8. Convergence Concepts: in Probability, in L^p and Almost Surely

Instructor: Alessandro Rinaldo

Associated reading: Sec 2.4, 2.5, and 4.11 of Ash and Doléans-Dade; Sec 1.5 and 2.2 of Durrett.

1 Convergence in Probability and the Weak Law of Large Numbers

The Weak Law of Large Numbers is a statement about sums of independent random variables. Before we state the WLLN, it is necessary to define convergence in probability.

Definition 1 (Convergence in Probability). *We say Y_n* converges in probability *to Y and write $Y_n \stackrel{P}{\to} Y$ if, $\forall \epsilon > 0$,*

$$P(\omega : |Y_n(\omega) - Y(\omega)| > \epsilon) \to 0, \quad n \to \infty.$$

Theorem 2 (Weak Law of Large Numbers). *Let $X, X_1, X_2, \ldots$ be a sequence of independent, identically distributed (i.i.d.) random variables with $E|X| < \infty$ and define $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\frac{S_n}{n} \stackrel{P}{\to} EX.$$

The proof of WLLN makes use of the independent condition through the following basic lemma.

Lemma 3. *Let X_1 and X_2 be independent random variables. Let f_i $(i = 1, 2)$ be measurable functions such that $\mathrm{E}|f_i(X_i)| < \infty$ for $i = 1, 2$, then $\mathrm{E}f_1(X_1)f_2(X_2) = \mathrm{E}f_1(X_1)\mathrm{E}f_2(X_2)$.*

The proof of Lemma 3 follows from Lemma 34 of lecture notes set 4 and Fubini's Theorem. The following corollary will be used in our proof of WLLN.

Corollary 4. *If X_1 and X_2 are independent random variables, and $\mathrm{Var}(X_i) < \infty$, then $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$.*

Proof: [Proof of WLLN] In this proof, we employ the common strategy of first proving the result under an L^2 condition (i.e. assuming that the second moment is finite), and then using truncation to get rid of the extraneous moment condition.

First, we assume $\mathrm{E}X^2 < \infty$. Because the X_i are iid,

$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\mathrm{Var}(X)}{n}.$$

By Chebychev's inequality, $\forall\ \epsilon > 0$,

$$\Pr\left(\left|\frac{S_n}{n} - \mathrm{E}X\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\mathrm{Var}(X)}{n\epsilon^2} \to 0.$$

Thus, $\frac{S_n}{n} \xrightarrow{P} EX$ under the finite second moment condition. To transition from L^2 to L^1, we use truncation. For $0 < t < \infty$ let

$$\begin{aligned} X_{tk} &= X_k \mathbf{1}_{(|X_k| \leq t)} \\ Y_{tk} &= X_k \mathbf{1}_{(|X_k| > t)} \end{aligned}$$

Then, we have $X_k = X_{tk} + Y_{tk}$ and

$$\begin{aligned} \frac{S_n}{n} &= \frac{1}{n}\sum_{k=1}^{n} X_{tk} + \frac{1}{n}\sum_{k=1}^{n} Y_{tk} \\ &= U_{tn} + V_{tn} \end{aligned}$$

Because $|\sum_k Y_{tk}| \leq \sum_k |Y_{tk}|$, we have

$$\mathrm{E}\left|\frac{1}{n}\sum_{k=1}^{n} Y_{tk}\right| \leq \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}|Y_{tk}| = \mathrm{E}(|X|\mathbf{1}_{(|X|>t)})$$

and by DCT,

$$\mathrm{E}(|X|\mathbf{1}_{(|X|>t)}) \to 0, \qquad t \to \infty.$$

Fix $1 > \epsilon > 0$, for any $0 \leq \delta \leq 1$ we can choose t such that

$$\mathrm{E}\left(|X|\mathbf{1}_{(|X|>t)}\right) = \mathrm{E}|Y_{t1}| < \epsilon\delta/6.$$

Let $\mu_t = \mathrm{E}(X_{t1})$ and $\mu = \mathrm{E}(X)$. Because $0 \leq \delta \leq 1$, then we also have

$$|\mu_t - \mu| \leq |\mathrm{E}(Y_{t1})| < \epsilon\delta/6 < \epsilon/3.$$

Let $B_n = \{|U_{tn} - \mu_t| > \epsilon/3\}$ and $C_n = \{|V_{tn}| > \epsilon/3\}$. Noting that $\mathrm{E}(X_{tk}^2) \leq t^2 < \infty$, we can apply the Weak Law of Large Numbers to U_{tn}. Thus, we choose $N > 0$ such that $\forall\ n > N$,

$$\Pr(B_n) = \Pr(|U_{nt} - \mu_t| > \epsilon/3) < \delta/2.$$

Now, by Markov's inequality, we also have

$$\Pr(C_n) = \Pr(|V_{tn}| > \epsilon/3) \leq \frac{3\mathrm{E}|V_{tn}|}{\epsilon} \leq \frac{3\mathrm{E}|Y_{t1}|}{\epsilon} \leq \delta/2\,.$$

But on $B_n^c \cap C_n^c = (B_n \cup C_n)^c$, we have $|U_{tn} - \mu_t| \leq \epsilon/3$ and $|V_{tn}| \leq \epsilon/3$, and therefore

$$\left|\frac{S_n}{n} - \mu\right| \leq |U_{tn} - \mu_t| + |V_{tn}| + |\mu_t - \mu| \leq \epsilon/3 + \epsilon/3 + \epsilon/3 \leq \epsilon.$$

Thus, $\forall\ n > N$,

$$\Pr\left(\left|\frac{S_n}{n} - \mathrm{E}X\right| > \epsilon\right) \leq \Pr(B_n \cup C_n) \leq \delta.$$

■

2 Convergence of Random Variables: almost sure, in probability and in L_p

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Recall that a sequence $\{X_n\}_n$ of random variables converges to the random variable X (all defined on that same probability space) when

$$P\left(\left\{\omega\colon\ \lim_n X_n(\omega) = X(\omega)\right\}\right) = 1,$$

or, equivalently, when for each $\epsilon > 0$,

$$P\left(\{\omega\colon |X_n - X| > \epsilon \quad i.o.\}\right) = 0.$$

Recall that the previous statement can be expressed as

$$P\left(\limsup_n A_{n,\epsilon}\right) = P\left(\bigcap_n \bigcup_{k\geq n} A_{k,\epsilon}\right) = 0,$$

where $A_{n,\epsilon} = \{\omega\colon |X_n(\omega) - X(\omega)| > \epsilon\}$.

For $p \geq 1$, we say that X_n converges to X in L^p when

$$\lim_n \mathbb{E}\left[|X_n - X|^p\right] = 0.$$

Since the L^p norms are increasing in p, convergence in L^p implies convergence in L^r for $r < p$. In the previous section we introduced convergence in probability. We now discuss the relationship among different notions of convergence.

Fact: Convergence in L^p is different from convergence a.s.

Example 5. *Let* $\Omega = (0,1)$ *with* P *being Lebesgue measure. Consider the sequence of functions* 1, $I_{(0,1/2]}$, $I_{(1/2,1)}$, $I_{(0,1/3]}$, $I_{(1/3,2/3]}$, $\ldots$*. These functions converge to 0 in* L^p *for all finite* p *since the integrals of their absolute values go to 0. But they clearly don't converge to 0 a.s. since every* ω *has* $f_n(\omega) = 1$ *infinitely often. These functions are in* L^∞*, but they don't converge to 0 in* L^∞*. because their* L^∞ *norms are all 1.*

Example 6. *Let* $\Omega = (0,1)$ *with* P *being Lebesgue measure. Consider the sequence of functions*

$$f_n(\omega) = \begin{cases} 0 & \text{if } 0 < \omega < 1/n, \\ 1/\omega & \text{if } 1/n \leq \omega < 1. \end{cases}$$

Each f_n *is in* L^p *for all* p*, and* $\lim_{n\to\infty} f_n(\omega) = 1/\omega$ *a.s. But the limit function is not in* L^p *for even a single* p*. Clearly,* $\{f_n\}_{n=1}^\infty$ *does not converge in* L^p*.*

Example 7. *Let* $\Omega = (0,1)$ *with* P *being Lebesgue measure. Consider the sequence of functions*

$$f_n(\omega) = \begin{cases} n & \text{if } 0 < \omega < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Then f_n *converges to 0 a.s. but not in* L^p *since* $\int |f_n|^p dP = n^{p-1}$ *for all* n *and finite* p*. In this case, the a.e. limit is in* L^p*, but it is not an* L^p *limit.*

Oddly enough convergence in L^∞ does imply convergence a.e., the reason being that L^∞ convergence is "almost" uniform convergence.

Proposition 8. *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space. Then* f_n *converges to* f *in* L^∞ *if and only if there exists a measurable set* A *such that* $\mu(A^c) = 0$ *and* $\lim_{n\to\infty} f_n = f$*, uniformly on* A*.*

We can extend convergence in probability to convergence in measure.

Definition 9 (Convergence in Measure). *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and let* f *and* $\{f_n\}_{n=1}^\infty$ *be measurable functions that take values in a metric space with metric* d*. We say that* f_n *converges to* f in measure *if, for every* $\epsilon > 0$*,*

$$\lim_{n\to\infty} \mu(\{\omega : d(f_n(\omega), f(\omega)) > \epsilon\}) = 0.$$

When μ *is a probability, convergence in measure is called* convergence in probability*, denoted* $f_n \xrightarrow{P} f$*.*

Convergence in measure is different from a.e. convergence. Example 5 is a classic example of a sequence that converges in measure (in probability in that example) but not a.e. Here is an example of a.e. convergence without convergence in measure (only possible in infinite measure spaces).

Example 10. *Let $\Omega = I\!R$ with μ being Lebesgue measure. Let $f_n(x) = I_{[n,\infty)}(x)$ for all n. Then f_n converges to 0 a.e. $[\mu]$. However, f_n does not converge in measure to 0, because $\mu(\{|f_n| > \epsilon\}) = \infty$ for every n.*

Example 7 is an example of convergence in probability but not in L^p. Indeed convergence in probability is weaker than L^p convergence.

Proposition 11. *If $\|X_n - X\|_p \to 0$ in L^p for some $p > 0$, then $X_n \overset{P}{\to} X$.*

Convergence in probability is also weaker than converges a.s.

Lemma 12. *If $X_n \to X$ a.s., then $X_n \overset{P}{\to} X$.*

Proof: Let $\epsilon > 0$. Let $C = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$, and define $C_n = \{\omega : d(X_k(\omega), X(\omega)) < \epsilon$, for all $k \geq n\}$. Clearly, $C \subseteq \bigcup_{n=1}^{\infty} C_{n,\epsilon}$. Because $\Pr(C) = 1$ and $\{C_n\}_{n=1}^{\infty}$ is an increasing sequence of events, $\Pr(C_n) \to 1$. Because $\{\omega : d(X_n(\omega), X(\omega)) > \epsilon\} \subseteq C_n^C$,

$$\Pr(d(X_n, X) > \epsilon) \to 0.$$

■

A partial converse of this lemma is true and will be proved later.

Lemma 13. *If $X_n \overset{P}{\to} X$, then there is a subsequence $\{X_{n_k}\}_{k=1}^{\infty}$ such that $X_{n_k} \overset{\text{a.s.}}{\to} X$.*

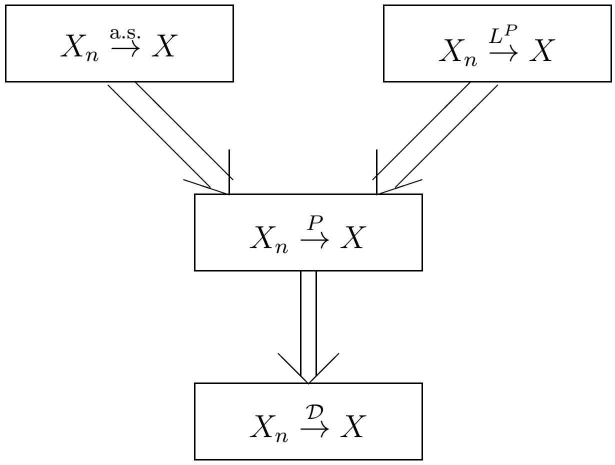
There is an even weaker form of convergence that we will discuss in detail later in the course.

Definition 14 (Convergence in Distribution). *A notion of convergence of a probability distribution on $\mathbb{R}$ (or more general space). We say $X_n \overset{\mathcal{D}}{\to} X$ if $\Pr(X_n \leq x) \to \Pr(X \leq x)$ for all x at which the RHS is continuous.*

Note that this is not really a notion of convergence of random variables, but the convergence of their distribution functions. This weak convergence appears in the central limit theorem.

Fact 15. $X_n \overset{\mathcal{D}}{\to} X \iff \mathrm{E}f(X_n) \to \mathrm{E}f(X)$ *for all bounded and continuous function f.*

The relationship between modes of convergence can be summarized as follows.


$X_n \xrightarrow{\text{a.s.}} X$
$X_n \xrightarrow{L^P} X$
$X_n \xrightarrow{P} X$
$X_n \xrightarrow{\mathcal{D}} X$