SDS 387
Linear Models

Fall 2025

Lecture 4 - Tue, Sept 11, 2025

Instructor: Prof. Ale Rinaldo

- Last time, the event

$$A_i = \{ |\hat{F}_n(x_i) - F_X(x_i)| < \varepsilon \}$$

should be

$$A_i = \{ |\hat{F}_n(x_i) - F_X(x_i)| < \varepsilon \text{ eventually} \}$$

as $n \to \infty$

- Then by SLLN $\mathbb{P}(A_i) = 1$ all i

$$\Rightarrow \mathbb{P}\left(\bigcap A_i\right) = \mathbb{P}\left(\left\{\max_i |\hat{F}_n(x_i) - F_X(x_i)| < \varepsilon \text{ eventually}\right\}\right) = 1$$

- Glivenko Cantelli:

empirical cdf

true cdf

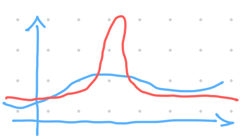
$$\|\hat{F}_n - F_X\|_\infty \xrightarrow{wp\ 1} 0$$

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

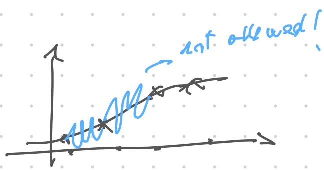
Aside: $f : \mathbb{R}^d \to \mathbb{R}$

The sup-norm of f is

$$\sup_{x \in \mathbb{R}^d} |f(x)|$$



PA/



Let $F = F_X$.

Let $\varepsilon > 0$ arbitrarily small. Then $\exists\, k = k(\varepsilon) \in \mathbb{N}_+$ and a set of points

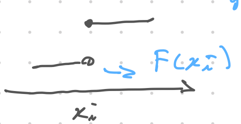
$$-\infty = x_0 < x_1 < \cdots < x_{k-1} < x_k = +\infty$$

s.t.

$$0 \leq F(x_i^-) - F(x_{i-1}) < \varepsilon \quad \text{all } i. \qquad (*)$$

$$\lim_{y \uparrow x_i} F(y)$$

[$F(x_i^-) < F(x_i)$ if there is positive mass at x_i: $\mathbb{P}(X = x_i) > 0$]



- Remark: points x at which $F(x) - F(x-) > \varepsilon$ are among the x_i's.

- Take any $x \in \mathbb{R}$. Then $\exists i$ s.t. $x_{i-1} \leq x < x_i$

Then

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

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

using $(*)$

Similarly

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

using $(*)$ and the fact that

$$F(x) \leq F(x_i^-) \leq F(x_{i-1}) + \varepsilon$$

Therefore, for any x:

$$|\hat{F}_n(x) - F(x)| \leq \max_i \Big\{ |\hat{F}_n(x_i^-) - F(x_i^-)|, \; |\hat{F}_n(x_{i-1}) - F(x_{i-1})| \Big\} + \varepsilon$$

$$= A + \varepsilon$$

as $n \to \infty$ $\quad A \xrightarrow{wp\, 1} 0$ by the arguments discussed last time

$\Downarrow$

$$\limsup_n |\hat{F}_n(x) - F(x)| \lesssim \varepsilon \quad wp\ 1$$

Because $\varepsilon > 0$ is arbitrary this limsup is 0. $\blacksquare$

Remark: this a great result but not a quantitative one.

Q: how fast is this convergence?

- As alluded last time, a more informative result is

DKW inequality $\quad \mathbb{P}(\|\hat{F}_n - F\|_\infty \geq \varepsilon) \leq 2\exp\{-2n\varepsilon^2\}$

See Massart's (1991) paper

- DKW inequality implies Glivenko Cantelli. This follows from Borel-Cantelli's First Lemma.

If $\{A_n\}$ is a sequence of events s.t. $\sum_n \mathbb{P}(A_n) < \infty$

then $\mathbb{P}(\limsup_n A_n) = 0$

$\hookrightarrow$ A_n happens i.o.

- Back to DKW let $A_n = \{\|\hat{F}_n - F\|_\infty \geq \varepsilon\}$. Then

$\mathbb{P}(A_n) \leq 2\exp\{-n2\varepsilon^2\}$ so $\sum_n \mathbb{P}(A_n) < \infty$.

$\hookrightarrow \mathbb{P}(\limsup A_n) = 0$

PF/ Borel–Cantelli's First Lemma.

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=n}^{\infty} A_m}_{B_n} := \bigcap_{n=1}^{\infty} B_n . \quad \text{Next}$$

B_n is a decreasing sequence (i.e. $B_n \supseteq B_{n+1}$)

By continuity of the probability:

$$\mathbb{P}\left(\bigcap_n B_n\right) = \lim_n \mathbb{P}(B_n)$$

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

↓ union bound

because $\sum_n \mathbb{P}(A_n) < \infty$.

↓

$$\mathbb{P}(\limsup_n A_n) = \mathbb{P}\left(\bigcap_n B_n\right) = 0$$

∎

$\lim_{n\to\infty} \mathbb{P}(d(X_n, X) \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$

- Final comment about convergence in probability.

It is important to realize that there needs to be some knowledge about the joint distribution of X_n and X, for all n

Example Consider the sequence $\{X_n\}$ s.t.

$$\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = \frac{1}{2} \quad \forall n$$

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

Q: $X_n \xrightarrow{P} X$? In fact, we cannot answer without further info about the joint distribution of (X_n, X)

Suppose $X_n \perp\!\!\!\perp X$. Then $X_n \xrightarrow{P}\!\!\!\!\!/\;\; X$.

independent

Because:

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X| = 1)$$

$$\varepsilon \in (0,1) \qquad = \frac{1}{2}\underbrace{\frac{1}{2}\frac{n+1}{n}}_{\mathbb{P}(X_n=1)} + \frac{1}{2}\underbrace{\frac{1}{2}\frac{n-1}{n}}_{\mathbb{P}(X_n=0)} = \frac{1}{2}$$

- On the other hand, assume:

$$\mathbb{P}(X_n = 1 \mid X = 1) = 1 \quad \text{and} \quad \mathbb{P}(X_n = 1 \mid X = 0) = \frac{1}{n}$$

[Aside: these conditionals are compatible with the marginals]
check!

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

$\varepsilon \in (0,1)$

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

- Convergence in prob. may not be as natural as we think!

$$X_n = Z \sim N(0,1) \qquad X = -Z \sim N(0,1)$$

all n

Then $X_n \xrightarrow{P}\!\!\!\!\!/\;\; X$ even though

$$X_n \stackrel{d}{=} X$$

equality in distribution

- L_p CONVERGENCE

For a univariate random variable X and $p \geq 1$ let

Lp norm of X

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

↳ this is a "norm":

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

$\|X\|_p \geq 0$

$\|aX\|_p = |a| \, \|X\|_p$ $a \in \mathbb{R}$

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

↳ triangle ineq

Remark if $p < 1$ this is not a norm as it fails triangle ineq.

Lp convergence $X_n \xrightarrow{L_p} X$ when

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

The case of $p = 2$ is by far the most common.

For example, in stats, if θ is a parameter of interest and $\hat{\theta}_n$ an estimator of θ, the mean squared error is

$$\|\hat{\theta}_n - \theta\|_2^2 = \mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right] = \underbrace{\left(\mathbb{E}[\hat{\theta}_n] - \theta\right)^2}_{\text{squared bias } \hat{\theta}_n} + \underbrace{\mathbb{E}\left[\left(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]\right)^2\right]}_{\text{variance of } \hat{\theta}_n}$$

- We can also take $p = \infty$!

$$\|X\|_\infty = \inf\{a : \mathbb{P}(X > a) = 0\}$$

essential supremum of X

- $\|X\|_p \uparrow \|X\|_\infty$ as $p \to \infty$

Properties of L_p norms:

i) $X, Y \in L_p$ $\quad X + Y \in L_p$

use the following fact (C_r inequality):

$$|x+y|^p \leq \begin{cases} |x| + |y| & 0 < p < 1 \\ 2^{p-1}\left(|x|^p + |y|^p\right) & p \geq 1 \end{cases}$$