SDS 387
Linear Models

Fall 2024

Lecture 3 - Tue, Sep 3, 2024

Instructor: Prof. Ale Rinaldo

- HW 1 : will be posted today. I will add more problems as we cover more problem

- Last time : Glivenko Cantelli Theorem (van der Vaart Thm 19.1)

$X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ from some distribution over $\mathbb{R}$ with c.d.f. F_X [remember that the cdf F_X is defined as:

$$x \in \mathbb{R} \longmapsto F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$$

Then F_X satisfies the properties:

i) it is non-decreasing $x \leq y \Rightarrow F_X(x) \leq F_X(y)$

ii) $\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to \infty} F_X(x) = 1$

iii) right-continuous : $\forall x \in \mathbb{R}$
$$\lim_{y \downarrow x} F_X(y) = F_X(x)$$

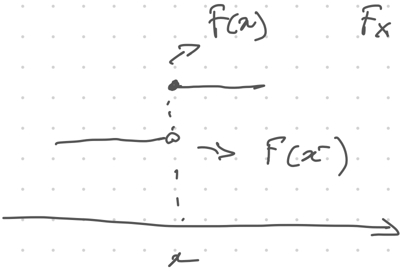
iv) it has left-limits : $\forall x \in \mathbb{R}$

$\lim_{y \uparrow x} F_X(x)$ exists and is denoted with

$F_X(x^-)$

In general, if F_X is discontinuous at x

$F_X(x^-) < F_X(x)$



v) F_X can have at most countably many points of discontinuity. HW

Consider the empirical cdf $\hat{F}_n(\cdot)$:

$$x \longmapsto \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \leq x\}$$

$\hat{F}_n$ is a cdf with $\hat{F}_n(x^-) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i < x\}$

- For each $x \in \mathbb{R}$, $n\hat{F}_n(x) \sim Bin(n, F_X(x))$

$$n\hat{F}_n(x^-) \sim Bin(n, F_X(x^-))$$

- By SLLN $\hat{F}_n(x) \xrightarrow{wp\ 1} F_X(x)$ all x

- Glivenko Cantelli Thm says that

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

↳ convergence is uniform in $x \in \mathbb{R}$

- This is non-trivial! If $\hat{F}_n(x_1) \xrightarrow{wp\ 1} F_X(x_1)$ and $\hat{F}_n(x_2) \xrightarrow{wp\ 1} F_X(x_2)$

then we can conclude that

$$\max_{i=1,2} |\hat{F}_n(x_i) - F_X(x_i)| \xrightarrow{wp\ 1} 0$$

In fact, if $\{x_i\}_{i=1,2,3,\ldots}$ is a possibly infinite sequence of points in $\mathbb{R}$ s.t.

$$|\hat{F}_n(x_i) - F_X(x_i)| \xrightarrow{wp\ 1} 0 \quad \text{all } i$$

then

$$\sup_{x_i} |\hat{F}_n(x_i) - F_X(x_i)| \xrightarrow{wp\ 1} 0.$$

Why? Because the intersection of countably many events of probability 1 is also an event of probability 1! To see this, let $\{A_n\}_{n=1,2,\ldots}$ be events s.t. $\mathbb{P}(A_n) = 1$ all n.

$$\mathbb{P}\left(\bigcap_n A_n\right) = 1 - \mathbb{P}\left(\left(\bigcap_n A_n\right)^c\right)$$

De Morgan Law

$$= 1 - \mathbb{P}\left(\bigcup_n A_n^c\right)$$

union bound

$$\text{Next } \mathbb{P}\left(\bigcup_n A_n^c\right) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n^c)$$

$$= \lim_{k \to \infty} \underbrace{\sum_{n=1}^{k} \underbrace{\mathbb{P}(A_n^c)}_{=0}}_{=0}$$

$$= 0$$

so $\mathbb{P}\left(\bigcap_n A_n\right) = 1$.

Proof of $\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_X(x) \right| \xrightarrow{wp\ 1} 0$

Let $\varepsilon > 0$. (→ small) $\exists\, k = k(\varepsilon) \in \mathbb{N}$ and points

$$-\infty = x_0 < x_1 < \dots < x_{k-1} < x_k = +\infty \quad \text{s.t.}$$

$$(*) \quad F_X(x_i^-) - F_X(x_{i-1}) < \varepsilon \quad \text{all } i$$

[points at which $F_X(x) - F_X(x^-) > \varepsilon$ are in this set]

For any x s.t. $x_{i-1} \le x < x_i$

$$\hat{F}_n(x) - F_X(x) \le \hat{F}_n(x_i^-) - F_X(x_{i-1})$$

$$\le \hat{F}_n(x_i^-) - F_X(x_i^-) + \varepsilon$$

by $(*)$.

Similarly,

$$\hat{F}_n(x) - F_X(x) \ge \hat{F}_n(x_{i-1}) - F_X(x_{i-1}) - \varepsilon$$

$F_X(x) \le F_X(x_i^-) \le F(x_{i-1}) + \varepsilon$ using $(*)$

So $\forall x \in \mathbb{R}$ $\exists\, (x_i, x_{i-1})$ s.t.

$$|\hat{F}_n(x) - F_X(x)| \le \max\{|\hat{F}_n(x_i^-) - F_X(x_i)|, |\hat{F}_n(x_{i-1}) - F_X(x_{i-1})|\} + \varepsilon$$

So

$$\limsup_{n\to\infty} \sup |\hat{F}_n(x) - F_X(x)| < \varepsilon \quad \text{wp } 1.$$

because $k = k(\varepsilon)$ is finite $\blacksquare$

- This a great result but not very useful because it is not quantitative. Q: "how fast does $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_X(x)| = \|\hat{F}_n - F_X\|_\infty \to 0$?"

- A stronger result is: DKW inequality with explicit constant by Massart

Dvoretzky - Kiefer - Wolfowitz

$$\mathbb{P}(\|\hat{F}_n - F_X\|_\infty \ge \varepsilon) \le 2\exp\{-2n\varepsilon^2\}$$

HW! any $\varepsilon > 0$ exponential prob inequality

Why does this imply Glivenko Cantelli? Because of First Borel Cantelli Lemma

Let $\{A_n\}$ be a sequence of events (not necessarily indep).

If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup A_n) = 0$.

Because $\exp\{-2n\varepsilon^2\}$ is summable in $n = 1, 2, \ldots$

Then $\mathbb{P}(\|\hat{F}_n - F_X\|_\infty > \varepsilon \text{ i.o.}) = 0$

$\Rightarrow \|\hat{F}_n - F_X\|_\infty \xrightarrow{\text{wp } 1} 0$

Pf/ Write $\limsup A_n = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=n}^{\infty} A_m}_{:= C_n} = \bigcap_{n=1}^{\infty} C_n$

Now $\{C_n\}$ is a decreasing sequence $C_n \supseteq C_{n+1}$

So $\mathbb{P}(\bigcap_n C_n) = \lim_n \mathbb{P}(C_n)$ by continuity of probability

Therefore if we can show that $\lim_n \mathbb{P}(C_n) = 0$ the result will follow.

Now $\mathbb{P}(C_n) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \to 0$ as $n \to \infty$

by union bound!

This complete the proof! ∎

- Back to convergence in probability:

$X_n \xrightarrow{P} X$ when $\forall \varepsilon > 0$ $\lim_{n\to\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0$

It is important to notice that this requires handling the joint distribution of X_n and X, for all n.

Example: Let $\{X_n\}$ be a sequence of Bernoulli r.v.'s s.t. $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = \frac{1}{2}\frac{n+1}{n}$

Let $X \sim \text{Bernoulli}(1/2)$

Q: Does $X_n \xrightarrow{P} X$?

A: who knows? It depends on joint of (X_n, X).

Suppose $\underset{\substack{\downarrow \\ \text{independent}}}{X_n \perp\!\!\!\perp X}$ all n. No!

Let $\varepsilon \in (0,1)$. $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X| = 1)$

$$= \frac{1}{4}\frac{n+1}{n} + \frac{1}{4}\frac{n-1}{n}$$

$$= \frac{1}{2}$$

On the other hand suppose that
$\mathbb{P}(X_n = 1 \mid X = 1) = 1$ and $\mathbb{P}(X_n = 1 \mid X = 0) = \frac{1}{n}$

Then $X_n \xrightarrow{P} X$.

First of all, we need to make sure this joint distribution is compatible with marginals. It is!
(For example $\mathbb{P}(X_n = 1) = \frac{1}{2}\frac{n+1}{n}$)

Next, $\varepsilon \in (0,1)$

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(X_n = 1 \mid X = 0)\,\mathbb{P}(X=0) + \mathbb{P}(X_n = 0 \mid X = 1)\,\mathbb{P}(X=1)$$

$$= \frac{1}{2n} \to 0 \quad \text{as} \quad n \to \infty$$

- Another example $X = Z \sim N(0,1)$

$$X_n = (-1)^n Z$$

Then $X_n \overset{d}{=} X$ but $X_n \overset{P}{\not\to} X$

equal in distribution

Lp CONVERGENCE

For a r.v. X (over $\mathbb{R}$) and $p \geq 1$ let

Not random

$$\|X\|_p = \left(\mathbb{E}[|X|^p]\right)^{1/p}$$

be the Lp norm of X. It is a norm over the space of r.v.'s with finite p-moments

[$\|X\|_p = 0$ iff $X = 0$ wp 1

$\|X\|_p \geq 0$]

$\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$

We say that $X_n \xrightarrow{L_p} X$ when

$$\|X_n - X\|_p \to 0 \quad \text{as} \quad n \to \infty$$

where $\{X_n\}$ is a sequence of r.v.'s with p-moments and X has finite p-moment

The most common case is $p = 2$

↳ mean squared error