IEOR 6711: Stochastic Models I

Fall 2012, Professor Whitt

Lecture Notes, Tuesday and Thursday, September 4 and 6

Laws of Large Numbers

1 Overview

We start 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 leads us to a careful definition of a **random variable** and an 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 The Classical Laws of Large Numbers

Theorem 2.1 *Let* $X_1, X_2, \ldots$ *be IID random variables. Let* $S_n \equiv X_1 + \cdots + X_n$*,* $n \geq 1$*,* $S_0 \equiv 0$*, be the associated partial sums. If* $E[|X_1|] < \infty$*, then*

(a) ***SLLN***

$$\frac{S_n}{n} \rightarrow E[X_1] \quad as \quad n \rightarrow \infty \quad w.\ p.\ 1\ .$$

(b) ***WLLN***

$$\frac{S_n}{n} \Rightarrow E[X_1] \quad as \quad n \rightarrow \infty\ ,$$

where $\equiv$ *denotes equality by definition, w.p.1 is convergence with probability 1 (almost sure convergence) and* $\Rightarrow$ *denotes convergence in distribution.*

Definition 2.1 (*convergence in distribution*) *There is convergence* $Y_n \Rightarrow Y$ *if the associated probability distributions of these random variables converge, i.e., if* $P_{Y_n} \rightarrow P_Y$ *as* $n \rightarrow \infty$ *or, equivalently, if the associated cumulative distribution functions (cdf's) converge, i.e., if* $F_{Y_n}(x) \rightarrow F_Y(x)$ *as* $n \rightarrow \infty$ *for all* x *that are continuity points of the function* $F_Y(x)$*, where* $F_Y(x) \equiv P(Y \leq x)$*. (More on this later.)*

Proof of (a) under extra condition $E[X_1^4] < \infty$. p. 56 of Ross. Draws heavily on Borel-Cantelli and thus §1.1.

Proof of (b) under extra condition $E[X_1^2] < \infty$. Two parts: (i)

$$E\left[\left(\frac{S_n}{n} - E[X_1]\right)^2\right] = Var\left(\frac{S_n}{n}\right) = \frac{Var(X_1)}{n} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty\ .$$

(ii) Convergence in mean squared (L_2) implies convergence in probability, which in turn implies convergence in distribution. In fact, convergence in probability is equivalent to convergence in distribution when the limit is constant (deterministic or non-random). Chebychev's inequality (which is a special case of Markov's inequality, Lemma 1.7.1 in the book) shows convergence in mean squared implies convergence in probability. ■

We remark that the IID condition can also be relaxed in a variety of ways; there is a large literature.

3 Random Variables and Functions of Random Variables

To properly understand why there are two versions of the LLN, it is necessary to understand what is a random variable and the possible modes of convergence for a sequence of random variables.

(i) What is a **random variable**?

A (real-valued) random variable, often denoted by X (or some other capital letter), is a **function** mapping a probability space (S, P) into the real line $\mathbb{R}$. This is shown in Figure 1. (It is also common to write Ω for the sample space and ω for an element in that set.) Associated with each point s in the domain S the function X assigns one and only one value $X(s)$ in the range $\mathbb{R}$. (The set of possible values of $X(s)$ is usually a proper subset of the real line; i.e., not all real numbers need occur. If S is a finite set with m elements, then $X(s)$ can assume at most m different values as s varies in S.)

As such, a random variable has a probability distribution. We usually do not care about the underlying probability space, and just talk about the random variable itself, but it is good to know the full formalism. The distribution of a random variable is defined formally in the obvious way

$$F(t) \equiv F_X(t) \equiv P(X \le t) \equiv P(\{s \in S : X(s) \le t\}) \ ,$$

where again $\equiv$ means "equality by definition," P is the probability measure on the underlying sample space S and $\{s \in S : X(s) \le t\}$ is a subset of S, and thus an *event* in the underlying sample space S. See Section 1.1 of Ross; he puts this out very quickly. (Key point: recall that P attaches probabilities to events, which are subsets of S.)

If the underlying probability space is discrete, so that for any event E in the sample space S we have

$$P(E) = \sum_{s \in E} p(s),$$

where $p(s) \equiv P(\{s\})$ is the *probability mass function* (pmf), then X also has a pmf p_X on a new sample space, say S_1, defined by

$$p_X(r) \equiv P(X = r) \equiv P(\{s \in S : X(s) = r\}) = \sum_{s \in \{s \in S : X(s) = r\}} p(s) \quad \text{for} \quad r \in S_1. \qquad (1)$$

Example 3.1 (*roll of two dice*) Consider a random roll of two dice. The natural sample space is

$$S \equiv \{(i, j) : 1 \le i \le 6, 1 \le j \le 6\},$$

where each of the 36 points in S is assigned equal probability $p(s) = 1/36$. The random variable X might record the sum of the values on the two dice, i.e., $X(s) \equiv X((i, j)) = i + j$. Then the new sample space is

$$S_1 = \{2, 3, 4, \ldots, 12\}.$$

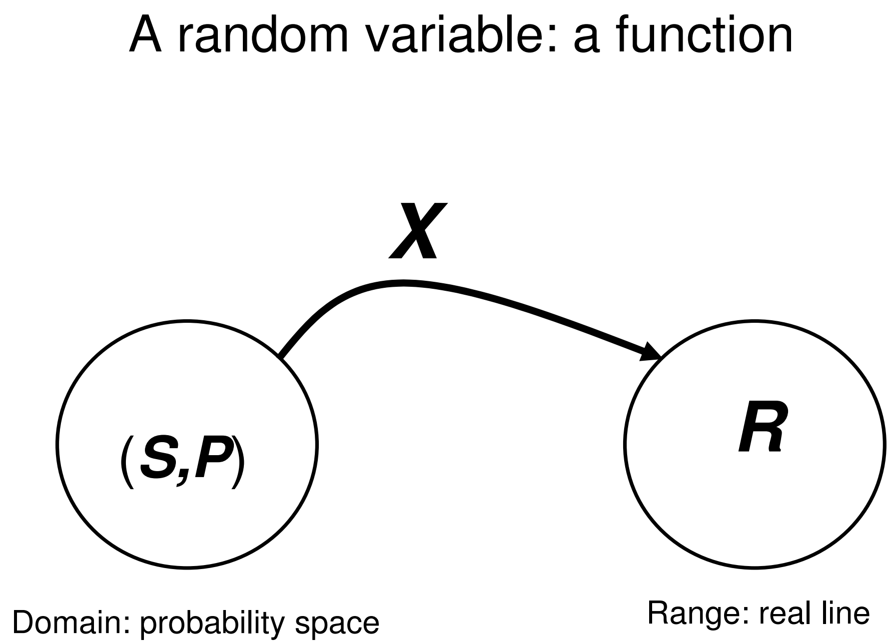

Figure 1: A (real-valued) random variable is a function mapping a probability space into the real line.

In this case, using formula (1), we get the pmf of X being $p_X(r) \equiv P(X = r)$ for $r \in S_1$, where

$$\begin{aligned} p_X(2) &= p_X(12) = 1/36, \\ p_X(3) &= p_X(11) = 2/36, \\ p_X(4) &= p_X(10) = 3/36, \\ p_X(5) &= p_X(9) = 4/36, \\ p_X(6) &= p_X(8) = 5/36, \\ p_X(7) &= 6/36. \end{aligned}$$

(ii) What is a **function of a random variable**?

Given that we understand what is a random variable, we are prepared to understand what is a function of a random variable. Suppose that we are given a random variable X mapping the probability space (S, P) into the real line $\mathbb{R}$ and we are given a function h mapping $\mathbb{R}$ into $\mathbb{R}$. Then $h(X)$ is a function mapping the probability space (S, P) into $\mathbb{R}$. As a consequence, $h(X)$ is itself a new random variable, i.e., a new function mapping (S, P) into $\mathbb{R}$, as depicted in Figure 2.

As a consequence, the distribution of the new random variable $h(X)$ can be expressed in different (equivalent) ways:

$$F_{h(X)}(t) \equiv P(h(X) \le t) \quad \equiv \quad P(\{s \in S : h(X(s)) \le t\}),$$

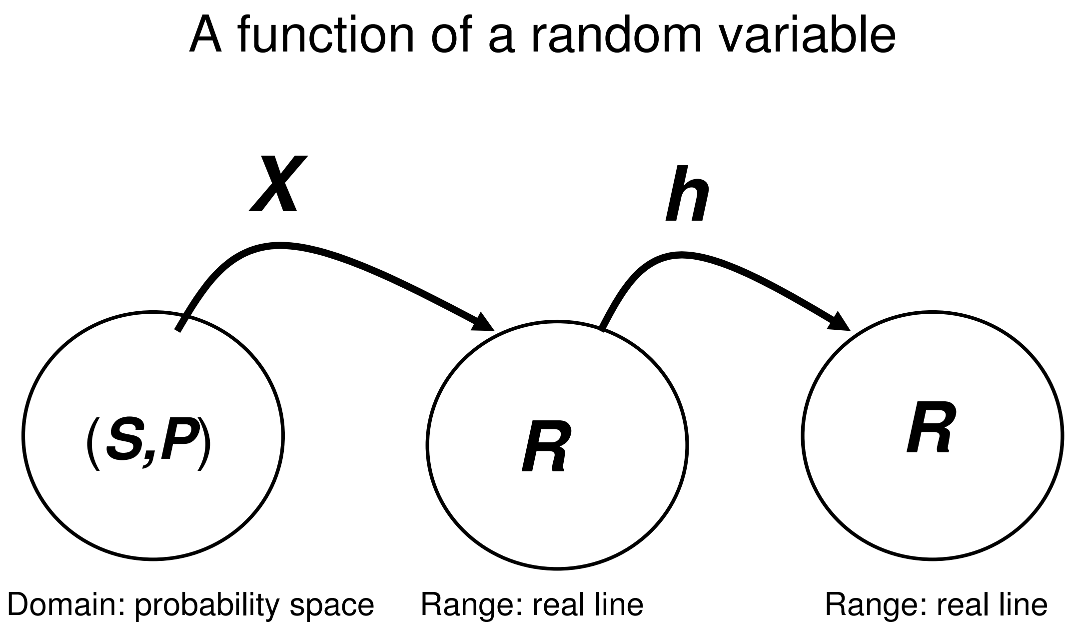

Figure 2: A (real-valued) function of a random variable is itself a random variable, i.e., a function mapping a probability space into the real line.

$$\begin{aligned} &\equiv P_X(\{r \in \mathbb{R} : h(r) \leq t\}), \\ &\equiv P_{h(X)}(\{k \in \mathbb{R} : k \leq t\}), \end{aligned}$$

where P is the probability measure on S in the first line, P_X is the probability measure on $\mathbb{R}$ (the distribution of X) in the second line and $P_{h(X)}$ is the probability measure on $\mathbb{R}$ (the distribution of the random variable $h(X)$ in the third line.

Example 3.2 (*more on the roll of two dice*) As in Example 3.1, consider a random roll of two dice. There we defined the random variable X to represent the sum of the values on the two rolls. Now let

$$h(x) = |x - 7|,$$

so that $h(X) \equiv |X - 7|$ represents the absolute difference between the observed sum of the two rolls and the average value 7. Then $h(X)$ has a pmf on a new probability space $S_2 \equiv \{0, 1, 2, 3, 4, 5\}$. In this case, using formula (1) yet again, we get the pmf of $h(X)$ being $p_{h(X)}(k) \equiv P(h(X) = k) \equiv P(\{s \in S : h(X(s)) = k\})$ for $k \in S_2$, where

$$\begin{aligned} p_{h(X)}(5) &= P(h(X) = 5) \equiv P(|X - 7| = 5) = 2/36 = 1/18, \\ p_{h(X)}(4) &= P(h(X) = 4) \equiv P(|X - 7| = 4) = 4/36 = 2/18, \\ p_{h(X)}(3) &= P(h(X) = 3) \equiv P(|X - 7| = 3) = 6/36 = 3/18, \\ p_{h(X)}(2) &= P(h(X) = 2) \equiv P(|X - 7| = 2) = 8/36 = 4/18, \end{aligned}$$

$$\begin{aligned} p_{h(X)}(1) &= P(h(X)=1) \equiv P(|X-7|=1) = 10/36 = 5/18, \\ p_{h(X)}(0) &= P(h(X)=0) \equiv P(|X-7|=0) = 6/36 = 3/18. \end{aligned}$$

In this setting we can compute probabilities for events associated with $h(X) \equiv |X-7|$ in three ways: using each of the pmf's p, p_X and $p_{h(X)}$.

(iii) How do we compute the **expectation** (or expected value) of a (probability distribution) or a random variable?

See Section 1.3. The expected value of a discrete probability distribution P is

$$\text{expected value} \quad = \quad \text{mean} \quad = \sum_k kP(\{k\}) = \sum_k kp(k) \ ,$$

where P is the probability measure on S and p is the associated pmf, with $p(k) \equiv P(\{k\})$. The expected value of a discrete random variable X can be written in two ways, as shown in the two lines below:

$$\begin{aligned} E[X] &= \sum_k kP(X=k) = \sum_k kp_X(k) \\ &= \sum_{s \in S} X(s)P(\{s\}) = \sum_{s \in S} X(s)p(s) \ . \end{aligned}$$

In the continuous case, with probability density functions (pdf's), we have corresponding formulas, but the story gets more complicated, involving calculus for computations. The expected value of a continuous probability distribution P with density f is

$$\text{expected value} \quad = \quad \text{mean} \quad = \int_{s \in S} xf(x)\, dx \ .$$

The expected value of a continuous random variable X with pdf f_X is

$$E[X] = \int_{-\infty}^{\infty} xf_X(x)\, dx = \int X(s)f(s)\, ds \ ,$$

where f is the pdf on S and f_X is the pdf "induced" by X on $\mathbb{R}$.

(iv) How do we compute the **expectation of a function of a random variable**?

Now we need to put everything above together. For simplicity, suppose S is a finite set, so that X and $h(X)$ are necessarily finite-valued random variables. Then we can compute the expected value $E[h(X)]$ in **three different ways**:

$$\begin{aligned} E[h(X)] &= \sum_{s \in S} h(X(s))P(\{s\}) = \sum_{s \in S} h(X(s))p(s) \\ &= \sum_{r \in \mathbb{R}} h(r)P(X=r) = \sum_{r \in \mathbb{R}} h(r)p_X(r) \\ &= \sum_{t \in \mathbb{R}} tP(h(X)=t) = \sum_{t \in \mathbb{R}} tp_{h(X)}(t) \ , \end{aligned}$$

where $p(s) \equiv P(\{s\})$ is the pmf associated with P on S, while $p_X(r)$ is the pmf of X and $p_{h(X)}(t)$ is the pmf of $h(X)$.

Similarly, we have the following expressions when all these probability distributions have probability density functions (the continuous case). First, suppose that the underlying probability distribution (measure) P on the sample space S has a probability density function (pdf) f. Then, under regularity conditions, the random variables X and $h(X)$ have probability density functions f_X and $f_{h(X)}$. Then we have:

$$\begin{aligned} E[h(X)] &= \int_{s\in S} h(X(s))f(s)\,ds \\ &= \int_{-\infty}^{\infty} h(r)f_X(r)\,dr \\ &= \int_{-\infty}^{\infty} t f_{h(X)}(t)\,dt \ . \end{aligned}$$

4 Implications for the LLN

So what does our study of random variables imply for the LLN? We see that the SLLN states that $S_n/n \to E[X_1]$ as $n \to \infty$ w.p.1. We have almost sure convergence of the sequence of functions $\{S_n/n\}$, all defined on the underlying sample space S. That is, for each s in a subset of S having probability 1, we have convergence of the sequence of numbers $S_n(s)/n \to E[X_1]$ as $n \to \infty$. All this is defined precisely in terms of the basic notion of the convergence of a sequence of numbers.

On the other hand, the WLLN states that $S_n/n \Rightarrow E[X_1]$ as $n \to \infty$. That means that the associated probability distributions converge. The statement concerns the probability distributions on the range of the functions (random variables). That is, $F_{S_n/n}(x) \to F_{E[X_1]}(x)$ and $n \to \infty$ for all x that are continuity points of the limiting cdf $F_{E[X_1]}(x)$. The only value that is *not* a continuity point is the mean itself $E[X_1]$, where the cdf has a unit jump.

We often are satisfied with the WLLN, because we often only care about the distributions of the random variables.

5 Moment Inequalities

We now turn to the proof of Theorem 2.1. To appreciate the extra conditions in Theorem 2.1 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.

6 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 $F_n(x) \equiv P(X_n \leq x)$ and $F(x) \equiv P(X \leq x)$. 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 $n \geq n_0$, $|x_n - x| < \epsilon$.

(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 $n \geq n_0$, $E[|X_n - X|] < \epsilon$.

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 $n \geq n_0$,

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

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

7 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:
$(iv) \to (v) \to (i), \quad (iii) \to (ii) \quad \text{and} \quad (iii) \to (v) \to (i)$

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

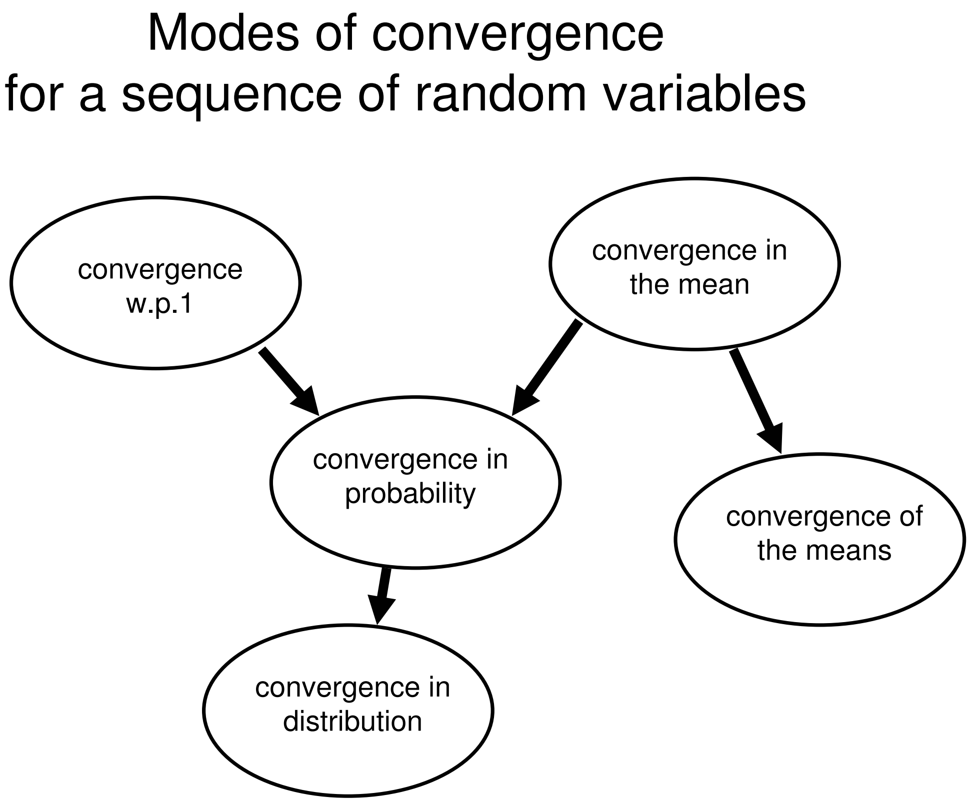

Figure 3: Relations among the five modes of convergence for a sequence of random variables: $X_n \to X$ as $n \to \infty$.

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 $n \to \infty$.

8 Proof of the SLLN

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 $\epsilon > 0$,

$$P(|S_n/n| > \epsilon \quad \text{infinitely often}) = 0. \tag{2}$$

We do so by showing that

$$\sum_{n=1}^{\infty} P(|S_n/n| > \epsilon) < \infty. \tag{3}$$

There are two remaining issues: (1) How do we establish this desired inequality in (3), and (ii) Why does that suffice to prove (2)?

To see why it suffices to prove (2), define the events

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

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 (2).

Now we turn to the second point: verifying (3). 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 $E[S_n^4] \leq K n^2$. (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 (3). (Note that this argument does not work if we use the second power instead of the fourth power.)

9 The Continuous Mapping Theorem

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

Theorem 9.1 *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.]