SDS 387
Linear Models

Fall 2025

Lecture 3 - Tue, Sept 9, 2025

Instructor: Prof. Ale Rinaldo

- HW 1 is out! It is on Canvas.

- Last time: convergence wp 1 vs convergence in probability

  $\Downarrow$

  in statistics, if $\{\hat{\theta}_n\}$ is a sequence of estimators of a parameter θ, then

  strong consistency $\leftarrow$ $\hat{\theta}_n \xrightarrow{wp\ 1} \theta$ and $\boxed{\hat{\theta}_n \xrightarrow{P} \theta}$

  are examples of consistency of $\hat{\theta}_n$.

- Law of large numbers (see e.g. Chapter 4 of Ferguson)

  If $\{X_n\}$ is a sequence of independent r.v.'s in $\mathbb{R}$ s.t. $\mathbb{E}[X_n] = \mu_n \in \mathbb{R}$. Then

  Assuming $\sum_{i=1}^{n} \frac{\mu_i}{n} = O(1)$

$$\lim_{n \to \infty} \sum_{i=1}^{n} \frac{(X_i - \mu_i)}{n} = 0 \begin{cases} \text{wp } 1 & \text{Strong LLN} \\ \text{in prob.} & \text{Weak LLN} \end{cases}$$

Typically $\mathbb{E}[X_n] = \mu$ all n, so the LLN becomes

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[]{\substack{\text{wp } 1 \\ \text{or} \\ P}} \mu$$

sample average of $X_1, \dots, X_n$

- For independent random vectors $\{X_n\}$ in say $\mathbb{R}^d$ (using $\|\cdot\|$ ← Euclidean norm) with $\mathbb{E}[X_n] = \mu \in \mathbb{R}^d$ all n

$$X_n \xrightarrow[]{\substack{\text{wp } 1 \\ \text{or} \\ P}} \mu \quad \text{iff} \quad X_n(j) \xrightarrow[]{\substack{\text{wp } 1 \\ \text{or} \\ P}} \mu(j)$$

j^{th} coordinate of X_n, $j = 1, \dots, d$

- <u>Remark</u>: here asymptotics is in n (sample size) but not in n and d!

- <u>Remark</u>: Strong LLN: proof is tricky.

Weak LLN: assuming existence of second moments (variance!), WLLN follows from Chebyshev's inequality:

Let $\{X_n\}$ be a sequence in $\mathbb{R}^d$ s.t.

$$X_i \sim (\mu_i, \Sigma_i)$$

↳ $d \times d$ covariance matrix

Then $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim (\bar{\mu}_n, \bar{\Sigma}_n)$ where $\bar{\Sigma}_n = \frac{1}{n^2}\sum_{i=1}^{n}\Sigma_i$

(if $\Sigma_i = \Sigma$ all i, $\bar{\Sigma}_n = \frac{\Sigma}{n}$)

To show that

$\bar{X}_n \xrightarrow{P} \bar{\mu}_n$ let's look at

arbitrarily small

$$\mathbb{P}(\|\bar{X}_n - \bar{\mu}_n\| \geq \varepsilon) \leq \frac{\mathbb{E}[\|\bar{X}_n - \bar{\mu}_n\|^2]}{\varepsilon^2}$$

Euclidean norm

$\|x\|^2 = \sum_{j=1}^{d} x(j)^2$

Chebyshev ineq.

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[X^2]}{\varepsilon^2}$$

by independence

$$= \frac{\mathbb{E}\left[\sum_{j=1}^{d}(\bar{X}_n(j) - \bar{\mu}_n(j))^2\right]}{\varepsilon^2}$$

Exercise

$\operatorname{tr}(\underset{d\times d}{A}) = \sum_{j=1}^{d} A(j,j)$

$$\underset{(*)}{=} \frac{\operatorname{tr}(\bar{\Sigma}_n)}{\varepsilon^2} \to 0 \text{ if } \operatorname{tr}(\bar{\Sigma}_n) \to 0 \text{ as } n \to \infty$$

- If $\Sigma_i = \Sigma$ all i then

$$\operatorname{tr}(\bar{\Sigma}_n) = \frac{\operatorname{tr}(\Sigma)}{n}$$

- <u>Remark</u>: We used independence in a fundamental way! in $(*)$

In HW you will study a situation in which the r.v.'s in the sequence are correlated.

Asymptotics statistics

Glivenko Cantelli Theorem [Thm 19.1 in Van der Vaart's book]

Let $X_1, X_2, \dots, X_n \overset{iid}{\sim}$ some distribution on $\mathbb{R}$, with c.d.f F_X. Remember that the cdf of a r.v. X is the function:

$$x \in \mathbb{R} \longmapsto F_X(x) = \mathbb{P}(X \le x)$$

i) it is non-decreasing $x \le x \to F_X(x) \le F_X(y)$

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

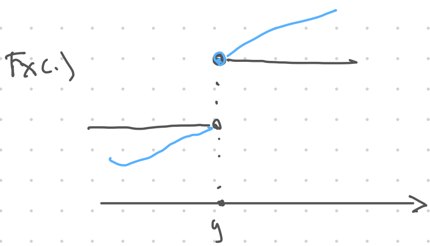
iii) right-continuous

$$\lim_{x \downarrow y} F_X(x) = F_X(y)$$

càdlàg

iv) has left limits:

$\lim\limits_{x \uparrow y} F_X(x)$ exists written as $F_X(y^-)$



- Points y at which

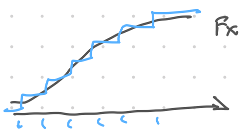
$$F_X(y^-) \ne F_X(y)$$

are points of discontinuity of F_X

<u>Fact</u>: the set of discontinuities of F_X is countable

- Given $X_1, \dots, X_n \overset{iid}{\sim} F_X$ we would like to estimate F_X. We can use the empirical cdf: it is a <u>random</u> function

$\hat{F}_n$ is piecewise constant



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

Natural estimator

$$\mathbb{E}\left[\hat{F}_n(x)\right] = \mathbb{P}(X \leq x) = F_X(x)$$

In fact:

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

and if x is a discontinuity point of F_X

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

- By SLLN, for any fixed $x \in \mathbb{R}$,

$$\hat{F}_n(x) \xrightarrow{\text{wp } 1} F_X(x)$$

$$\hat{F}_n(x^-) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i < x\} \xrightarrow{\text{wp } 1} F_X(x^-)$$

$\hookrightarrow$ we can estimate F_X pointwise use SLLN

- A much more valuable result is this:

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

$\downarrow$ sup norm distance

- This is not trivial!! It is non-trivial because it requires control over <u>uncountably</u> many values of x.

If $\{x_n\}$ is a possibly infinite sequence of points in $\mathbb{R}$ s.t.

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

for all i, then

$$\sup_{x_i} \left| F_n(x_i) - F_X(x_i) \right| \xrightarrow{wp\ 1} 0.$$

Why? Because the intersection of countably many events of prob. 1 is also an event of prob. 1 ! i.e. if $\mathbb{P}(A_n) = 1$ all n then $\mathbb{P}(\bigcap_n A_n) = 1$

$\Updownarrow$

if $\mathbb{P}(A_n) = 0$ all n

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

In detail,

$$\mathbb{P}\left(\bigcap_n A_n\right) = 1 - \mathbb{P}\left(\left(\bigcap_n A_n\right)^c\right)$$ De Morgan's Law

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

Next $\mathbb{P}\left(\bigcup_n A_n^c\right) \leq \sum_n \mathbb{P}(A_n^c)$

union bound (aka sub-additivity)

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

$$= 0$$

$\hookrightarrow \mathbb{P}(\bigcap_n A_n) = 1$

Thus, letting $A_i = \{ |\hat{F}_n(x_n) - F_X(x_n)| < \varepsilon, \text{ eventually} \}$

Then $\bigcap_n A_n = \{ \max_{x_i} |F_n(x_n) - F_X(x_i)| < \varepsilon$ eventually $\}$

- Proof of Glivenko - Cantelli: next time

- A stronger result: DKW inequality with explicit constant due to Massart

  DKW = Dvoretzky - Kiefer - Wolfowitz

$$\boxed{\mathbb{P}\left(\| \hat{F}_n - F_X \|_\infty \geq \varepsilon \right) \leq 2 \exp\{ -2n\varepsilon^2 \}}$$

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

ε: arbitrarily small >0

finite sample ineq. (no asymptotics!)

- DKW implies Glivenko Cantelli!