36710-36752 Advanced Probability Overview Fall 2020

5. L^p Spaces and Weak Law of Large Numbers

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 L^p Spaces

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. For $p \geq 1$, consider $\mathcal{L}^p$, the set of all functions f from Ω to $\mathbb{R}$ such that $|f|^p$ is integrable. This set has many properties of a metric space, but it has one problem that we shall see shortly. Let $\|f\|_p = \left(\int |f|^p d\mu\right)^{1/p}$. It is easy to see that $\|f\|_p \geq 0$ for all $f \in \mathcal{L}^p$. For $p \geq 1$, we will also show that $\|f+g\|_p \leq \|f\|_p + \|g\|_p$, whenever $f, g \in \mathcal{L}^p$. Also, if $f \in \mathcal{L}^p$ and $a \in \mathbb{R}$, then $af \in \mathcal{L}^p$ and $\|af\|_p = |a|\|f\|_p$, so $\mathcal{L}^p$ is a real vector space. The only problem is that $\|\cdot\|_p$ is not a true norm because $\|f\|_p = 0$ does not imply that $f = 0$ in the vector space. Every f that equals 0 a.e. $[\mu]$ will have $\|f\|_p = 0$. Hence, we modify $\mathcal{L}^p$ to create a new set L^p which consists of the equivalence classes of elements of $\mathcal{L}^p$ under the equivalence relation $f \sim g$ if and only if $f = g$ a.e. $[\mu]$. We define $\|[f]\|_p = \|f\|_p$, where $[f]$ stands for the equivalence class to which f belongs. Because $g \in [f]$ implies $\|g\|_p = \|f\|_p$, the norm is now well-defined on L^p. So, L^p is a normed linear space, hence a metric space. We will also show (later in this class) that it is complete, and hence a Banach space. People tend to ignore the fact that L^p is a set of equivalence classes rather than a set of functions, and we will do the same.

To be precise, we should include more information in the name of L^p. Indeed, we should refer to the space as $L^p(\Omega, \mathcal{F}, \mu)$, since each measure space has its own L^p space.

The special case of $p = \infty$ is handled separately. A function (equivalence class) f is in L^∞ if the function is essentially bounded.

Definition 1 (Essential Supremum). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f : \Omega \to \overline{\mathbb{R}}$ be a measurable function. Let $\mathcal{A}$ be the collection of all elements A of $\mathcal{F}$ such that $\mu(A^C) = 0$. The* essential supremum *of f is*

$$\operatorname{ess\,sup} f = \inf_{A \in \mathcal{A}} \sup_{\omega \in A} |f(\omega)|$$

or equivalently

$$\operatorname{ess\,sup} f = \sup\{t : \mu(\omega : |f(\omega)| \geq t) > 0\} = \inf\{t \geq 0 \colon \mu(\omega : |f(\omega)| > t) = 0\}.$$

If $\operatorname{ess\,sup} f < \infty$, we say that f is essentially bounded. *The L^∞ norm of f is $\|f\|_\infty = \operatorname{ess\,sup} f$.*

The essential supremum of f is the least upper bound of $|f|$ on sets whose complements have measure 0. It is the smallest number c such that $|f| \leq c$ a.e. $[\mu]$.

The special L^p space in which $\Omega = \mathbb{Z}^+$, $\mathcal{F} = 2^\Omega$, and μ is counting measure is called ℓ^p.

Example 2. *For each $1 \leq p < \infty$, $f \in \ell^p$ if and only if its pth power is an absolutely summable sequence. A sequence f is in ℓ^∞ if and only if it is bounded.*

For finite measure spaces, $L^{p_1} \subseteq L^{p_2}$ whenever $p_1 \geq p_2$. In fact, for probability soaces, $\|X\|_{p_1} \geq \|X\|_{p_2}$ if $p_1 \geq p_2$. This follows from Hölder's inequality.

Proposition 3 (Jensen's inequality). *If f is a convex function, then* $\mathrm{E}[f(X)] \geq f(\mathrm{E}(X))$.

For infinite measure spaces, odd behavior is possible.

Example 4. *Consider $L^p(\mathbb{R}, \mathcal{B}^1, \lambda)$, where λ is Lebesgue measure. Let*

$$f(x) = \begin{cases} x^{-3/8} & \text{for } 0 < x < 1, \\ x^{-1} & \text{for } 1 \leq x < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Then $f \notin L^1$, $f \in L^2$, and $f \notin L^3$.

We will prove most results for $L^p(\Omega, \mathcal{F}, \mu)$ where μ is a σ-finite measure. We will use the results mainly for probability spaces.

Lemma 5 (Linearity of L^p). *If $X, Y \in L^p$, then $X + Y \in L^p$.*

Proof: For all $p > 0$, we have

$$|a+b|^p \leq (|a| + |b|)^p \leq (2\max(|a|, |b|))^p = 2^p \max(|a|^p, |b|^p) \leq 2^p(|a|^p + |b|^p).$$

It follows that

$$\int |X+Y|^p d\mu \leq 2^p \left[\int |X|^p d\mu + \int |Y|^p d\mu \right].$$

■

Some of the results on L^p spaces involve the concept of conjugate indices.

Definition 6 (Conjugate Indices). *For each $p \in (1, \infty)$, let q be the unique value in $(1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. For $p = 1$, let $q = \infty$. Then p and q are* conjugate indices.

Theorem 7 (Hölder's inequality). *Let $X \in L^p$ and $Y \in L^q$, where p and q are conjugate indices. Then $XY \in L^1$ and $\int |XY| d\mu \leq \|X\|_p \|Y\|_q$.*

Proof: If either p or q is ∞, the result is obvious. So assume that neither p nor q is ∞. Define $U = |X|^p$ and $V = |Y|^q$, so that $U, V \in L^1$. Since the weighted geometric mean of two numbers is never more than the weighted arithmetic mean (with the same weights), we have

$$\left[\frac{U}{\int U d\mu}\right]^{1/p}\left[\frac{V}{\int V d\mu}\right]^{1/q} \leq \frac{1}{p}\frac{U}{\int U d\mu} + \frac{1}{q}\frac{V}{\int V d\mu}.$$

It follows that

$$\int U^{1/p}V^{1/q}d\mu \leq \left(\int U d\mu\right)^{1/p}\left(\int V d\mu\right)^{1/q} < \infty.$$

The first of these inequalities can be rewritten $\int |XY|d\mu \leq \|X\|_p\|Y\|_q$. The second one implies that $XY \in L^1$. ■

Example 8 (Cauchy-Schwarz inequality). *Let $p = q = 2$ in Theorem 7 to get that $X, Y \in L^2$ implies*

$$\int |XY|d\mu \leq \sqrt{\int X^2 d\mu \int Y^2 d\mu}.$$

If μ is a probability, this is the familiar Cauchy-Schwarz inequality.

Theorem 9 is the triangle inequality for L^p norms.

Theorem 9 (Minkowski's inequality). *If $X, Y \in L^p$, then*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Proof: The proofs are simple for $p = 1$ and $p = \infty$. So assume $p \in (1, \infty)$. First, let q be the conjugate index and notice that $(p-1)q = p$. Hence $|X+Y|^{p-1} \in L^q$ and $\||X+Y|^{p-1}\|_q = \|X+Y\|_p^{p/q}$. Write

$$|X+Y|^p = |X+Y||X+Y|^{p-1} \leq |X||X+Y|^{p-1} + |Y||X+Y|^{p-1}.$$

Theorem 7 says that

$$\int |X+Y|^p d\mu \leq (\|X\|_p + \|Y\|_p)\||X+Y|^{p-1}\|_q.$$

Rewrite this as

$$\|X+Y\|_p^p \leq (\|X\|_p + \|Y\|_p)\|X+Y\|_p^{p/q}.$$

Divide both sides by $\|X+Y\|_p^{p/q}$ to get the desired result, because $p - p/q = 1$. ■

We give two more simple but useful inequalities for integrals and expectations.

Proposition 10 (Markov inequality). *If f is a nonnegative measurable function, then $\mu(\{\omega : f(\omega) \geq c\}) \leq \int f d\mu/c$. In particular, let X be a nonnegative random variable. Then* $\Pr(X \geq c) \leq \mathrm{E}(X)/c$.

There is also a famous corollary.

Corollary 11 (Tchebychev inequality). *Let X be a random variable and have finite mean μ. Then* $\Pr(|X - \mu| \geq c) \leq \mathrm{Var}(X)/c^2$.

Exercise 12. *Show that there is some random variable X for which* $\Pr(|X - \mu| \geq c) = \mathrm{Var}(X)/c^2$*. Thus, without additional assumptions, Tchebychev's inequality cannot be improved.*

2 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 13 (Convergence in Probability). *We say Y_n* converges in probability *to Y and write $Y_n \xrightarrow{P} Y$ if,* $\forall \epsilon > 0$,

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

Theorem 14 (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*

$$\frac{S_n}{n} \xrightarrow{P} EX.$$

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

Lemma 15. *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 15 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 16. *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, $\forall\ \epsilon > 0$,

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

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 $\mathrm{E}(X_{tk}^2) \le t^2 < \infty$, 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

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

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

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

Thus, $\forall\ n > N$,

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

∎

3 Convergence of Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. We have already discussed convergence a.s., in the context of what a.s. means. Each L^p space has a sense of convergence. Since the L^p norms are increasing in p, convergence in L^p implies convergence in L^r for $r < p$. In WLLN we introduced convergence in probability. We now discuss the relationship between different notions of convergence.

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

Example 17. *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 18. *Let $\Omega = (0,1)$ with P being Lebesgue measure. Consider the sequence of functions*

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

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 19. *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 20. *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space. If* f_n *converges to* f *in* L^∞*, then* $\lim_{n\to\infty} f_n = f$*, a.e.* $[\mu]$.

We can extend convergence in probability to convergence in measure.

Definition 21 (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 17 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 22. *Let* $\Omega = \mathbb{R}$ *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 19 is an example of convergence in probability but not in L^p. Indeed convergence in probability is weaker than L^p convergence.

Proposition 23. *If* $\|X_n - X\|_p \to 0$ *in* L^p *for some* $p > 0$*, then* $X_n \xrightarrow{P} X$.

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

Lemma 24. *If* $X_n \to X$ *a.s., then* $X_n \xrightarrow{P} X$.

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

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

Definition 26 (Convergence in Distribution). *A notion of convergence of a probability distribution on $\mathbb{R}$ (or more general space). We say $X_n \xrightarrow{\mathcal{D}} X$ 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 27. $X_n \xrightarrow{\mathcal{D}} X \iff \mathrm{E}f(X_n) \to \mathrm{E}f(X)$ *for all bounded and continuous function f.*

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

$$\begin{tikzcd}
\boxed{X_n \xrightarrow{\text{a.s.}} X} \arrow[dr, Rightarrow] & & \boxed{X_n \xrightarrow{L^P} X} \arrow[dl, Rightarrow] \\
& \boxed{X_n \xrightarrow{P} X} \arrow[d, Rightarrow] & \\
& \boxed{X_n \xrightarrow{\mathcal{D}} X} &
\end{tikzcd}$$