SDS 387
Linear Models

Fall 2025

Lecture 5 - Tue, Sept 16, 2025

Instructor: Prof. Ale Rinaldo

- HW 1, Q2 (a): we need extra assumption, e.g. $|X_n| \leq a$ for all n, with probability 1. (boundedness assumption)

  $\downarrow$ now bonus problem!

  It can be relaxed, e.g. by assuming a finite 4^{th} moment and using Paley - Zygmund.

- In the proof of Glivenko - Cantelli, we set

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

$\hookrightarrow$ $\exists\, n_i$ s.t. the inequality is true for all $n \geq n_i$

$\hookrightarrow$ random!

Since $\mathbb{P}(A_i) = 1$ all i,

$$\mathbb{P}\Big(\underbrace{\bigcap_i A_i}\Big) = 1$$

$$\{ \forall i \;\; \exists \overset{\text{random}}{n_i} \text{ s.t. } |\hat{F}_n(x_i) - F_X(x_i)| < \varepsilon \text{ all } n \geq n_i \}$$

If there are finitely many i, take $n^* = \max n_i$
random

- Last time: L_p convergence

$X_n \xrightarrow{L_p} X$ means $\|X_n - X\|_p \to 0$ as $n \to \infty$

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

$p \geq 1$

$p=2$ is the most common, gives mean squared (or risk) convergence

- Properties of L_p spaces (the L_p space is the space of all r.v.'s X s.t. $\|X\|_p < \infty$)

1) $X, Y \in L_p$ $\quad X+Y \in L_p$ $\qquad$ C_r-inequality

because for every $x, y \in \mathbb{R}$

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

2) Minkowski:

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

3) Hölder inequality:

if $X \in L_p$ and $Y \in L_q$ where $p, q \geq 1$ are conjugate $\frac{1}{p} + \frac{1}{q} = 1$

then

$XY \in L_1$ and

$$\|XY\|_1 \le \|X\|_p \, \|Y\|_q$$

When $p = q = 2$ this is known as Cauchy-Schwartz inequality

$$\mathbb{E}\left[|XY|\right] \le \sqrt{\mathbb{E}[X^2]}\,\sqrt{\mathbb{E}[Y^2]}$$

4) if $1 \le p \le q$ then $\|X\|_p \le \|X\|_q$

Proof uses another important inequality:

<u>Jensen's inequality</u>: if f is a convex function then

$$\mathbb{E}\left[f(X)\right] \ge f\left(\mathbb{E}[X]\right)$$

$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$

$\alpha \in [0,1]$

assuming $\mathbb{E}[|X|] < \infty$

if f is concave

$$\mathbb{E}\left[f(X)\right] \le f\left(\mathbb{E}[X]\right)$$

So

$$\mathbb{E}\left[|X|^p\right] = \mathbb{E}\left[|X|^{p \underbrace{\frac{q}{p}\frac{p}{q}}_{=1}}\right]$$

$$\le \left(\mathbb{E}\left[|X|^{p\frac{q}{p}}\right]\right)^{p/q}$$

by Jensen because $x \mapsto |x|^{p/q}$ is concave
$x \ge 0$

$$= \mathbb{E}\left[|X|^q\right]^{p/q}$$

and we are done!

Recall last time we define the L_∞ norm of X as

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

essential supremum

You can show that

$$\|X\|_p \uparrow \|X\|_\infty \quad \text{as} \quad p \to \infty$$

Exercise

- Back to convergence: L_p converges is stronger (i.e. it implies) convergence in probability:

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\|X_n - X\|_p^p}{\varepsilon^p}$$

arbitrarily small $\varepsilon > 0$

Markov's ineq

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

$X \geq 0$ r.v.
$\varepsilon > 0$

- In general, L_p convergence and $\xrightarrow{wp\ 1}$ do not imply each other. Examples: let $U \sim \text{Uniform}(0,1)$

1)
$$X_n = f_n(U) = \begin{cases} 0 & 0 \leq U < \frac{1}{n} \\ \frac{1}{U} & \frac{1}{n} \leq U \leq 1 \end{cases}$$

Then

$$X_n \xrightarrow{wp\ 1} \frac{1}{U} \quad \text{but} \quad \frac{1}{U} \notin L_p \quad \text{any } p \geq 1$$

2)
$$X_n = g_n(U) = \begin{cases} n & 0 \leq U < \frac{1}{n} \\ 0 & \frac{1}{n} \leq U \leq 1 \end{cases}$$

$$X_n \xrightarrow{wp\ 1} 0 \quad \text{but} \quad \|X_n\|_p^p = n^{p-1} \to \infty$$

CONVERGENCE IN DISTRIBUTION OR WEAK CONVERGENCE

- Weakest form of stochastic convergence.

<u>Def</u> Let $\{X_n\}$ be a sequence of random variables and X a random variable taking values in $\mathbb{R}$. Let F_{X_n} and F_X be the cdf of X_n and X, respectively.

X_n converges in distribution to X, $X_n \xrightarrow{d} X$ or $X_n \rightsquigarrow X$, when for every $x \in \mathbb{R}$ s.t. F_X is continuous at x,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

<u>Remarks</u> : i) this notion does not impose any restriction on the joint distribution of $\{X_n\}$ and X. So, recall the examples from last time:

$X_n \sim \text{Bernoulli}\left(\frac{1}{2}\frac{n+1}{n}\right)$

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

$X_n \xrightarrow{d} X$ but we cannot say anything about convergence in prob.

Similarly if $X = Z \sim N(0,1)$ and

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

then $\quad X_n \xrightarrow{d} X \quad$ because $X_n \overset{d}{=} X$

but $\quad X_n \overset{p}{\not\to} X$

2) $X_n \overset{d}{=} X$ means that X_n and X have the same distribution and so, from the point of view of convergence in distribution, they are identical. But their realization can be very different.

3) The restriction that $F_{X_n}(x) \longrightarrow F_X(x)$ for a continuity point of F_X is necessary!

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

Example:

$$X_n = \begin{cases} \begin{cases} 1 - \frac{1}{n} & 1/2 \\ 0 - \frac{1}{n} & 1/2 \end{cases} & n \text{ odd} \\ \begin{cases} 1 + \frac{1}{n} & 1/2 \\ 0 + \frac{1}{n} & 1/2 \end{cases} & n \text{ even} \end{cases}$$

with prob

$$X_n \xrightarrow{d} X \sim \text{Bernoulli}(1/2)$$

Take $x = 1$, not a continuity point of F_X

Then $\quad F_{X_n}(1) = \begin{cases} 1 & n \text{ odd} \\ 1/2 & n \text{ even} \end{cases} \quad$ so

$F_{X_n}(1)$ does not converge.

4) Convergence in distribution is quite general.

$$F_{X_n}(x) = \begin{cases} 0 & x < -n \\ \dfrac{\Phi(x) - \Phi(-n)}{\Phi(n) - \Phi(-n)} & -n \le x \le n \\ 1 & x > n \end{cases}$$

Φ cdf of $N(0,1)$

$$X_n \xrightarrow{d} Z \sim N(0,1)$$

- Recall that a cdf F is a function s.t.

  1) non-decreasing and non-negative

  2) right-continuous $\lim_{y \downarrow x} F(y) = F(x)$

     with left limits $\lim_{y \uparrow x} F(y)$ exists $(F(x-))$

  3) $\lim_{y \to -\infty} F(y) = 0$ $\lim_{y \to \infty} F(y) = 1$

Then F is the cdf of a random variable say X s.t. for any $a < b$

$$\mathbb{P}(a < X \le b) = F(b) - F(a)$$
$$\mathbb{P}(X \in (a,b])$$

cdf's are representations of prob. measures.

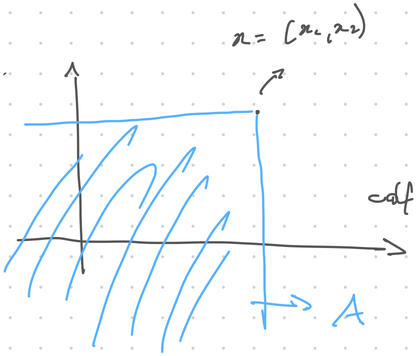
- In $\mathbb{R}^d$, if X is a random vector then one can define its cdf as

$$F_X(x) = \mathbb{P}(X \leq x) \qquad x \in \mathbb{R}^d$$

elementwise inequality:

$$X(j) \leq x(j) \quad \forall\, j = 1, \ldots, d$$

$$= \mathbb{P}\left(\bigcap_{j=1}^{d} \{ X(j) \leq x(j) \} \right)$$



cdf at x is $\mathbb{P}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in A \right)$

It is not difficult to show that F_X satisfies properties 1), 2) and 3) above provided that $\leq$ is taken element-wise

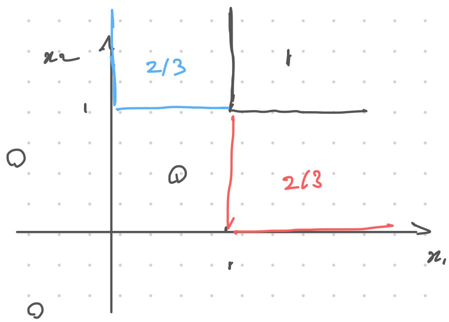
$\left(\text{i.e. if } x = \begin{bmatrix} x(1) \\ \vdots \\ x(d) \end{bmatrix} \quad y = \begin{bmatrix} y(1) \\ \vdots \\ y(d) \end{bmatrix} \right.$

$x \leq y$ means $x(j) \leq y(j)$ all j $\Big)$

- A function on $\mathbb{R}^d$ that satisfies properties 1), 2) and 3) however does not necessarily define a prob. distribution!

Example :

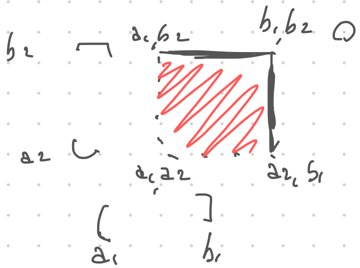
$$F(x_1, x_2) = \begin{cases} 1 & x_1, x_2 \geq 1 \\ 2/3 & x_1 \geq 1 \quad 0 \leq x_2 < 1 \\ 2/3 & x_2 \geq 1 \quad 0 < x_1 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



Let $A = (a_1, b_1] \times (a_2, b_2]$

If X has cdf F

Then



$$\mathbb{P}(X \in A) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$$

If F is a cdf the above quantity ≥ 0

But take $a_1 = a_2 = 1 - \varepsilon \rightarrow$ small

$b_1 = b_2 = 1$

The expression above evaluates to

$$1 - \frac{2}{3} - \frac{2}{3} + 0 = -\frac{1}{3} < 0$$