SDS 387
Linear Models

Fall 2024

Lecture 2 - Thu, Sep 29, 2024

Instructor: Prof. Ale Rinaldo

- Last time: convergence wp 1

$\{X_n\}$ and X random variables in $\mathbb{R}^d$

$$X_n \xrightarrow{wp\,1} X \quad \text{when}$$

with prob 1

$$\mathbb{P}\Big(\lim_n \underbrace{d(X_n, X)}_{\|X_n - X\|} = 0 \Big) = 1$$

- Requires a handle of joint distribution of $\{X_n\}$ and X

Think of $(\{X_n\}, X)$ as a random variable whose realizations are pairs of $(\{x_n\}, x)$

$(\mathbb{R}^d)^\infty$ ↙ ↘ $\mathbb{R}^d$

$X_n \xrightarrow{wp\,1} X$ when the prob. of seeing a realization s.t. the limit does not exists is zero!

Equivalently $X_n \xrightarrow{wp1} X$ iff $\forall \varepsilon > 0$ (→ small)

$$\mathbb{P}\left(\|X_n - X\| < \varepsilon \text{ eventually} \right) = 1$$

$\exists N$ (random) s.t. $\forall n \geq N$ $\|X_n - X\| < \varepsilon$

or

$$\mathbb{P}\left(\|X_n - X\| \geq \varepsilon \text{ infinitely often} \right) = 0$$

$\exists$ infinitely many (random) indices n s.t. $\|X_n - X\| \geq \varepsilon$

- for $\varepsilon > 0$ (small) let $A_{\varepsilon,n} = \{ \|X_n - X\| < \varepsilon \}$

Then $X_n \xrightarrow{wp1} X$ is equivalent to:

HW

$$\underbrace{\mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{\varepsilon,m} \right)}_{\liminf A_{\varepsilon,n}} = 1$$

$\liminf A_{\varepsilon,n} \iff \|X_n - X\| < \varepsilon$ eventually

and

$$\underbrace{\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A^c_{\varepsilon,m} \right)}_{\limsup A_{\varepsilon,n}} = 0$$

$\limsup A_{\varepsilon,n} \iff \|X_n - X\| \geq \varepsilon$ infinitely often

- **Convergence in probability**

This is a weaker notion of stochastic convergence that is central to statistical inference

$$\boxed{X_n \xrightarrow{P} X \quad \text{when} \quad \underbrace{\lim_n \mathbb{P}\left(\|X_n - X\| > \varepsilon \right) = 0}_{d(X_n, X)} \quad \forall \varepsilon > 0 \text{ (→ small)}}$$

This result does not require control of the joint distribution of $\{X_n\}$ and X but only of X_n and X, $\forall n$.

<u>Thm</u> Convergence wp 1 implies convergence in probability

Pf/ Let $C = \{\lim_n X_n = X\}$. Then $X_n \xrightarrow{wp1} X$ is equivalent to $\mathbb{P}(C) = 1$. Let $\varepsilon > 0$ small and let $C_n = \{\|X_k - X\| \leq \varepsilon, \forall k \geq n\}$

Then $C \subseteq \bigcup_{n=1}^{\infty} C_n$. So $\mathbb{P}(\bigcup_n C_n) = 1$.

But $C_n \subseteq C_{n+1}$ $\forall n$ $\Rightarrow \mathbb{P}(C_n) \to 1$ as $n \to \infty$

Therefore $\mathbb{P}(C_n^c) \to 0$ as $n \to \infty$. ∎

<u>Example</u> (the typewriter sequence)

Let $U \sim \text{Uniform}(0,1)$. Define $\{X_n\}$ as follows. For every $n \in \mathbb{N}_+$ we have that

$$2^k \leq n < 2^{k+1} \quad \text{where} \quad k = \lfloor \log_2 n \rfloor$$

So define

$$X_n = f_n(U) = \begin{cases} 1 & \text{if } U \in \left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right] \\ 0 & \text{otherwise} \end{cases}$$

So $X_1 = 1$

$X_2 = 1$ if $U \in [0, 1/2]$

$X_3 = 1$ if $U \in [1/2, 1]$

$X_4 = 1$ if $U \in [0, 1/4]$

$X_5 = 1$ if $U \in [1/4, 1/2]$

$X_6 = 1$ if $U \in [1/2, 3/4]$

Now for $\varepsilon > 0$

$$\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}\left(U \in \left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right]\right) = \frac{1}{2^k} \to 0 \text{ as } n \to \infty$$

$$k = \lfloor \log_2 n \rfloor$$

$\Downarrow$

$X_n \xrightarrow{P} 0$!

Is it true $X_n \xrightarrow{wp\,1} 0$? No!

$$\{ u \in (0,1) : f_n(u) > \varepsilon \ \text{ i.o.} \} = (0,1)$$

<u>Example</u> Let $\{U_n\} \overset{iid}{\sim}$ Uniform $[0,1]$ and let

$$X_n = \begin{cases} 1 & U_n \in [0, 1/n] \\ 0 & \text{otherwise} \end{cases} := \mathbb{1}\{U_n \in [0, 1/n]\}$$

$X_n \xrightarrow{P} 0$ because $\forall \varepsilon > 0$ $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(U_n \in [0, 1/n])$

$$= \frac{1}{n} \to 0 \text{ as } n \to \infty$$

Does $X_n \xrightarrow{wp\,1} 0$? No!

$$\mathbb{P}(|X_n| < \varepsilon \text{ eventually}) = \mathbb{P}(\liminf_n \underbrace{A_{\varepsilon,n}}_{\{|X_n| < \varepsilon\}})$$

$$= \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_{\varepsilon,m}\right)$$

$$\leq \sum_{n=1}^{\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_{\varepsilon,m}\right)$$

<u>Fact</u>: if $\{B_n\}$ is a sequence of events

$$\mathbb{P}\left(\bigcup_n B_n\right) \leq \sum_n \mathbb{P}(B_n)$$

countable sub-additivity of prob.

(countable additivity means "$=$" if B_n's are pairwise disjoint)

Next $\quad \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_{\varepsilon,m}\right) = \lim_{k\to\infty} \underbrace{\mathbb{P}\left(\bigcap_{m=n}^{k} A_{\varepsilon,m}\right)}_{\text{by independence}}$

Facts (continuity of probabilities):

if $B_n \downarrow B$ and $B = \bigcap_n B_n$

$\mathbb{P}(B) = \lim_n \mathbb{P}(B_n)$

if $B_n \uparrow B$ $\quad B = \bigcup_n B_n$

$\mathbb{P}(B) = \lim_n \mathbb{P}(B_n)$

$$= \prod_{m=n}^{k} \left(1 - \frac{1}{m}\right)$$

$$= \lim_{k\to\infty} \prod_{m=n}^{k} \left(1 - \frac{1}{m}\right)$$

$$\leq \lim_{k\to\infty} \exp\left\{- \sum_{m=n}^{k} \frac{1}{m}\right\}$$

$1 - x \leq e^{-x}$

$$= 0$$

because $\quad \sum_{m=1}^{\infty} \frac{1}{m} = \lim_{n\to\infty} \underbrace{\sum_{m=1}^{n} \frac{1}{m}}_{\sim \log n} \sim \lim_{n\to\infty} \log n = \infty$

So, $\quad \mathbb{P}(|X_n| < \varepsilon \text{ eventually}) \leq \sum_{n=1}^{\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_{\varepsilon,m}\right)$

$$= \lim_{k\to\infty} \sum_{n=1}^{k} \underbrace{\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_{\varepsilon,m}\right)}_{=0}$$

$$= 0$$

This is the proof of

Borel-Cantelli second Lemma If $\{A_n\}$ is a collection of independent events and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ then

$$\mathbb{P}\left(\limsup_n A_n\right) = 1$$

cumulative distribution function

Application :

→ van der Vaart Thm 19.1

Glivenko - Cantelli Let $X_1, \dots, X_n \overset{iid}{\sim}$ from a distribution over $\mathbb{R}$ with c.d.f. F. Let $\hat{F}_n$ be the empirical cdf $x \in \mathbb{R} \longmapsto \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \leq x\}$

[Of course $n \hat{F}_n(x) \sim \text{Bin}(n, F(x))$ so $\hat{F}_n(x) \xrightarrow{(*)} F(x)$ $* =$ wp1 and $\xrightarrow{P}$ by LLN]

$$\| \hat{F}_n - F \|_\infty = \sup_{x \in \mathbb{R}} | \hat{F}_n(x) - F(x) | \xrightarrow{wp\ 1} 0$$

Empirical cdf is a strong estimator of the entire cdf F !!