2. Convergence with probability one, and in probability. Other types of convergence. Ergodic theorem

2.1. **Definitions.** In this Lecture, we consider different type of convergence for a sequence of random variables $X_n, n \geq 1$. Since $X_n = X_n(\omega)$, we may consider the convergence for fixed ω°: $X_n(\omega^\circ) \to \xi(\omega^\circ),\ n \to \infty$. That type of convergence might be not valid for all $\omega \in \Omega$.

Convergence with probability one. Let $A = \{\omega : \lim_n X_n(\omega) = X(\omega)\}$ and $A^c = \Omega \setminus A$. If $P(A^c) = 0$, a sequence $X_n(\omega), n \geq 1$ is said to be obey the limit $X(\omega)$ with probability one: $X_n \xrightarrow[n\to\infty]{P-a.s.} X$.

Obviously, the convergence for all ω's provides the convergence with probability one. Let $X_n(\omega) = (1 - \frac{1}{n})X(\omega)$. Then $X_n(\omega)$ converges to $X(\omega)$ for all ω's and so with probability one.

Convergence in probability. A sequence $X_n(\omega), n \geq 1$ of random variables converges in probability to a random variable $X(\omega)$, if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\big(\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\big) = 0.$$

Convergence in L^1-norm. A sequence $X_n(\omega), n \geq 1$ of random variables, with $\boldsymbol{E}|\xi_n| < \infty,\ n \geq 1$, converges in L^1-norm to a random variable $X(\omega)$, with $\boldsymbol{E}|X(\omega)| < \infty$, if

$$\lim_{n\to\infty} \boldsymbol{E}|X_n - X| = 0.$$

Convergence in L^2-norm. A sequence $X_n(\omega), n \geq 1$ of random variables, with $\boldsymbol{E}X_n^2 < \infty,\ n \geq 1$, converges in the L^2-n0rm to a random variable $X(\omega)$ with $\boldsymbol{E}X^2(\omega) < \infty$, if

$$\lim_{n\to\infty} \boldsymbol{E}\big(X_n - X\big)^2 = 0.$$

Convergence in distribution (law). A sequence $X_n(\omega),\ n \geq 1$ of random variables converges in distribution to a random variable $X(\omega)$, if for any $x \in \mathbb{R}$, being the point of continuity of the distribution function $F(x) = P(X(\omega) \leq x)$,

$$\lim_{n\to\infty} P(X_n(\omega) \leq x) = P(X(\omega) \leq x);$$

or, equivalently, for any bounded and continuous function f

$$\lim_{n\to\infty} \boldsymbol{E}f(X_n) = \boldsymbol{E}f(X).$$

2.2. **List of relationships for different type of convergence.**

1. Convergence with probability one provides convergence in probability and in law.

Proof. Define the indicator function $I(a \geq b) = \begin{cases} 1 & a \geq b, \\ 0 & \text{otherwise.} \end{cases}$ If $X_n \xrightarrow{P-\text{a.s.}} X$, then $|X_n - X|$ converges to zero with probability one and $I(|X_n - X| \geq \varepsilon) \to 0$ as well. Taking the expectation we find

$$\boldsymbol{E} I(|X_n - X| \geq \varepsilon) = P(|X_n - X| \geq \varepsilon) \to 0.$$

The convergence in law follows due the definition. The convergence in L^1 and L^2 norm is failed when first or second moments of X_n do not exist.

□

2. Convergence in probability provides convergence in law only.

Proof. We apply here the known fact. If ξ_n, $n \geq 1$ converges in probability to ξ, then for any bounded and continuous function f we have

$$\lim_{n \to \infty} \boldsymbol{E} f(\xi_n) = \boldsymbol{E}(\xi).$$

To convince ourselves that the convergence in probability does not provide the convergence with probability one, we consider the following example. Let $X_n, n \geq 1$, be a sequence of independent random variables such that $X_n = 1$ with probability $1/n$ and $X_n = 0$ with probability $1 - (1/n)$. That sequence converges to zero in probability since for any $\varepsilon > 0$

$$P(X_n \geq \varepsilon) \leq P(X_n = 1) = (1/n) \to 0, \ n \to \infty.$$

We show that at the same time this sequence does not converges to zero with probability one. To establish that fact, we assume that the convergence with probability one holds true and then obtain the contradiction. If the convergence with probability one holds true, then for large number m

$$P\big(\max_{n \geq m} X_n = 1\big) \to 0, \ m \to \infty.$$

Notice that $\big\{\max_{n \geq m} X_n = 1\big\} = \cup_{n \geq m}\big\{X_n = 1\big\}$. Hence, taking into consideration the fact that $(X_n)_{n \geq 1}$ is the sequence of independent random variables and

$$P\big(\cup_{n \geq m}\{X_n = 1\}\big) = 1 - P\big(\cap_{n \geq m}\{X_n = 0\}\big)$$
$$= 1 - \prod_{n \geq m} P\big(X_n = 0\big) = 1 - \prod_{n \geq m}\big(1 - \frac{1}{n}\big) \equiv 1,$$

we arrive at announced contradiction. □

3. Convergence on L^1-norm provides convergence in probability and in law.

Proof. It follows by he Chebyshev inequality:

$$P\Big(|X_n - X| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon}\boldsymbol{E}|X_n - X|.$$

□

4. Convergence on L^2-norm provides convergence in L^1-norm, in probability and in law.

Proof. The use of the Cauchy-Schwartz inequality

$$\boldsymbol{E}|X_n - X| \leq \sqrt{\boldsymbol{E}|X_n - X|^2}$$

and the Chebyshev inequality that

$$P\Big(|X_n - X| \geq \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\boldsymbol{E}|X_n - X|^2$$

provides the result.

□

5. Convergence in law does not provide convergence of any other types.

Proof. The convergence in law does not guarantee that X_n and X are "comparable", that is $X_n - X$ is well defined. □

Different types of convergence have different types of limits for the same sequence $(X_n)_{n\geq 1}$. Assume $(X_n)_{n\geq 1}$ converges simultaneously P-a.s., in probability, and in L^1- , L^2-norms. Denote for a definiteness by X, X^p, X', and X'' the limits respectively. A remarkable fact is given below.

Lemma 2.1.

$$P\big(X = X^p = X' = X''\big) = 1.$$

Proof. It suffices to show that $P(X \neq X^p) = 0$. Write

$$P(X \neq X^p) = P(|X - X^p| > 0)$$

and notice that $P(|X - X^p| > 0) = \lim_{\varepsilon\to 0} P(|X - X^p| > \varepsilon)$. On the other hand, since $|X_n - X| \xrightarrow[n\to\infty]{P-a.s.} 0$, also $I(|X_n - X| > \frac{\varepsilon}{2}) \xrightarrow[n\to\infty]{P-a.s.} 0$ and so, by taking the expectation we, find that $P(|X_n - X| > \frac{\varepsilon}{2}) \to 0$. The use now the triangular inequality $|X - X^p| \leq |X - X_n| + |X_n - X^p|$ for any $\varepsilon > 0$ provides

$$P\big(|X - X^p| > \varepsilon\big) \leq P\Big(|X - X_n| > \frac{\varepsilon}{2}\Big) + P\Big(|X_n - X^p| > \frac{\varepsilon}{2}\Big) \xrightarrow[n\to\infty]{} 0$$

and, thus, $\lim_{\varepsilon\to 0} P(|X - X^p| > \varepsilon) = 0$. □

Theorem 2.2. [Cauchy criteria] *A sequence of random variables* $(X_n)_{n\geq 1}$ *(*$\boldsymbol{E}X_n^2 < \infty$*) converges to a random variable* X *(*$\boldsymbol{E}X^2 < \infty$*) in* L^2*-norm, iff*

$$\lim_{n,m\to\infty} \boldsymbol{E}(X_n - X_m)^2 = 0. \tag{2.1}$$

Proof. The implication that L^2-norm convergence provides (2.1) is obvious by the inequality $\boldsymbol{E}(X_n - X_m)^2 \leq 2\big(\boldsymbol{E}(X_n - X)^2 + \boldsymbol{E}(X_m - X)^2\big)$.

To establish the inverse implication, we choose a subsequence n_k such that $\boldsymbol{E}(X_{n_{k+1}} - X_{n_k})^2 \leq (1/2^{2k})$. Since by the Cauchy-Schwarz inequality we have $\boldsymbol{E}|X_{n_{k+1}} - X_{n_k}| \leq (1/2^k)$ it holds

$$\sum_k \boldsymbol{E}|X_{n_{k+1}} - X_{n_k}| < \infty.$$

Hence $\sum\limits_k |X_{n_{k+1}} - X_{n_k}| < \infty$ with probability one. Therefore the random variable $X = X_{n_1} + \sum\limits_k \big(X_{n_{k+1}} - X_{n_k}\big)$ is well defined and is a candidate to be the limit, in L^2-norm, for the sequence $(X_n)_{n\geq 1}$. Write

$$(X_n - X)^2 \leq 2(X_n - X_{n_j})^2 + 2(X_{n_j} - X)^2$$

and notice that $\lim_{n,n_j} \boldsymbol{E}(X_n - X_{n_j})^2 = 0$ by (2.1). Further, we have that

$$(X_{n_j} - X)^2 = \Big(\sum_{k=j+1}^{\infty} \big(X_{n_{k+1}} - X_{n_k}\big)\Big)^2$$

and so

$$\begin{aligned}
\boldsymbol{E}(X_{n_j} - X)^2 &\leq \sum_{k=j+1,\ell=i+1}^{\infty} \boldsymbol{E}\big|X_{n_{k+1}} - X_{n_k}\big|\big|X_{n_{k+1}} - X_{n_k}\big| \\
&\leq \sum_{k=j+1,\ell=i+1}^{\infty} \boldsymbol{E}\big|X_{n_{k+1}} - X_{n_k}\big|^2 \boldsymbol{E}\big|X_{n_{k+1}} - X_{n_k}\big|^2 \\
&\leq \Big(\sum_{k=j+1}^{\infty} \frac{1}{2^{j+1}}\Big)^2 \to 0,\ k \to \infty.
\end{aligned}$$

□

Corollary 2.3. *Let* $\xi_n, n \geq 1$ *be a sequence of random variables with* $\boldsymbol{E}\xi_n \equiv 0$*,* $\boldsymbol{E}\xi_n^2 = \sigma_n^2$*, and* $\boldsymbol{E}\xi_n\xi_m = 0, n \neq m$*. The series* $\sum\limits_n \xi_n$ *converges in* L^2*-norm, iff* $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$*.*

Lemma 2.4. *If* $\lim_{n\to\infty} \boldsymbol{E}(\xi_n - \xi)^2 = 0$*, then* $\lim_{n\to\infty} \boldsymbol{E}\xi_n = \boldsymbol{E}\xi$*,* $\lim_{n\to\infty} \boldsymbol{E}\xi_n^2 = \boldsymbol{E}\xi^2$*.*

Proof. Since $|\boldsymbol{E}\xi_n - \boldsymbol{E}\xi| \le \boldsymbol{E}|\xi_n - \xi)|$ and by the Cauchy-Schwarz inequality $\boldsymbol{E}|\xi_n - \xi)| \le \sqrt{\boldsymbol{E}|\xi_n - \xi|^2}$, the first statement holds true.

Further, $|\boldsymbol{E}\xi_n^2 - \boldsymbol{E}\xi^2| \le \boldsymbol{E}|\xi_n - \xi||\xi_n + \xi)| \le \sqrt{\boldsymbol{E}|\xi_n - \xi|^2}\sqrt{\boldsymbol{E}(\xi_n + \xi)^2}$. Recall that the first term in the right side of the above inequality converges to zero while the second one is evaluated above as follows:

$$\begin{aligned} \boldsymbol{E}(\xi_n + \xi)^2 &= \boldsymbol{E}(\xi_n - \xi + 2\xi)^2 \le 2\boldsymbol{E}(\xi_n - \xi)^2 + 8\boldsymbol{E}\xi^2 \\ &\to 8\boldsymbol{E}\xi^2, \ n \to \infty. \end{aligned}$$

□

2.4. **Sufficient condition for the convergence with probability one.** (Not included in test).

A sequence of random variables $(X_n)_{n\ge1}$ *converges with probability one to a random variable* $X := X_1 + \sum_{n=1}^{\infty}\big(X_{n+1} - X_n\big)$*, if there exists a decreasing to zero sequence of positive numbers* $(\varepsilon_n)_{n\ge1}$ *such that*

$$\begin{aligned} &\sum_{n=1}^{\infty} \varepsilon_n < \infty \\ &\sum_{n=1}^{\infty} P(|X_{n+1} - X_n| > \varepsilon_n) < \infty. \end{aligned} \tag{2.2}$$

2.5. **Law of large numbers for stationary process.**

2.5.1. Let $(X_n)_{n\ge1}$ be a stationary process in the wide sense, ($\boldsymbol{E}X_n \equiv 0$) with a correlation function K. The following result is readily verified:

$$\lim_n \frac{1}{n}\sum_{j=1}^{n} K(j)\Big(1 - \frac{j}{n}\Big) = 0 \Rightarrow \lim_n \boldsymbol{E}\Big(\frac{1}{n}\sum_{j=1}^{n} X_j\Big)^2 = 0. \tag{2.3}$$

Write

$$\begin{aligned} \boldsymbol{E}\Big(\frac{1}{n}\sum_{j=1}^{n} X_j\Big)^2 &\le \frac{2}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{k} K(k-\ell) \\ &= \frac{2}{n^2}\sum_{k=1}^{n}\sum_{q=0}^{k-1} K(q) = \frac{2}{n^2}\sum_{k=1}^{n}\sum_{q=0}^{n} I(k \ge q+1)K(q) \\ &= \frac{2}{n^2}\sum_{q=0}^{n} K(q)[n-q] = \frac{2}{n}\sum_{q=0}^{n} K(q)\Big(1 - \frac{q}{n}\Big). \end{aligned}$$

□

2.5.2. The Birkhoff-Khinchin Theorem. (Not included in test) *Let* $(X_n)_{n\geq 1}$ *be a strictly stationary process with* $\boldsymbol{E}|X_1| < \infty$. *Then* $\frac{1}{n}\sum_{j=1}^{n} X_j$, $n \geq 1$, *converges with probability one to* $\boldsymbol{E}X_1$, *if the process is ergodic, and to* $\boldsymbol{E}(X_1|\mathfrak{F})$, *where* $\boldsymbol{E}(X_1|\mathfrak{F})$ *is the conditional expectation given the* σ*-algebra of the invariant sets, in a general case.*