CSE 840: Computational Foundations of Artificial Intelligence Nov. 22, 2023

Convergence of Random Variables and Borel Cantelli

Instructor: Vishnu Boddeti *Scribe: Max Resch, Shafkat Kabir*

1 Convergence of Random Variables

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

Toss the coin, record if I see a head.

$\hat{P}_n(\{H\}) = \frac{num.of heads}{n}$, where $\hat{P}_n(H)$ is a random variable.

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

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

where (Ω, A, P) is a probability space.

1.1 Levels of Strength of Convergence

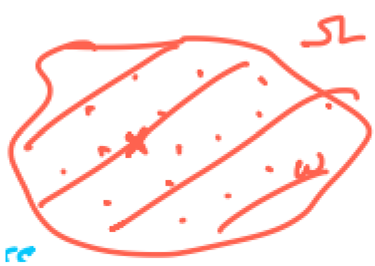
(0) Point-wise convergence or "Sure Convergence"

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

(1) $(x_i)_{i\in\mathbb{N}}$ converges to X almost surely (a.s): $\iff$

$P(\{w \in \Omega | \lim_{i\to\infty} x_i(w) = x(w)\}) = 1$

Notation: $x_i \to x$ a.s.



Note: This is used in the proof of Strong Law of Large Numbers

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

$\iff \forall\epsilon > 0$

$P(\{w \in \Omega||x_i(w) - x(w)| > \epsilon\}) \to 0$

Note: This is used in the proof of Weak Law of Large Numbers and 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 A.

case (1): $\lim x_i(w) = x(w) \iff$

$\forall k \in \mathbb{N}\ \exists\ N \in \mathbb{N}\ \forall n > N : |x_n(w) - x(w)| < \frac{1}{k}$

So we get: $\{w|x_i(w) \to x(w)\}$

$= \bigcap_{k\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq\mathbb{N}} \{w||x_i(w) - x(w)| < \frac{1}{k}\}$

where $\bigcap_{k\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq\mathbb{N}}$ are the countable unions and intersections

$\Rightarrow x_i(w), x(w)$ are measurable, $|x_i(w) - x(w)|$ is measurable so $\{...\} \in A$

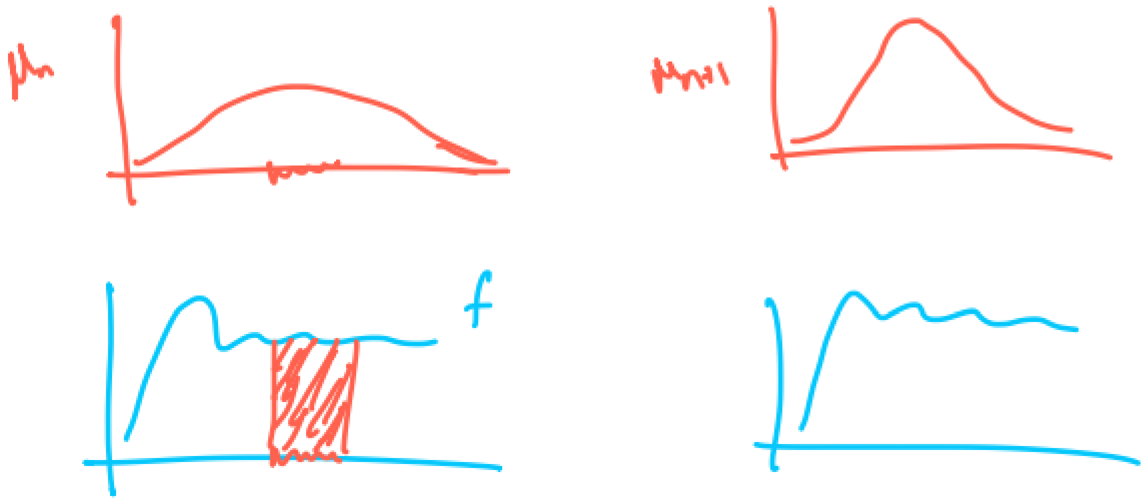
(3) $x_n \to$ in L^p ("in the p-th mean"):

$\iff x_n, x \in L^p$ and $||x_i - x||_p \to 0$

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

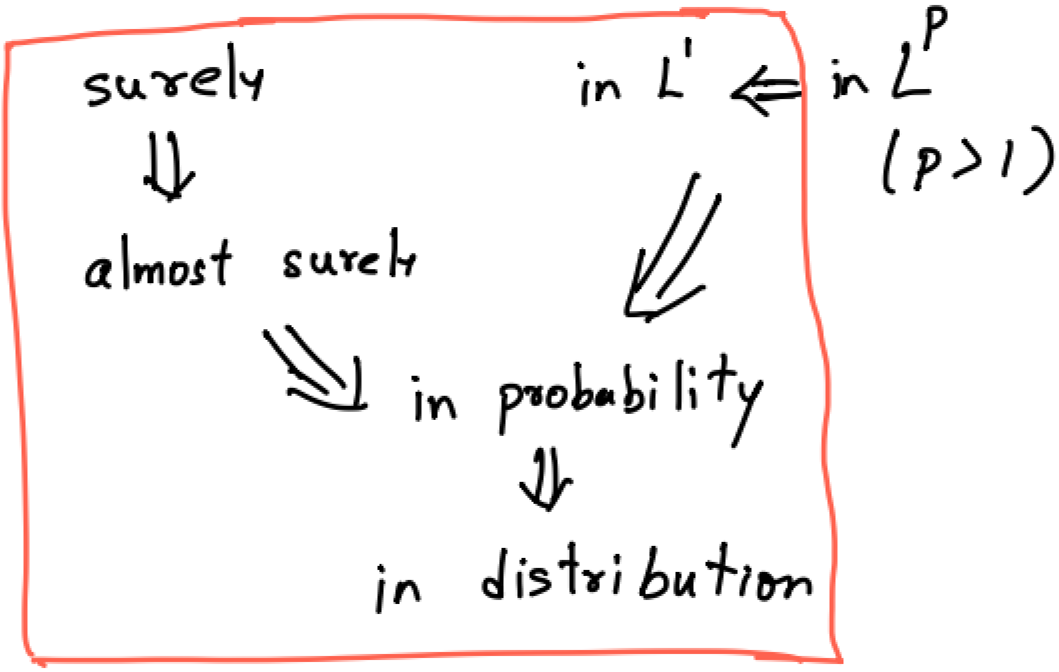
$C_b(\mathbb{R}^n) :=$ space of bounded continuous functions.

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



(5) $x_i, x : (\Omega, A, P) \to \mathbb{R}^n$. The sequence x_n converges in distribution to x: $\iff$ the distribution P_{x_n} converges to P_x weakly.

Note: This is used in the proof of the Central Limit Theorem. This is the weakest form of convergence.



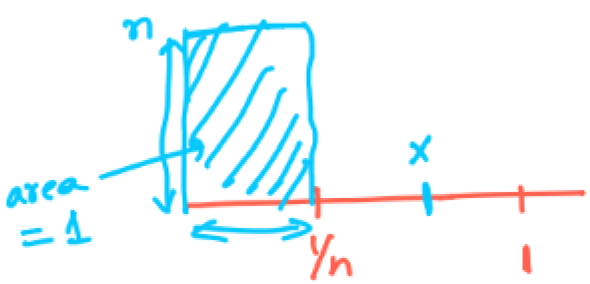
We have the following implications:

Remark 1. *All the other relations now shown in general not true.*

Example. Converge a.s., in probability, but not in L'.

$X_n : \mathbb{R} \to \mathbb{R}$

$$X_n(w) = \begin{cases} n & \text{to } 0 \leq x \leq \frac{1}{n}, \\ 0 & \text{otherwise} \end{cases}$$



$\forall x > 0 : \quad X_n(w) \to 0$

Can see that, a.s, in probability.

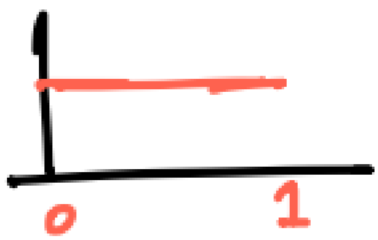
But: no L' convergence.

$||x_n(w) - x||_1 = \int ||x_n(w) - x||_1 = 1$

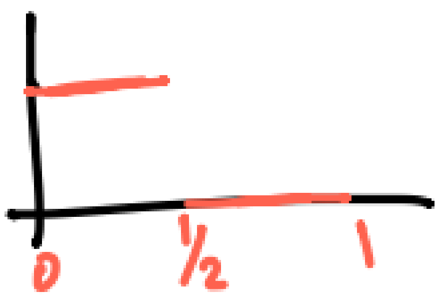
Example. converges in probability, but not a.s.

"sliding blocks"

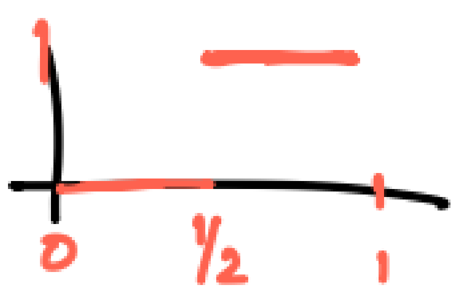
$f_1 = \mathbb{1}_{[0,1]}$



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

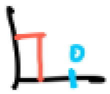


$f_3 = \mathbb{1}_{[\frac{1}{2},1]}$

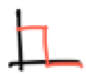


Similarly,

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



Example. converges in distribution, but not in probability.

$X_n = [0,1] \to \mathbb{R},\ x_1 = x_2 ......\mathbb{1}_{[0,\frac{1}{2}]}$ 

$x = \mathbb{1}_{[\frac{1}{2},1]}$ 

Obviously $x_n \nrightarrow x$ in probability but,

$P_{x_1} = \frac{1}{2}(\delta_0 + \delta_1) = P_{x_2} = P_{x_3} = ... = P_x$

So in distribution, $x_n \to x$.

Borel-Cantelli Theorem

(Ω, A, P) probability space, $(A_n)_n$ sequence of events in the probability space.

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

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

Proposition: X_n, X r.v. on (Ω, A, P).

$X_n \to X$ a.s. $\iff \forall_{\varepsilon>0} : P(|X_n - X| > \varepsilon \text{ i.o.}) = 0$

Theorem: Consider a sequence of events

$(A_n)_n \subset A$.

(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 $P(|X_n - X| > \frac{1}{n}) < \delta_n$, and assume that $\sum_{n=1}^{\infty} \delta_n < \infty$.

Then $P(|X_n - X| > \frac{1}{n} \text{ i.o.}) = 0$. Then you can use Borel-Cantelli to prove that

$P(|X_n - X| > \frac{1}{n} \text{ i.o.}) = 0$, thus $X_n \to X$ a.s.