Convergence of random variables

Coin, $P(\{H\})$ estimate?

Toss the coin, record if I see a head.

$$\underbrace{\hat{P}_n(\{H\})}_{\text{R.V.}} = \frac{\#\text{ of heads}}{n}$$

$$\hat{P}_n(\{H\}) \xrightarrow{?} P(\{H\})$$

Consider $X_i : \Omega \to \mathbb{R}$, $i \in \mathbb{N}$, $X : \Omega \to \mathbb{R}$.

$(\Omega, \mathcal{A}, P)$ a probability space.

(0) Point-wise convergence or sure convergence

$$X_n(\omega) \to X(\omega), \ \forall \omega \in \Omega$$

(1) $(X_i)_{i \in \mathbb{N}}$ converges to X almost surely $:\Leftrightarrow$

$$P\left(\{\omega \in \Omega \mid \lim_{i \to \infty} X_i(\omega) = X(\omega)\}\right) = 1$$

Notation: $X_i \longrightarrow X$ a.s.

Used in proof of Strong Law of Large Numbers

(2) $(X_i)_{i\in\mathbb{N}}$ converges to X in probability:

$$\Leftrightarrow \quad \forall \varepsilon > 0 \quad P\left(\{\omega\in\Omega \mid |X_i(\omega) - X(\omega)| > \varepsilon\}\right) \to 0$$

Used in proof of Weak Law of Large Numbers. convergence of empirical estimators.

Let us check that these definitions make sense. We need to prove that the events in (1) and (2) are measurable and are in $\mathcal{A}$.

Case (1): $\lim X_i(\omega) = X(\omega)$

$$\Leftrightarrow \quad \forall k\in\mathbb{N} \quad \exists N\in\mathbb{N} \quad \forall n > N: \quad |X_n(\omega) - X(\omega)| < 1/k$$

So we get: $\{\omega \mid X_i(\omega) \to X(\omega)\}$

$$= \underbrace{\bigcap_{k\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N}}_{\text{countable unions and intersections.}} \{\omega \mid |X_i(\omega) - X(\omega)| < 1/k\}$$

$X_i(\omega)$, $X(\omega)$ are measurable

$\Rightarrow$ $|X_i(\omega) - X(\omega)|$ is measurable

so $\{\dots\}\in\mathcal{A}$

(3) $X_n \to X$ in L^p ("in the p-th mean"):

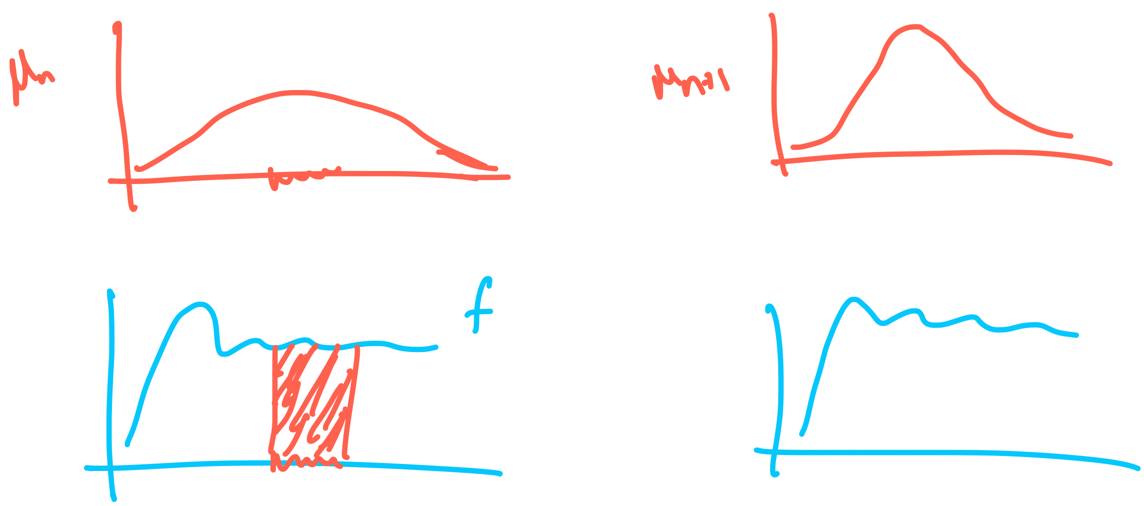
$\Leftrightarrow$ $X_n, X \in L^p$ and $E(|X_n - X|^p) \to 0$

(4) Let $M^1(\mathbb{R}^n)$ be the set of all probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Assume $(\mu_n)_n \subset M^1(\mathbb{R}^n)$, $\mu \in M^1(\mathbb{R}^n)$.

$C_b(\mathbb{R}^n) :=$ space of bounded continuous funcs.

$\mu_n \to \mu$ <u>weakly</u> $:\Leftrightarrow$

$$\forall f \in C_b(\mathbb{R}^n): \int f \, d\mu_n \to \int f \, d\mu$$



(5) $X_i, X : (\Omega, \mathcal{A}, P) \to \mathbb{R}^n$. The sequence X_n converges <u>in distribution</u> to X $:\Longleftrightarrow$ the distribution P_{X_n} converges to P_X weakly.

Used in the proof of the Central Limit Theorem. Weakest form of convergence.

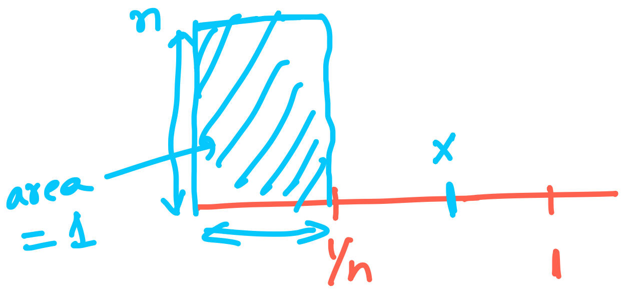
We have the following implications

$$\begin{tikzcd}
\text{surely} \arrow[d, Rightarrow] & & \text{in } L^1 \arrow[dl, Rightarrow] & \text{in } L^p \ (p>1) \arrow[l, Rightarrow] \\
\text{almost surely} \arrow[r, Rightarrow] & \text{in probability} \arrow[d, Rightarrow] & & \\
& \text{in distribution} & &
\end{tikzcd}$$

<u>Remark</u>: All the other relations not shown are in general not true.

Example (converge a.s., in prob, but not in L^1)

$X_n : \mathbb{R} \to \mathbb{R}$ $\qquad X_n(\omega) = \begin{cases} n & \text{to } 0 \le x \le 1/n \\ 0 & \text{otherwise} \end{cases}$



$\forall x > 0 : \; X_n(\omega) \to 0$

Can see that, a.s, in probability

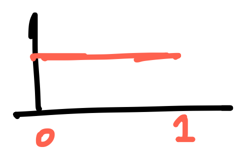
But: no convergence in L^1

$$\|X_n(\omega) - X\|_1 = \int \|X_n(\omega) - X\|_1 = 1$$

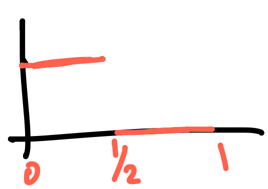
Example (converges in prob, in L^1, but not a.s.)

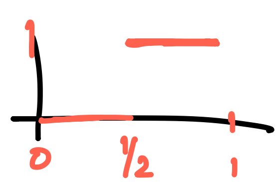
"sliding blocks"

$f_1 = \mathbb{1}_{[0,1]}$

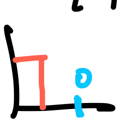


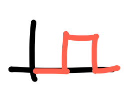
$f_2 = \mathbb{1}_{[0,1/2]}$, $f_3 = \mathbb{1}_{[1/2,1]}$

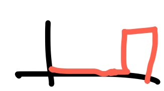




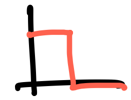
$f_4 = \mathbb{1}_{[0,1/3)}$, $f_5 = \mathbb{1}_{[1/3,2/3)}$, $f_6 = \mathbb{1}_{[2/3,1]}$

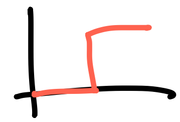






Example (converges in distribution, but not in prob.)

- $X_n = [0,1] \to \mathbb{R}$, $\quad X_1 = X_2 = \cdots \quad \mathbb{1}_{[0,1/2)}$ 

$X = \mathbb{1}_{[1/2,1]}$ 

Obviously $X_n \not\to X$ in prob. but

$$P_{X_1} = \frac{1}{2}(\delta_0 + \delta_1) = P_{X_2} = P_{X_3} = \cdots = P_X$$

So in distribution $X_n \to X$

Borel-Cantelli Theorem

$(\Omega, \mathcal{A}, P)$ prob. space, $(A_n)_n$ sequence of events in the prob. space.

$$P(A_n \text{ infinitely often}) := P(A_n \text{ i.o.})$$

$$= P(\{\omega \in \Omega \mid \omega \in A_n \text{ for infinitely many } n\})$$

<u>Proposition</u>: X_n, X r.v. on $(\Omega, \mathcal{A}, P)$.

$$X_n \to X \text{ a.s.} \iff \forall \varepsilon > 0: P(\{|X_n - X| > \varepsilon \text{ infinite often}\}) = 0$$

<u>Theorem</u>: Consider a sequence of events $(A_n)_n \subset \mathcal{A}$.

(1) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$

(2) If $\sum_{n=1}^{\infty} P(A_n) = \infty$, and if $(A_n)_n$ are independent then $P(A_n \text{ i.o.}) = 1$

Application in learning theory:

Assume that $P(|X_n - X| > 1/n) < \delta_n$, and assume that $\sum_{n=1}^{\infty} \delta_n < \infty$. Then you can use Borel-Cantelli to prove that

$$P(|X_n - X| > 1/n \ \text{i.o.}) = 0,$$

thus $X_n \to X$ a.s.