is a constant. That explains why the WLLN can be expressed as convergence in distribution.

For all the relations that do not hold, it suffices to give two counterexamples. These examples involve specifying the underlying probability space and the random variables X_n and X, which are (measurable) functions mapping this space into the real line $\mathbb{R}$. In class we used pictures to describe the functions, where the x axis represented the domain and the y-axis represented the range. In both examples, the underlying probability space is taken to be the interval $[0, 1]$ with the uniform probability distribution.

Counterexample 1: To see that convergence with probability 1 (w.p.1) (iv) does not imply convergence of means (ii) (and thus necessarily convergence of means (iii)), let the underlying probability space be the unit interval $[0, 1]$ with the uniform distribution (which coincides with Lebesgue measure). Let $X = 0$ w.p.1 and let $X_n = 2^n$ on the interval $(a_n, a_n + 2^{-n})$ where $a_n = 2^{-1} + 2^{-2} + \cdots + 2^{-(n-1)}$ with $a_1 = 0$, and let $X_n = 0$ otherwise. Then $P(X_n \to X \equiv 0) = 1$, but $E[|X_n - X|] = EX_n = 1$ for all n, but $EX = 0$. (To see that indeed $P(X_n \to X \equiv 0) = 1$, note that the interval on which X_k is positive for any $k > n$ has probability going to 0 as $n \to \infty$.) We could make $E[X_n]$ explode (diverge to $+\infty$ if we redefined X_n to be $n2^n$ where it is positive.

From the example above, it follows that convergence in probability (v) does not imply convergence of means (ii) and that convergence in distribution (i) does not imply convergence of means (ii).

However, convergence in distribution (i) does imply convergence of means (ii) under *extra regularity conditions*, namely, under *uniform integrability*. See p. 32 of Billingsley, *Convergence of Probability Measures*, 1968, for more discussion. If $X_n \Rightarrow X$ and if $\{X_n : n \geq 1\}$ is uniformly integrable (UI), then $E[X_n] \to E[X]$ as $n \to \infty$. The online sources give more.

For a sample paper full of UI issues, see
Wanmo Kang, Perwez Shahabuddin and WW. Exploiting Regenerative Structure to Estimate Finite Time Averages via Simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 17, No. 2, April 2007, Article 8, pp. 1-38.

Available at: http://www.columbia.edu/∼ww2040/recent.html

Counterexample 2: To see that convergence in the mean (iii) does not imply convergence w.p.1 (iv), again let the underlying probability space be the unit interval $[0, 1]$ with the uniform distribution (which coincides with Lebesgue measure). Let $X = 0$ w.p.1. Somewhat like before, let $X_n = 1$ on the interval $(a_n, a_n + n^{-1})$ where $a_n = a_{n-1} + (n-1)^{-1}$ mod 1, with $a_1 = 0$, and let $X_n = 0$ otherwise. (The mod1 means that there is "wrap around" from 1 back to 0.) (To see that indeed $P(X_n \to X \equiv 0) = 0$, note that the $X_k = 1$ infinitely often for each sample point. On the other hand, $E[|X_n - X|] = EX_n = 1/n \to 0$ as $n \to \infty$.

4 Reducing Everything to Convergence w.p.1

An important unifying concept is the Skorohod Representation Theorem; it shows how convergence in distribution can be expressed in terms of convergence w.p.1. See homework 1. Also see Section 1.3 of the Internet Supplement to my book, available for downloading from my web page. It might also be helpful to read the introduction (3 pages) to my paper, "Some Useful Functions for Functional Limit Theorems." *Mathematics of Operations Research*, vol. 5, No. 1, February 1980, pp. 67-85. This is also available for downloading. Both these sources point to the power of these ideas in the general context of stochastic processes instead of just real-valued random variables.

5 Proof of the SLLN

5.1 Moment Inequalities

We now turn to the proof of Theorem **??**. To appreciate the extra conditions in Theorem **??** above, you want to know the following moment inequality.

Theorem 5.1 *Suppose that $r > p > 0$. If $E[|X|^r] < \infty$, then $E[|X|^p] < \infty$.*

Theorem 5.1 follows from:

Theorem 5.2 *Suppose that $r > p > 0$. Then*

$$E[|X|^p] < (E[|X|^r])^{p/r} \ .$$

Theorem 5.2 follows from *Hölder's inequality*:

Theorem 5.3 *Suppose that $p > 1$, $q > 1$ and $p^{-1} + q^{-1} = 1$. Suppose that $E[|X|^p] < \infty$ and $E[|Y|^q] < \infty$. Then*

$$E[|XY|] \le (E[|X|^p])^{1/p} (E[|Y|^q])^{1/q} .$$

Hölder's inequality in Theorem 5.3 can be proved by exploiting the concavity of the logarithm. To apply Hölder's inequality, replace X by X^p and Y by 1. Let the exponents be $r/p > 1$ and q such that $(1/q) + (p/r) = 1$.

Google the law of large numbers and Hölder's inequality.

5.2 The Proof in the Appendix to Chapter 1 in Ross

The proof is given on pages 56-58 of Ross. The key idea is to apply the Borel-Cantelli lemma, Proposition 1.1.2 on p. 4, as Ross indicates on the bottom of p. 57. For further discussion, let the mean be 0. We apply the Borel-Cantelli Lemma to show that, for all $\epsilon > 0$,

$$P(|S_n/n| > \epsilon \quad \text{infinitely often}) = 0. \tag{1}$$

We do so by showing that

$$\sum_{n=1}^{\infty} P(|S_n/n| > \epsilon) < \infty. \tag{2}$$

There are two remaining issues: (1) How do we establish this desired inequality in (2), and (ii) Why does that suffice to prove (1)?

To see why it suffices to prove (1), define the events

$$A_k \equiv \{s \in S : |S_n(s)/n| > 1/k \quad \text{infinitely often}\}, \quad k \ge 1.$$

Here "infinitely often" mean for infinitely many n. We observe that

$$\{s \in S : S_n(s)/n \to 0 \quad \text{as} \quad n \to \infty\}^c = \bigcup_{k=1}^{\infty} A_k.$$

However, by the continuity property of increasing sets, Proposition 1.1.1 on p. 2 of Ross, which itself is equivalent to the fundamental countable additivity axiom, we have

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{K\to\infty} P\left(\bigcup_{k=1}^{K} A_k\right) = \lim_{K\to\infty} P\left(A_K\right).$$

Hence, it suffices to show that $P(A_K) = 0$ for each K, which is equivalent to (1).

Now we turn to the second point: verifying (2). Now we exploit Markov's inequality, using the fourth power, getting

$$P(|S_n/n| > \epsilon) < \frac{E[S_n^4]}{n^4 \epsilon^4}.$$

We then show that $E[S_n^4] \le K n^2$. (Details are given in Ross.) Hence,

$$P(|S_n/n| > \epsilon) < K'/n^2$$

for some new constant K'. Thus we can deduce the required (2). (Note that this argument does not work if we use the second power instead of the fourth power.)

6 The Continuous Mapping Theorem

a. continuous-mapping theorem [see Section 3.4 of my book online]

Theorem 6.1 *If $X_n \Rightarrow X$ as $n \to \infty$ and f is a continuous function, then*

$$f(X_n) \Rightarrow f(X) \quad as \quad n \to \infty \ .$$

This can be approved by applying the Skorohod representation theorem (homework 1): We start by replacing the original random variables by the new random variables, say X_n^* and X^*, on a new underlying probability space that converge w.p.1, and have the property that the probability law of X_n^* coincides with the probability law of X_n for each n and the probability law of X^* coincides with the probability law of X. Then we get convergence $f(X_n^*) \to f(X^*)$ w.p.1 by the assumed continuity. But this w.p.1 convergence implies convergence in distribution: $f(X_n^*) \Rightarrow f(X^*)$ because w.p.1 convergence always implies convergence in distribution (or in law). However, since the asterisk random variables have the same distributions as the random variables without the asterisks, we have the desired $f(X_n) \Rightarrow f(X)$.

[see Theorem 0.1 in Problem 3 (d) in Homework 1; also see the Skorohod Representation Theorem at the end of Section 3.2 in my book; see Section 1.3 of the Internet Supplement for a proof in the context of separable metric spaces]

b. If $X_n \to X$ and $Y_n \to Y$, does $(X_n, Y_n) \to (X, Y)$? [In general, the answer is no, but the answer is yes if either X or Y is deterministic. The answer is also yes if X_n is independent of Y_n for all n. See Section 11.4 of my book online.]

(c) Some questions:

Suppose that

$$V_n \Rightarrow N(2,3), \quad W_n \Rightarrow N(1,5), \quad X_n \Rightarrow 7 \quad \text{and} \quad Y_n \Rightarrow 3 \ .$$

(Convergence in distribution as $n \to \infty$)

(a) Does $V_n^2 \Rightarrow N(2,3)^2$? [yes, by continuous mapping theorem; See Section 3.4 of my book]

(b) Is $N(2,3)^2 \stackrel{\mathrm{d}}{=} N(4,9)$, where $\stackrel{\mathrm{d}}{=}$ means *equal in distribution*?
[no, square of normal has chi squared distribution]

(c) Does $\sqrt{V_n} e^{(V_n^3 - 12V_n^2)} \Rightarrow \sqrt{N(2,3)} e^{(N(2,3)^3 - 12N(2,3)^2)}$
[yes, again by continuous mapping theorem]

(d) Does $V_n + W_n \Rightarrow N(3,8)$?
[not necessarily, if V_n and W_n are independent then true]

(e) Does $V_n + X_n \Rightarrow N(9,3)$? [yes, See Section 11.4 of my book, we first have $(V_n, X_n) \Rightarrow (V, X)$]

(f) Does $V_n^4 (e^{(X_n^2 + Y_n^3)} - W_n V_n^{13} + 6 \Rightarrow N(2,3)^4 (e^{(49+27)} - N(1,5) N(2,3)^{13} + 6$?
[yes, under extra assumptions to get vector convergence: $(V_n, W_n, X_n, Y_n) \Rightarrow (V, W, X, Y)$, where random variables on the right are independent (only important for random ones), independence suffices; then apply continuous mapping theorem]

(g) Answer questions (a) - (e) above under the condition that $\Rightarrow$ in all the limits is replaced by convergence with probability one (w.p.1).

[almost same answers]

(h) Answer questions (a) - (e) above under the condition that $\Rightarrow$ in all the limits is replaced by convergence in probability.

[almost same answer, because there is a w.p.1 representation of convergence in probability: $X_n \to X$ in probability if and only if for all subsequences $\{X_{n_k}\}$ there exists a further subsubsequence $\{X_{n_{k_j}}\}$ such that for this subsequence there is convergence w.p.1 to X.]