Convergence types

Hans Karlsen

UiB: 2021-02-10 16:28:04

Basic probability theory

1. Convergence types

i) Convergence of a real valued sequence.

ii) Convergence of a p-dimensional sequence of real numbers.

iii) Convergence in probability.

iv) Almost sure convergence.

v) Convergence in quadratic mean, i.e. convergence in L^2.

vi) Convergence in quadratic mean, i.e. convergence in L^p for $p \in (0, \infty)$.

vii) Convergence in distribution.

Ordinary sequences

A sequence of real numbers $\{a_n,\, n \geq 0\}$ converges to the real number a if the distance between a_n and a goes to zero as n goes to infinity. In that case we write $\lim a_n = a$ or $a_n \to a$.

Definition 1. [o-notation]
Let $\{a_n,\, n \geq 0\}$ be a convergent sequence that converges to zero. Then $a_n = \mathcal{o}(1)$. Let $\{a_n,\, n \geq 0\}$ be a bounded sequence not necessarily convergent. Then $a_n = \mathcal{O}(1)$.

Remark 1. This means that «$\mathcal{o}(1)$» represents a sequence that converge to zero while «$\mathcal{O}(1)$» denotes a bounded sequence.

In o-notation the convergence of $\{a_n\}$ to a can be written as

$$a_n = a + \mathcal{o}(1)$$

Example 1. If $a_n = \mathcal{o}(1)$ and $b_n = \mathcal{O}(1)$ then $c_n \stackrel{\text{def}}{=} a_n b_n = \mathcal{o}(1)$.

Convergence of a sequence of p-dimensional vectors of real numbers

The sequence of p-dimensional vectors of real numbers $\{\mathbf{a}_n,\, n \geq 0\}$ converge to the p-dimensional vector $\mathbf{a}$ if Euclidean distance between $\mathbf{a}_n$ and $\mathbf{a}$ goes to zero:

$$\lim_{n \longrightarrow \infty} \mathbf{a}_n = \mathbf{a} = \quad \Leftrightarrow \quad \lim_{n \longrightarrow \infty} \|\mathbf{a} - \mathbf{a}_n\| = 0.$$

This means that we have convergence towards $\mathbf{a}$ for $\{\mathbf{a}_n,\, n \geq 0\}$ if

$$\|\mathbf{a} - \mathbf{a}_n\| = \mathcal{O}(1).$$

PROPOSITION 1. The sequence of p-dimensional vectors of real numbers $\{\mathbf{a}_n, n \geq 0\}$ converge to the p-dimensional vector $\mathbf{a}$ iff each component in the vector sequence converge to the corresponding component in vector $\mathbf{a}$;

$$a_{nj} = a_j + o(1), \qquad j = 1, \ldots, p.$$

PROOF OF PROPOSITION 1.

$$\max_j |a_{nj} - a_j|^2 \leq \|\mathbf{a} - \mathbf{a}_n\|^2 = \sum_{j=1}^{p} |a_{nj} - a_j|^2 \leq p \max_j |a_{nj} - a_j|^2.$$

$\square$

Convergence in propability

Definition 2. The sequence of stochastic variables $\{X_n,\, n \geq 0\}$ converges in probability to X if for all $\epsilon > 0$,

$$\mathbb{P}\big(|X_n - X| > \epsilon\big) = \mathcal{o}(1).$$

We can write this as

$$X_n = X + \mathcal{o}_{\mathbb{P}}(1).$$

Remark 2. Convergence in probability depends only on the sequence of marginal distributions of (X_n, X). If X is a constant, it depends only on sequence of marginal distributions of $\{X_n,\, n \geq 0\}$.

Problem 1.1. A p-dimensonal vector sequence $\{\mathbf{X}_n\}$ convergence in probability to $\mathbf{X}$ if the Euclidean distance between the vectors goes to zero inprobability. Explain that is true if and only if each component converge in probability.

Almost sure convergence

Let $\{X_n,\, n \geq 0\}$ be sequence of stochastic variabkes defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ togheter with X. Then

$$X_n \xrightarrow[n]{\text{a.s.}} X$$

if for all outcome ω outside an exceptional set E, the real sequence $\{x_n\}$ with $x_n \stackrel{\text{def}}{=} X_n(\omega)$ converges to the real number $x = x(\omega)$. The exceptional set E has probability 0. This means that

$$\mathbb{P}(X_n \xrightarrow[n]{} X) = 1.. \tag{1.1}$$

This means that we have pointwise convergence except on a null set.

Technical description of almost sure convergence

We need some operational criteria for almost sure convergence,
By definition (1.1) holds iff $\mathbb{P}(A) = 1$. We can (and must) reformulate A;

$$
\begin{aligned}
A &= \left\{ \begin{array}{l} \text{all outomes } \omega \text{ such that the distance between} \\ X_n(\omega) \text{ and } X(\omega) \text{ goes to zero} \end{array} \right\} \\
&= \{\omega\colon (\forall \epsilon > 0)(\exists n)(\forall m \geq n)\colon |X(\omega) - X_m(\omega)| < \epsilon\} \\
&= \bigcap_{\epsilon>0} \bigcup_{m\geq 1} \bigcap_{m \geq n} \{|X_m - X| \leq \epsilon\}.
\end{aligned}
$$

What is impressing is this triplet of set operations.

Operational results for as-convergence

Fortunantely, we have a more straightforward operational expression:

Proposition 2. We have almost sure convergence iff for all $\epsilon > 0$

$$\mathbb{P}\Big(\bigcup_{m\geq n}\{|X_m - X| > \epsilon\}\Big) = o(1) \quad \text{with respect to } n\ . \tag{1.2}$$

Proof of Proposition 2.
We start by taking the complement of the set A,

$$A^c = \bigcup_{\epsilon>0} \bigcap_{n\geq 1} \bigcup_{m\geq n} \{|X_m - X| > \epsilon\}.$$

The sequence of sets in the first union is increasing as ϵ is decreasing and the sequence of sets in the first insection is decreasing as n is increasing. For the last union we again have an increasing sequence of sets. Therefore

$$\begin{aligned}&\mathbb{P}(A^c)\\ &\quad = \lim_{\epsilon\downarrow 0} \mathbb{P}\Big(\bigcap_{n\geq 1} \bigcup_{m\geq n} \{|X_m - X| > \epsilon\}\Big) = \lim_{\epsilon\downarrow 0} \lim_{m\uparrow\infty} \mathbb{P}\Big(\bigcup_{m\geq n} \{|X_m - X| > \epsilon\}\Big).\end{aligned}$$

We have as-convergence iff $\mathbb{P}(A^c) = 0$ which is equivalent with (1.2) $\square$

«Almost sure» is stronger that «in probability»

COROLLARY 1. As-convergence implies convergence in probability.

PROOF.

$$\{|X_n - X| > \epsilon\} \subseteq \bigcup_{m \geq n} \{|X_m - X| > \epsilon\}.$$

If the probability of the union of sets at the right hand side of the set inclusion goes to zero, then necessarily the same must be true for the left hand side.
□

Borel Cantelli

A key result when it comes to proving strong convergence is the following:

THEOREM 1. [Borel Cantelli]
Let $\{A_n, n \geq 1\}$ be a sequence of events. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\mathbb{1}_{A_n} \xrightarrow[n]{\text{a.s.}} 0$.

PROOF OF THEOREM 1.
Let $Y = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$. Then $\mathbb{E}Y = \sum \mathbb{P}(A_n) < \infty$ and

$$\infty > \mathbb{E}Y \geq \mathbb{E}Y\mathbb{1}(Y = \infty) = \infty \times \mathbb{P}(Y = \infty) = \infty \times 0 = 0.$$

Hence $B = \left\{Y = \sum \mathbb{1}_{A_n} < \infty\right\}$ has probability 1.

For $\omega \in B$ we must have that $\mathbb{1}_{A_n}(\omega) \equiv 0$ for n large enough since an indicator is either zero or one. But then $\xrightarrow[n]{} \mathbb{1}_{A_n}(\omega) = 0$.

$\square$

Corollary 2. If for all $\epsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty$. Then $X_n \xrightarrow[n]{\text{a.s.}} X$.

Proof of Corollary 2.
Let $A_n = \{|X_n - X| > \epsilon\}$ in Theorem 1. Then with probability 1 we have that $|X_n - X| \leq \epsilon$. Since $\epsilon > 0$ is arbitrary, the limit must be zero.tBC $\square$

Remark 3. This result is an important ingredient in a proof of SLLN.

Remark 4. This corollary also follows from Proposition 2.

Remark

Almost sure convergence of $\{X_n,\, n \geq 0\}$ towards X depends on the simultaneous tail distributions of sequence together with X. This is in contrast to convergence in probability. However, the condition

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty$$

depends only on the sequence of marginal distributions.

Markovs inequality

PROPOSITION 3. Let ϕ be an increasing function with $\phi(x) > 0$ foe $x > 0$. Then

$$\mathbb{P}\big(\phi(X) \geq \epsilon\big) \leq \frac{1}{\phi(\epsilon)} \mathbb{E}\,\phi(X).$$

Proof of Proposition 3.

$$\begin{aligned}\mathbb{E}\,\phi(X) &\geq \mathbb{E}\,\phi(X)\mathbb{1}(\phi(X) \geq \epsilon) \geq \mathbb{E}\,\phi(\epsilon)\mathbb{1}(\phi(X) \geq \epsilon) = \phi(\epsilon)\mathbb{E}\,\mathbb{1}(\phi(X) \geq \epsilon)\\ &= \phi(\epsilon)\mathbb{P}(\phi(X) \geq \epsilon).\end{aligned}$$

$\square$

Remark 5. Note the use of an indicator here. So, the message is to avoid the distribution function and integral sign whenever that is possible.

Chebyshev and mean square convergence

COROLLARY 3. Chebyshev's inequality

$$\mathbb{P}\big(\big(|X-\mu| \geq \epsilon\big) \leq \frac{1}{\epsilon^2}\sigma^2.$$

PROOF.
Let $X' = X - \mu$ and $\phi(x) = x^2$. $\square$

COROLLARY 4. If $X_n \xrightarrow[n]{L^2} X$. Then $X_n \xrightarrow[n]{\mathbb{P}} X$.

PROOF.
Use Chebyshev's inequality $\square$

Convergence in distribution

Definition 3. The sequence $\{X_n, n \geq 0\}$ converges in distribution to X if the corresponding sequence of distribution function converges weakly which means that $F_n(x) = F(x) + \mathcal{O}(1)$ for at continuity points x for F. Here $F_n = F_{X_n}$ and $F = F_X$.

Remark 6. The most important example is of course all variants of the central limit theorem.

Remark 7. Other examples of convergence of distribution functions are the converges of the binomial distribution to a Poisson distribution and the convergence of the geomtric distribution to the exponential distribution.

XYZ-lemma

LEMMA.

$$X = Y + Z$$

Then for $\delta > 0$

$$F_Y(x-\delta) - \mathbb{P}(|Z| > \delta) \leq F_X(x) \leq F_Y(x+\delta) + \mathbb{P}(|Z| > \delta)$$

REMARK 8. This is a squeeze pre-theorem (sandwhich theorem).

Proof of XYZ-lemma

PROOF.
We have

$$\begin{aligned}\{X \leq x\} &= \{Y + Z \leq x\} \\ &= \{Y + Z \leq x, |Z| \leq \delta\} \bigcup \{Y + Z \leq x, |Z| > \delta\} \\ &\subseteq \{Y \leq x + \delta\} \bigcup \{|Z| > \delta\}\end{aligned}$$

which implies

$$\begin{aligned}F_X(x) = \mathbb{P}(X \leq x) &\leq \mathbb{P}\big(Y \leq x + \delta\big) + \mathbb{P}\big(|Z| > \delta\big) \\ &= F_Y(x + \delta) + \mathbb{P}\big(|Z| > \delta\big).\end{aligned}$$

For the lower bound, $Y' = X + (-Z)$, $X' = Y$ and$x' = x - \delta$. Then $X' = Y' + Z$.

□

Convergence in probability implies convergence in distribution

PROPOSITION 4. Convergence in probability implies convergence in distribution.

$\xrightarrow[n]{\mathbb{P}}$ implies $\overset{d}{\underset{n}{\Longrightarrow}}$

Proof of Proposition 4.
Suppose that $X_n \xrightarrow[n]{\mathbb{P}} X$. Then $X_n = X + (X_n - X) = Y + Z_n$ with $Y = X$ and $Z_n = X_n - X$. By the XYZ-lemma,

$$F_X(x-\delta) - \mathbb{P}(|X - X_n| > \delta) \leq F_n(x) \leq F(x+\delta) + \mathbb{P}(|X - X_n| > \delta).$$

Let x be a continuity point for F_X. We first let n go to infinity and then let δ go to zero,

$$\begin{aligned}&\lim_{\delta\downarrow 0} F_X(x-\delta) - \overline{\lim_n}\, \mathbb{P}(|X - X_n| > \delta)\\ &\qquad \leq \underline{\lim}\, F_n(x) \leq \overline{\lim}\, F_n(x) \leq \lim_{\delta\downarrow 0} F_X(x+\delta) + \overline{\lim_n}\, \mathbb{P}(|X - X_n| > \delta)\end{aligned}$$

Hence

$$F_X(x) \leq \underline{\lim}\, F_n(x) \leq \overline{\lim}\, F_n(x) \leq F_X(x)$$

which means that $\lim_{n\to\infty} F_n(x) = F(x)$. $\square$

Cramér - Slutsky

PROPOSITION 5. [Cramér -slutsky]
Let $Y_n \underset{n}{\overset{d}{\Longrightarrow}} X$.

i) Suppose that $X_n = Y_n + Z_n$ and $Z_n \underset{n}{\overset{\mathbb{P}}{\longrightarrow}} 0$ Then $X_n \underset{n}{\overset{d}{\Longrightarrow}} X$.

ii) If $X_n = Y_n Z_n$ and $Z_n \underset{n}{\overset{\mathbb{P}}{\longrightarrow}} 1$ Then $X_n \underset{n}{\overset{d}{\Longrightarrow}} X$.

PROOF OF PROPOSITION 5.
For i) we use the XYZ -lemma and for ii) we must in addition know that convergence in distribution implies that the sequence is bounded in probability.
$\square$

MA(q)

Example 2. Let $\{Z_t,\, t \in \mathbb{Z}\}$ be iid with finite expectation and variance σ_Z^2.

$$X_t = \mu + \sum_{j=0}^{q} \theta_j Z_{t-j}$$

with q finite.

Then

$$\overline{X}_n \xrightarrow[n]{\text{a.s.}} \mu,$$
$$\sqrt{n}(\overline{X}_n - \mu) \xRightarrow[n]{d} \mathcal{N}\Big(0, \sigma^2(\sum \theta_j)^2\Big).$$

The proof is a challenge in HW2.

Example

Assume we have observations $X_1, \ldots, X_n$ from a time series. The empirical ACVF is

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \overline{X}_n)(X_t - \overline{X}_n), \qquad h = 0, \ldots, n-1.$$

with $\widehat{\gamma}(h) = \widehat{\gamma}(-h)$ for $h < 0$ and $\widehat{\gamma}(h) \equiv 0$ for $|h| \geq n$.

Consistensy

The estimator does not depend on the mean and we can therefore assume that $\mu = 0$. Let $h \geq 0$ be fixed, $m = n - h$ and

$$\widetilde{\gamma}(h) = \frac{1}{m} \sum_{t=1}^{m} X_{t+h} X_t.$$

More details

$$
\begin{aligned}
\widehat{\gamma}(h) &= \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h}-\overline{X}_n)(X_t-\overline{X}_n) \\
&= n^{-1}\sum_{t=1}^{n-h}X_{t+h}X_t - n^{-1}\sum_{t=1}^{n-h}X_t\overline{X}_n - n^{-1}\sum_{t=1}^{n-h}X_{t+h}\overline{X}_n + n^{-1}\sum_{t=1}^{n-h}\overline{X}_n{}^2 \\
&= \frac{m}{n}\widetilde{\gamma}(h) - \frac{m}{n}\overline{X}_m\overline{X}_n - \big(\overline{X}_n - \frac{h}{n}\overline{X}_h\big)\overline{X}_n + \frac{m}{n}\overline{X}_n{}^2 \\
&= \frac{m}{n}\widetilde{\gamma}(h) - \Big(\frac{m}{n}\overline{X}_m - \frac{h}{n}\overline{X}_h - \frac{h}{n}\overline{X}_n\Big)\overline{X}_n.
\end{aligned}
$$

From the last line we can get the bias, but the variance is harder. However, most important is the asymptotic variance.