Lecture-14: L^p convergence of random variables

1 L^p space

Definition 1.1 (L^p space). Consider a probability space $(\Omega, \mathcal{F}, P)$. For any $p > 1$, we say that a random variable $X \in L^p$, if $(\mathbb{E}|X|^p)^{\frac{1}{p}} < \infty$, and we can define a norm

$$\|X\|_p \triangleq (\mathbb{E}|X|^p)^{\frac{1}{p}}.$$

Theorem 1.2 (Minkowski's inequality). *Norm on the L^p satisfies the triangle inequality. That is, if $X, Y \in L^p$, then $X + Y \in L^p$ and*

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

Proof. We first show that $X + Y \in L^p$. To this end, it suffices to show that $\|X+Y\|_p$ is bounded if $\|X\|_p$ and $\|Y\|_p$ are bounded. From the convexity of $g(x) = |x|^p$ for $p \geqslant 1$, we get

$$|X+Y|^p \leqslant \frac{1}{2}|2X|^p + \frac{1}{2}|2Y|^p = 2^{p-1}(|X|^p + |Y|^p).$$

Taking expectation on both sides, it follows from the linearity of expectation, $\|X+Y\|_p \leqslant 2^{p-1}(\|X\|_p + \|Y\|_p)$.

We now show that $\|\|_p$ is a norm and satisfies the triangle inequality. From the triangle equality $|X+Y| \leqslant |X| + |Y|$ for two random variables $X, Y \in L^p$, the linearity of expectation, and the Hölder inequality for pair of random variables $X, (X+Y)^{p-1}$ and $Y, (X+Y)^{p-1}$, we get

$$\mathbb{E}|X+Y|^p \leqslant \mathbb{E}|X|\,|X+Y|^{p-1} + \mathbb{E}|Y|\,|X+Y|^{p-1} \leqslant ((\mathbb{E}|X|^p)^{\frac{1}{p}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}})(\mathbb{E}|X+Y|^p)^{1-\frac{1}{p}}.$$

□

Theorem 1.3. *For a probability space $(\Omega, \mathcal{F}, P)$ and $q \geqslant p \geqslant 1$, we have $L^q \subseteq L^p$.*

Proof. Consider $q \geqslant p \geqslant 1$, and a random variable $X \in L^q$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Applying Hölder's inequality to the product of random variables $|X|^p \cdot 1$ with conjugate variables $p' \triangleq \frac{q}{p} \geqslant 1$ and $q' \triangleq \frac{q}{q-p} \geqslant 1$, we get $\mathbb{E}|X|^p = \mathbb{E}[|X|^{\frac{q}{p'}} \cdot 1] \leqslant (\mathbb{E}|X|^q)^{\frac{1}{p'}}$. □

Example 1.4 (Mean square error). Consider a sequence of random variables $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ such that

$$m \triangleq \mathbb{E}X_n, \qquad \rho_k \triangleq \text{cov}(X_n X_{n+k}) \text{ for all } n, k \in \mathbb{N}.$$

The **best linear predictor** of X_{n+1} based on $X_1, \dots, X_n$ is given by $\hat{X}_{n+1} = \sum_{i=1}^n \alpha_i X_i$ for $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ such that the **mean square error** is minimized. Taking $\alpha_0 = -1$, we have

$$\mathbb{E}\left|X_{n+1} - \hat{X}_{n+1}\right|^2 = \min_{\alpha \in \mathbb{R}^n}\Big(\rho_0 + m^2 + \sum_{i=1}^n \alpha_i^2(\rho_0 + m^2) - 2\sum_{i=1}^n \alpha_i((m^2 + \rho_{n+1-i}) - \sum_{k=1}^{n-i} \alpha_{i+k}(m^2 + \rho_k))\Big).$$

Taking derivatives with respect to coefficients $\alpha \in \mathbb{R}^n$, we get

$$\alpha_i(\rho_0 + m^2) = m^2 + \rho_{n+1-i} - \sum_{k=1}^{n-i} \alpha_{i+k}(m^2 + \rho_k), \quad i \in [n].$$

2 L^p convergence

Definition 2.1 (Convergence in L^p). A sequence $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ of random variables **converges in** L^p to a random variable $X_\infty : \Omega \to \mathbb{R}$, if

$$\lim_n \mathbb{E}\,|X_n - X_\infty|^p = 0.$$

Proposition 2.2 (Convergences L^p implies in probability). *Consider a sequence of random variables $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ such that $\lim_n X_n = X_\infty$ in L^p, then $\lim_n X_n = X_\infty$ in probability.*

Proof. Let $\epsilon > 0$, then from the Markov's inequality applied to random variable $|X_n - X|^p$, we have

$$P\{|X_n - X_\infty| > \epsilon\} \leqslant \frac{\mathbb{E}\,|X_n - X_\infty|^p}{\epsilon}.$$

□

Example 2.3 (Convergence in probability doesn't imply in L^p). Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ such that $\lambda([a,b]) = b - a$ for all $0 \leqslant a \leqslant b \leqslant 1$. We define the scaled indicator random variable $X_n : \Omega \to \{0,1\}$ such that

$$X_n(\omega) = 2^n \mathbb{1}_{[0,\frac{1}{n}]}(\omega).$$

Then, $\lim_n X_n = 0$ in probability, since for any $1 > \epsilon > 0$, we have

$$P\{|X_n| > \epsilon\} = \frac{1}{n}.$$

However, we see that $\mathbb{E}\,|X_n|^p = \frac{2^{np}}{n}$.

Theorem 2.4 (L^2 weak law of large numbers). *Consider a sequence of uncorrelated random variables $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ such that $\mathbb{E}X_n = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$. Defining the sum $S_n \triangleq \sum_{i=1}^n X_i$ and the empirical mean $\bar{X}_n \triangleq \frac{S_n}{n}$, we have $\lim_n \bar{X}_n = \mu$ in L^2 and in probability.*

Proof. This follows from the fact that $\mathrm{Var}\,\bar{X}_n = \mathbb{E}(\bar{X}_n - \mu)^2 = \frac{1}{n^2}\mathbb{E}(S_n - n\mu)^2 = \frac{\sigma^2}{n}$. Convergence in L^p implies convergence in probability, and hence the result holds. □

Example 2.5 (Convergence in L^p doesn't imply almost surely). Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ such that $\lambda([a,b]) = b - a$ for all $0 \leqslant a \leqslant b \leqslant 1$. For each $k \in \mathbb{N}$, we consider the sequence $S_k = \sum_{i=1}^k i$, and define integer intervals $I_k \triangleq \{S_{k-1} + 1, \dots, S_k\}$. Clearly, the intervals $(I_k : k \in \mathbb{N})$ partition the natural numbers, and each $n \in \mathbb{N}$ lies in some I_k, such that $n = S_{k-1} + i$ for $i \in [k]$. Therefore, for each $n \in \mathbb{N}$, we define indicator random variable $X_n : \Omega \to \{0,1\}$ such that

$$X_n(\omega) = \mathbb{1}_{[\frac{i-1}{k},\frac{i}{k}]}(\omega).$$

For any $\omega \in [0,1]$, we have $X_n(\omega) = 1$ for infinitely many values since there exist infinitely many (i,k) pairs such that $\frac{(i-1)}{k} \leqslant \omega \leqslant \frac{i}{k}$, and hence $\limsup_n X_n(\omega) = 1$ and hence $\lim_n X_n(\omega) \neq 0$. However, $\lim_n X_n(\omega) = 0$ in L^p, since

$$\mathbb{E}\,|X_n|^p = \lambda\{X_n(\omega) \neq 0\} = \frac{1}{k_n}.$$

3 L^1 convergence theorems

Theorem 3.1 (Monotone Convergence Theorem). *Consider a probability space $(\Omega, \mathcal{F}, P)$ and a non-decreasing non-negative random sequence $X : \Omega \to \mathbb{R}_+^{\mathbb{N}}$ such that $X_n \in L^1$ for all $n \in \mathbb{N}$. Let $X_\infty(\omega) = \sup_n X_n(\omega)$ for all $\omega \in \Omega$, then*

$$\mathbb{E}X_\infty = \sup_n \mathbb{E}X_n.$$

Proof. From the monotonicity of sequence X and the monotonicity of expectation, we have $\sup_n \mathbb{E}X_n \leqslant \mathbb{E}X_\infty$. Let $\alpha \in (0,1)$ and $Y : \Omega \to \mathbb{R}_+$ a non-negative simple random variable such that $Y \leqslant X_\infty$. We define

$$E_n \triangleq \{\omega \in \Omega : X_n(\omega) \geqslant \alpha Y\} \in \mathcal{F}.$$

From the monotonicity of sequence X, the sequence of events $(E_n \in \mathcal{F} : n \in \mathbb{N})$ are monotonically non-decreasing such that $\cup_{n \in \mathbb{N}} E_n = \Omega$. It follows that

$$\alpha \mathbb{E}[Y \mathbb{1}_{E_n}] \leqslant \mathbb{E}[X_n \mathbb{1}_{E_n}] \leqslant \mathbb{E}X_n.$$

We will use the fact that $\lim_n \mathbb{E}[Y \mathbb{1}_{E_n}] = \mathbb{E}[Y]$, then $\alpha \mathbb{E}Y \leqslant \sup_n \mathbb{E}X_n$. Taking supremum over all $\alpha \in (0,1)$ and all simple functions $Y \leqslant X_\infty$, we get $\mathbb{E}X_\infty \leqslant \sup_n \mathbb{E}X_n$. □

Theorem 3.2 (Fatou's Lemma). *Consider a probability space $(\Omega, \mathcal{F}, P)$ and a non-negative random sequence $X : \Omega \to \mathbb{R}_+^{\mathbb{N}}$. Let $X_\infty(\omega) \triangleq \liminf_n X_n(\omega)$ for all $\omega \in \Omega$, then*

$$\mathbb{E}X_\infty \leqslant \liminf_n \mathbb{E}X_n.$$

Proof. We define $Y_n \triangleq \inf_{k \geqslant n} X_k$ for all $n \in \mathbb{N}$. It follows that $Y : \Omega \to \mathbb{R}_+^{\mathbb{N}}$ is a non-negative non-decreasing sequence of random variables, and $X_\infty = \sup_n Y_n = \lim_n Y_n$. By Motonone convergence theorem applies to Y, we have $\mathbb{E}X_\infty = \sup_n \mathbb{E}Y_n$. The result follows from the monotonicity of expectation, and the fact that $Y_n \leqslant X_k$ for all $k \geqslant n$, to get $\mathbb{E}Y_n \leqslant \inf_{k \geqslant n} \mathbb{E}X_k$. □

Theorem 3.3 (Dominated Convergence Theorem). *Let $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ be a random sequence defined on a probability space $(\Omega, \mathcal{F}, P)$. If $\lim_n X_n = X_\infty$ a.s. and there exists a $Y : \Omega \to \mathbb{R}_+$ such that $Y \in L^1(\mathcal{F})$ and $|X_n| \leqslant Y$ a.s., then $\mathbb{E}X_\infty = \lim_n \mathbb{E}X_n$.*

Proof. From the hypothesis, we have $Y + X_n \geqslant 0$ a.s. and $Y - X_n \geqslant 0$ a.s. Therefore, from Fatou's Lemma and linearity of expectation, we have

$$\mathbb{E}Y + \mathbb{E}X_\infty \leqslant \liminf_n \mathbb{E}(Y + X_n) = \mathbb{E}Y + \liminf_n \mathbb{E}X_n, \quad \mathbb{E}Y - \mathbb{E}X_\infty \leqslant \liminf_n \mathbb{E}(Y - X_n) = \mathbb{E}Y - \limsup_n \mathbb{E}X_n.$$

Therefore, we have $\limsup \mathbb{E}X_n \leqslant \mathbb{E}X_\infty \leqslant \liminf \mathbb{E}X_n$, and the result follows. □

4 Uniform integrability

Definition 4.1 (uniform integrability). A family $(X_t \in L^1 : t \in T)$ of random variables indexed by T is **uniformly integrable** if

$$\lim_{a \to \infty} \sup_{t \in T} \mathbb{E}[|X_t| \mathbb{1}_{\{|X_t| > a\}}] = 0.$$

Example 4.2 (Single element family). If $|T| = 1$, then the family is uniformly integrable, since $X_1 \in L^1$ and $\lim_a \mathbb{E}[|X_1| \mathbb{1}_{\{|X_t| > a\}}] = 0$. This is due to the fact that $(X_n \triangleq |X| \mathbb{1}_{\{|X| \leqslant n\}} : n \in \mathbb{N})$ is a sequence of increasing random variables $\lim_n X_n = X$. From monotone convergence theorem, we get $\lim_n \mathbb{E}|X_n| = \mathbb{E}\lim_n |X_n|$. Therefore,

$$\lim_a \mathbb{E}[|X| \mathbb{1}_{\{|X| > a\}}] = \mathbb{E}|X| - \lim_a \mathbb{E}[|X| \mathbb{1}_{\{|X| \leqslant a\}}] = 0.$$

Proposition 4.3. *Let $X \in L^p$ and $(A_n : n \in \mathbb{N}) \subset \mathcal{F}$ be a sequence of events such that $\lim_n P(A_n) = 0$, then*

$$\lim_n \left\| |X| \mathbb{1}_{A_n} \right\|_p = 0.$$

Example 4.4 (Dominated family). If there exists $Y \in L^1$ such that $\sup_{t \in T} |X_t| \leqslant |Y|$, then the family of random variables $(X_t : t \in T)$ is uniformly integrable. This is due to the fact that

$$\sup_{t \in T} \mathbb{E}[|X| \mathbb{1}_{\{|X|>a\}}] \leqslant \mathbb{E}[|Y| \mathbb{1}_{\{|Y|>a\}}].$$

Example 4.5 (Finite family). then the family of random variables $(X_t : t \in T)$ is uniformly integrable. This is due to the fact that $\sup_{t \in T} |X_t| \leqslant \sum_{t \in T} |X_t| \in L^1$.

Theorem 4.6 (Convergence in probability with uniform integrability implies in L^p). *Consider a sequence of random variables $(X_n : n \in \mathbb{N}) \subset L^p$ for $p \geqslant 1$. Then the following are equivalent.*

(a) The sequence $(X_n : n \in \mathbb{N})$ converges in L^p, i.e. $\lim_n \mathbb{E}|X_n - X|^p = 0$.

(b) The sequence $(X_n : n \in \mathbb{N})$ is Cauchy in L^p, i.e. $\lim_{m,n\to\infty} \mathbb{E}|X_n - X_m|^p = 0$.

(c) $\lim_n X_n = X$ in probability and the sequence $(|X_n|^p : n \in \mathbb{N})$ is uniformly integrable.

Proof. For a random sequence $(X_n : n \in \mathbb{N})$ in L^p, we will show that $(a) \implies (b) \implies (c) \implies (a)$.

$(a) \implies (b)$: We assume the sequence $(X_n : n \in \mathbb{N})$ converges in L^p. Then, from Minkowski's inequality, we can write

$$(\mathbb{E}|X_n - X_m|^p)^{\frac{1}{p}} \leqslant (\mathbb{E}|X_n - X|^p)^{\frac{1}{p}} + (\mathbb{E}|X_m - X|^p)^{\frac{1}{p}}.$$

$(b) \implies (c)$: We assume that the sequence $(X_n : n \in \mathbb{N})$ is Cauchy in L^p, i.e. $\lim_{m,n\to\infty} \mathbb{E}|X_n - X_m|^p = 0$. Let $\epsilon > 0$, then for each $n \in \mathbb{N}$, there exists N_ϵ such that for all $n, m \geqslant N_\epsilon$

$$\mathbb{E}|X_n - X_m|^p \leqslant \frac{\epsilon}{2}.$$

Let $A_a = \{\omega \in A : |X_n| > a\}$. Then, using triangle inequality and the fact that $\mathbb{1}_{A_a} \leqslant 1$, from the linearity and monotonicity of expectation, we can write for $n \geqslant N_\epsilon$

$$(\mathbb{E}[|X_n|^p \mathbb{1}_{\{|X_n|>a\}}])^{\frac{1}{p}} \leqslant (\mathbb{E}[|X_{N_\epsilon}|^p \mathbb{1}_{A_a}])^{\frac{1}{p}} + (\mathbb{E}[|X_n - X_{N_\epsilon}|^p])^{\frac{1}{p}} \leqslant (\mathbb{E}[|X_{N_\epsilon}|^p \mathbb{1}_{A_a}])^{\frac{1}{p}} + \frac{\epsilon}{2}.$$

Therefore, we can write $\sup_n \mathbb{E}[|X_n|^p \mathbb{1}_{\{|X_n|>a\}}] \leqslant \sup_{m \leqslant N_\epsilon} \mathbb{E}[|X_m|^p \mathbb{1}_{A_a}] + \frac{\epsilon}{2}$. Since $(|X_n|^p : n \leqslant N_\epsilon)$ is finite family of random variables in L^1, it is uniformly integrable. Therefore, there exists $a_\epsilon \in \mathbb{R}_+$ such that $\sup_{m \leqslant N_\epsilon} (\mathbb{E}[|X_m|^p \mathbb{1}_{A_a}])^{\frac{1}{p}} < \frac{\epsilon}{2}$. Taking $a' = \max\{a, a_\epsilon\}$, we get $\sup_n (\mathbb{E}[|X_n|^p \mathbb{1}_{\{|X_n|>a'\}}])^{\frac{1}{p}} \leqslant \epsilon$. Since the choice of ϵ was arbitrary, it follows that

$$\lim_{a\to\infty} \sup_n (\mathbb{E}[|X_n|^p \mathbb{1}_{\{|X_n|>a'\}}])^{\frac{1}{p}} = 0.$$

The convergence in probability follows from the Markov inequality, i.e.

$$P\left\{|X_n - X_m|^p > \epsilon\right\} \leqslant \frac{1}{\epsilon} \mathbb{E}|X_n - X_m|^p.$$

$(c) \implies (a):$ Since the sequence $(X_n : n \in \mathbb{N})$ is convergent in probability to a random variable X, there exists a subsequence $(n_k : k \in \mathbb{N}) \subset \mathbb{N}$ such that $\lim_k X_{n_k} = X$ a.s. Since $(|X_n|^p : n \in \mathbb{N})$ is a family of uniformly integrable sequence, by Fatou's Lemma

$$\mathbb{E}|X|^p \leqslant \liminf_k \mathbb{E}\left|X_{n_k}\right|^p \leqslant \sup_n \mathbb{E}|X_n|^p < \infty.$$

Therefore, $X \in L^1$, and we define $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ for any $\epsilon > 0$. From Minkowski's inequality, we get

$$\|X_n - X\|_p \leqslant \left\|(X_n - X)\mathbb{1}_{\{|X_n - X|^p \leqslant \epsilon\}}\right\|_p + \left\|X_n \mathbb{1}_{A_n(\epsilon)}\right\|_p + \left\|X \mathbb{1}_{A_n(\epsilon)}\right\|_p.$$

We can check that $\left\|(X_n - X)\mathbb{1}_{A_n^c(\epsilon)}\right\|_p \leqslant \epsilon$. Further, since $\lim_n X_n = X$ in probability, $(A_n : n \in \mathbb{N}) \subset \mathcal{F}$ is decreasing sequence of events, and since $X_n, X \in L^1$, we have $\lim_n \left\|X_n \mathbb{1}_{A_n(\epsilon)}\right\| = \lim_n \left\|X \mathbb{1}_{A_n(\epsilon)}\right\| = 0.$

□