Convergence: a.s., i.p., and L^p

Sayan Mukherjee

Why convergence ?

Given a probability space $(\Omega, \mathcal{F}, \mathbf{P}_\theta)$ with $\theta \in \Theta$ given $X_1, ..., X_n$ defined on Ω our objective is to infer θ.

$\hat{\theta}_n(X_1, ..., X_n)$ is a statistic or estimator and $\hat{\theta}(x_1, ..., x_n)$ is the estimate.

Definition 0.0.1 (Weak consistency) *An estimator $\hat{\theta}_n$ is weakly consistent if $\forall \theta \in \Theta$ and $\forall \varepsilon > 0$*

$$\lim_{n \to \infty} \mathbf{P}_\theta(|\theta_n - \theta| > \varepsilon) = 0.$$

This is written as $\hat{\theta}_n \stackrel{\mathbf{P}_\theta}{\to} \theta$.

Definition 0.0.2 (Strong consistency) *An estimator $\hat{\theta}_n$ is strongly consistent if $\forall \theta \in \Theta$ and $\forall \varepsilon > 0$*

$$\lim_{n \to \infty} \mathbf{P}_\theta(\omega : \theta_n(\omega) \neq \theta(\omega)) = 0.$$

This is written as $\hat{\theta}_n \stackrel{a.s.}{\to} \theta$.

Types of convergence

Given random variables $X_1, ..., X_n$ and X defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ the following types of convergence can happen.

Definition 0.0.3 (In probability) $X_n \stackrel{p}{\to} X$ *if* $\forall \varepsilon > 0$

$$\lim_{n \to \infty} \mathbf{P}(|X_n - X| > \varepsilon) = 0.$$

Definition 0.0.4 (In probability) $X_n \stackrel{a.s.}{\to} X$ *if*

$$\lim_{n \to \infty} \mathbf{P}(\omega : X_n(\omega) \neq X(\omega)) = 0.$$

Definition 0.0.5 (In L^p) $X_n \stackrel{L^p}{\to} X$ *if*

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

Definition 0.0.6 (In distribution) $X_n \stackrel{d}{\to} X$ *if*

$$\lim_{n \to \infty} |\mathbf{E}f(X_n) - \mathbf{E}f(X)| = 0,$$

for every bounded continuous function f. Note this requires convergence at all points of continuity of $F_X(x)$ and not at all points X.

A sequence $X_1, ..., X_n, ...$ is a Cauchy sequence if $\forall \varepsilon$ there exists $n_0(\varepsilon)$ such that for $m, n > n_0$

$$|X_n - X_m| < \varepsilon.$$

The remainder of the lecture relates these different types of convergences to each other.

Almost sure convergenc can be related to convergence in probability of Cauchy sequences.

Theorem 0.0.1 $X_n \stackrel{a.s.}{\to} X$ *iff* $\forall \varepsilon > 0$

a) $\lim_{n \to \infty} \mathbf{P}(\sup_{k > n} |X_k - X| \geq \varepsilon) = 0$

b) $\{X_n, n \geq 1\}$ *is Cauchy with probability 1 iff*

$$\lim_{n\to\infty} \mathbf{P}(\sup_{k>0} |X_{n+k} - X_n| \geq \varepsilon) = 0.$$

Corollary 0.0.1

$$\begin{aligned} \mathbf{P}\left(\sup_{k\geq n} |X_k - X| \geq \varepsilon\right) &= \mathbf{P}\left(\bigcup_{k\geq n} (|X_k - X| \geq \varepsilon)\right) \\ &\leq \sum_{k\geq n} \mathbf{P}(|X_k - X| \geq \varepsilon). \end{aligned}$$

So a sufficient condition for $X_n \stackrel{a.s.}{\to} X$ *is* $\forall \varepsilon > 0$

$$\sum_{k=1}^{\infty} \mathbf{P}(|X_k - X| \geq \varepsilon) < \infty.$$

Proof.

$$\begin{aligned} A_n^\varepsilon &= \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\} \\ A^\varepsilon &= \limsup A_N^\varepsilon. \end{aligned}$$

For all $\varepsilon > 0$

$$\sum_{k=1}^{\infty} \mathbf{P}(|X_k - X| \geq \varepsilon) < \infty \Rightarrow \mathbf{P}(A^\varepsilon) = 0 \Rightarrow \mathbf{P}(\omega : X_n(\omega) \not\to X(\omega)) = 0 \, \square.$$

Relations between types of convergence

The following theorem summarizes the relation between the different types of convergence.

Theorem 0.0.2 *The following implications hold*

1) $X_n \stackrel{a.s.}{\to} X \Rightarrow X_n \stackrel{p}{\to} X,$

2) $X_n \stackrel{L^p}{\to} X \Rightarrow X_n \stackrel{p}{\to} X, p > 0,$

3) $X_n \stackrel{p}{\to} X \Rightarrow X_n \stackrel{d}{\to} X.$

Proof. Proof of (1)

$$\begin{aligned} 0 &= \mathbf{P}([|X_n - X| > \varepsilon] i.o.) \\ &= \mathbf{P}(\lim \sup_{n\to\infty} (|X_n - X| > \varepsilon)) \\ &= \lim_{N\to\infty} \mathbf{P}\left(\bigcup_{n\geq N} (|X_n - X| > \varepsilon)\right) \\ &\geq \lim_{n\to\infty} \mathbf{P}(|X_n - X| > \varepsilon). \end{aligned}$$

Proof of (2)

$$\mathbf{P}(|X_n - X|^p \geq \varepsilon) \leq \frac{\mathbf{E}|X_n - X|^{2p}}{\varepsilon^2}.$$

Proof of (3), left to reader. $\square$

The remainder of the lecture illustrates why the converses to the implications do not hold. This also gives us a hierarchy of the implications.

Example 1. $X_n \stackrel{p}{\to} X \not\Rightarrow X_n \stackrel{a.s.}{\to} X$ and $X_n \stackrel{L^p}{\to} X \not\Rightarrow X_n \stackrel{a.s.}{\to} X$

$\Omega = [0,1], \quad \mathcal{F} = \mathcal{B}([0,1]), \mathbf{P} = \lambda$

Define the sets $A_n^i = [\frac{i-1}{n}, \frac{i}{n}]$ and the random variables $X_n^i = \mathbf{1}_{A_n^i}(\omega), \quad 1,..,n, n \geq 1$.
We now define a sequence

$$\{X_1^1; X_2^1, X_2^2; X_3^1, X_3^2, X_3^3; ,....\}.$$

This sequence converges in probability, it converges in L^p (for $0 < p < \infty$ but it does not converge at any point $\omega \in [0,1]$, observe that $X_n(\omega) = 1$ for infinite values of n , again $X = 0$.

Example 2. $X_n \stackrel{a.s.}{\to} X \Rightarrow X_n \stackrel{p}{\to} X \not\Rightarrow X_n \stackrel{L^p}{\to} X$

$\Omega = [0,1], \quad \mathcal{F} = \mathcal{B}([0,1]), \mathbf{P} = \lambda$

$$X_n(\omega) = \begin{cases} e^n & 0 \leq \omega \leq \frac{1}{n} \\ 0 & \omega > \frac{1}{n}. \end{cases}$$

For this sequence $X_n \stackrel{a.s.}{\to} X$, $X_n \stackrel{p}{\to} X$ but the sequence does not converge in L^p

$$\mathbf{E}|X_n|^p = \frac{e^{np}}{n} \to \infty.$$

Example 3. $X_n \stackrel{L^p}{\to} X \not\Rightarrow X_n \stackrel{a.s.}{\to} X$

$\{X_n\}$ are iid random variables with

$$\mathbf{P}(X_n = 1) = p_n, \quad \mathbf{P}(X_n = 0) = 1 - p_n$$

the following hold as $n \to \infty$

$$\begin{aligned} X_n \stackrel{p}{\to} 0 \quad &\Leftrightarrow \quad p_n \to 0 \\ X_n \stackrel{L^p}{\to} 0 \quad &\Leftrightarrow \quad p_n \to 0 \\ X_n \stackrel{a.s.}{\to} 0 \quad &\Rightarrow \quad \sum_{i=1}^{\infty} p_n < \infty. \end{aligned}$$

If $p_n = \frac{1}{n}$ then $X_n \stackrel{p}{\to} 0$ and $X_n \stackrel{L^p}{\to} 0$ but $X_n \stackrel{a.s.}{\not\to} 0$.