Lecture-15: $L^p$ convergence of random variables

1 $L^p$ convergence

**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 $\mathbb{E}|X|^p < \infty$, and we can define a norm

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

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

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

*Proof.* From the triangle equality $|X + Y| \leq |X| + |Y|$ 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

$$\mathbb{E}|X + Y|^p \leq \mathbb{E}|X|^p|X + Y|^{p-1} + \mathbb{E}|Y|^p|X + Y|^{p-1} \leq \left((\mathbb{E}|X|^p)^{\frac{1}{p}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}}\right)(\mathbb{E}|X + Y|^p)^{\frac{1}{p}}.$$  

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 $p \geq 1$, we get

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

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

**Example 1.3 ($L^q \subseteq L^p$ for $q \geq p \geq 1$).** Consider $q \geq p \geq 1$, 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 $p' \triangleq \frac{q}{p} \geq 1$ and $q' \triangleq \frac{q}{q-p} \geq 1$, we get the result

$$\mathbb{E}|X|^p = \mathbb{E}(|X|^{p' \cdot 1}) \leq (\mathbb{E}|X|^q)^{\frac{p}{p'}}.$$  

**Definition 1.4 (Convergence in $L^p$).** A sequence $(X_n : n \in \mathbb{N})$ of random variables converges in $L^p$ to a random variable $X$, if

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

**Example 1.5 (Mean square error).** Consider a sequence of random variables $(X_n : n \in \mathbb{N})$ such that

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

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

$$\mathbb{E}|X_{n+1} - \hat{X}_{n+1}|^2 = \min_{\alpha \in \mathbb{R}^n} \mathbb{E}(X_{n+1} - \sum_{i=1}^n \alpha_i X_i)^2.$$
This is achieved for the coefficients \( \alpha \in \mathbb{R}^n \), such that
\[
\alpha_i \mathbb{E}[X_i^2] = \mathbb{E}[X_i X_{n+1}], \quad i \in [n].
\]
That is, \( \alpha_i = \frac{\theta_{i+1} - \theta_{i} + m^2}{\rho_{i} + m^2} \) for all \( i \in [n] \).

**Proposition 1.6 (Convergences \( L^p \) implies in probability).** Consider a sequence of random variables \( (X_n : n \in \mathbb{N}) \) such that \( \lim_{n \to \infty} X_n = X \) in \( L^p \), then \( \lim_{n \to \infty} X_n = X \) in probability.

**Proof.** Let \( \epsilon > 0 \), then from the Markov’s inequality applied to random variable \( |X_n - X|^p \), we have
\[
P \{ |X_n - X| > \epsilon \} \leq \frac{\mathbb{E} |X_n - X|^p}{\epsilon^p}.
\]

**Example 1.7 (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 \( 0 \leq a \leq b \leq 1 \). We define the scaled indicator random variable \( X_n : \Omega \to \{0,1\} \) such that
\[
X_n(\omega) = 2^n \mathbb{I}_{[0,\frac{1}{2}]}(\omega).
\]
Then, \( \lim_{n \to \infty} X_n = 0 \) in probability, since for any \( 1 > \epsilon > 0 \), we have
\[
P \{ |X_n| > \epsilon \} = \frac{1}{n}.
\]
However, we see that \( \mathbb{E} |X_n|^p = \frac{2^n}{n} \).

**Theorem 1.8 (\( L^2 \) weak law of large numbers).** Consider a sequence of uncorrelated random variables \( (X_n : n \in \mathbb{N}) \) such that \( \mathbb{E} X_n = \mu \) and \( \text{Var}(X_n) = \sigma^2 \). 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 \( \lim_{n \to \infty} \bar{X}_n = \mu \) in \( L^2 \) and in probability.

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

**Example 1.9 (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 \( 0 \leq a \leq b \leq 1 \). For each \( k \in \mathbb{N} \), we consider the sequence \( S_k = \sum_{i=1}^{k} \mathbb{I}_i \), and define integer intervals \( I_k \triangleq \{ S_{k-1} + 1, \ldots, S_k \} \). 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
\[
X_n(\omega) = \mathbb{I}_{[\frac{i-1}{k}, \frac{i}{k}]}(\omega).
\]
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 \( \frac{i-1}{k} \leq \omega \leq \frac{i}{k} \), 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
\[
\mathbb{E}|X_n|^p = \lambda\{X_n(\omega) \neq 0\} = \frac{1}{k^n}.
\]
2 Uniform integrability

Definition 2.1 (uniform integrability). A family \((X_t : t \in T)\) 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 2.2 (Single element family). If \(|T| = 1\), then the family is uniformly integrable, since \(X_t \in L^1\) and \(\lim_a \mathbb{E}[|X_t| \mathbb{1}_{(|X_t| > a)}] = 0\). This is due to the fact that \((X_n) = |X| \mathbb{1}_{(|X| \leq n)} : n \in \mathbb{N}\) 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,
\[
\lim_a \mathbb{E}[|X| \mathbb{1}_{(|X| > a)}] = \mathbb{E}|X| - \lim_a \mathbb{E}[|X| \mathbb{1}_{(|X| \leq a)}] = 0.
\]

Proposition 2.3. 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
\[
\lim_n \|X \mathbb{1}_{A_n}\|_p = 0.
\]

Example 2.4 (Dominated family). If there exists \(Y \in L^1\) such that \(\sup_{t \in T} |X_t| \leq |Y|\), then the family of random variables \((X_t : t \in T)\) is uniformly integrable. This is due to the fact that
\[
\sup_{t \in T} \mathbb{E}[|X_t| \mathbb{1}_{(|X_t| > a)}] \leq \mathbb{E}[|Y| \mathbb{1}_{(|Y| > a)}].
\]

Example 2.5 (Finite family). then the family of random variables \((X_t : t \in T)\) is uniformly integrable. This is due to the fact that \(\sup_{t \in T} |X_t| \leq \sum_{t \in T} |X_t| \in L^1\).

Theorem 2.6 (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 \(p \geq 1\). 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
\[
(\mathbb{E}|X_n - X_m|^p)^{\frac{1}{p}} \leq (\mathbb{E}|X_n - X|^p)^{\frac{1}{p}} + (\mathbb{E}|X_m - X|^p)^{\frac{1}{p}}.
\]

\((b) \implies (c)\): 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 \(N_\epsilon\) such that for all \(n, m \geq N_\epsilon\)
\[
\mathbb{E}|X_n - X_m|^p \leq \frac{\epsilon}{2}.
\]
Let $A_a = \{ \omega \in A : |X_n| > a \}$. Then, using triangle inequality and the fact that $\mathbb{1}_{A_a} \leq 1$, from the linearity and monotonicity of expectation, we can write for $n \geq N_\varepsilon$

$$
(\mathbb{E}[|X_n|^p \mathbb{1}_{\{|X_n| > a\}}])^{\frac{1}{p}} \leq \left( (\mathbb{E}[|X_n|_p \mathbb{1}_{A_a}])^{\frac{1}{p}} + (\mathbb{E}[|X_n - X_n|_p])^{\frac{1}{p}} \right) + \mathbb{E}[|X_n|_p \mathbb{1}_{A_a}] \geq (\mathbb{E}[|X_n|_p \mathbb{1}_{A_a}] + \mathbb{E}[|X_n|_p \mathbb{1}_{A_a}])^{\frac{1}{p}} + \mathbb{E}[|X_n|_p \mathbb{1}_{A_a}] + \varepsilon \cdot \frac{1}{2}.
$$

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

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

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

$$
P\{ |X_n - X_m|^p > \varepsilon \} \leq \frac{1}{\varepsilon} \mathbb{E}|X_n - X_m|^p.
$$

$(c) \implies (a)$: Since the sequence $(X_n : n \in \mathbb{N})$ is convergent in probability to a random variable $X$, there exists a subsequence $(X_{n_k} : k \in \mathbb{N}) \subset \mathbb{N}$ 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

$$
\mathbb{E}|X|^p \leq \liminf_{k} \mathbb{E}|X_{n_k}|^p \leq \sup_{n} \mathbb{E}|X_n|^p < \infty.
$$

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

$$
\|X_n - X\|_p \leq \left( \|X_n - X\|_p \mathbb{1}_{\{|X_n - X| \leq \varepsilon\}} + \|X_n \mathbb{1}_{A_n(\varepsilon)}\|_p + \|X \mathbb{1}_{A_n(\varepsilon)}\|_p \right).
$$

We can check that $\left( \|X_n - X\|_p \mathbb{1}_{A_n(\varepsilon)}\|_p \right) \leq \varepsilon$. 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 $\lim_n \|X_n \mathbb{1}_{A_n(\varepsilon)}\| = \lim_n \|X \mathbb{1}_{A_n(\varepsilon)}\| = 0$. 

\[\square\]