SDS 387
Linear Models

Fall 2024

Lecture 1 - Tue, Sep 27, 2024

Instructor: Prof. Ale Rinaldo

RECAP OF DETERMINISTIC CONVERGENCE & NOTATION

- We will be working in Euclidean spaces, i.e. $\mathbb{R}^d$ ($d=1$ or higher, but <u>fixed</u>).
  ↳ NO HIGH-DIM MODELS for most of the class

- In $\mathbb{R}^d$, we have an inner product: $x = \begin{bmatrix} x(1) \\ \vdots \\ x(d) \end{bmatrix} \in \mathbb{R}^d$ and $y \in \mathbb{R}^d$ then
$$\langle x, y \rangle = x^T y = y^T x = \sum_{j=1}^{d} x(j)\, y(j)$$

- Euclidean norm: $x \in \mathbb{R}^d$
$$\|x\| = \sqrt{x^T x} = \sqrt{\sum_{j=1}^{d} x^2(j)}$$

- We can use other norms $\left(\text{e.g. } \|x\|_p = \left(\sum_{j=1}^{d} |x(j)|^p \right)^{1/p} \right.$ $p \geq 1$

$$\left. \|x\|_\infty = \max_j |x(j)| \right)$$

- In general, much of what we say is applicable to <u>metric spaces</u> : $(\mathcal{X}, d)$

set

distance function. $\left\{ \begin{array}{l} d: \mathcal{X} \times \mathcal{X} \to [0, \infty] \\ d(x,x) = 0 \\ d(x,y) = d(y,x) \\ d(x,y) \le d(x,z) + d(z,y) \end{array} \right.$

$(\mathbb{R}^d, \|\cdot\|)$

↳ Euclidean norm $d(x,y) = \|x - y\|$

- Let $\{x_n\}_{n=1,2,\dots}$ be a sequence of points in $(\mathcal{X}, d)$ and $x^* \in \mathcal{X}$. Then $x_n \to x^*$ as $n \to \infty$ when $\lim_{n \to \infty} d(x_n, x^*) = 0$

↓ DETERMINISTIC CONVERGENCE

<u>Question</u> : what if we have a sequence of random variables: how do we define convergence?

<u>Notation</u> for Euclidean sequences.

Let $\{r_n\}$ be a sequence of positive numbers and $\{x_n\}$ be a sequence of points in $\mathbb{R}^d$.

big-oh • $x_n = O(r_n) \iff \exists\, C > 0$ s.t. $\frac{\|x_n\|}{r_n} \le C \quad \forall n$

↳ $x_n = O(1)$ means $\{x_n\}$ is <u>bounded</u>

little-oh • $x_n = o(r_n) \iff \forall \varepsilon > 0 \quad \exists N$ s.t. $\frac{\|x_n\|}{r_n} \le \varepsilon \quad \forall n > N$

small

↳ $x_n = o(1)$ means $x_n \to 0$

big-omega

- $x_n = \Omega(r_n) \iff \exists c > 0$ s.t. $\frac{\|x_n\|}{r_n} \geq c \quad \forall n$

$x_n = \Omega(1)$ means ??

little omega

- $x_n = \omega(r_n) \iff \forall M > 0 \quad \exists N$ s.t. $\frac{\|x_n\|}{r_n} \geq M \quad \forall n > N$ (large)

$x_n = \omega(1)$?

- $x_n = \Theta(r_n) \iff$ $x_n = O(r_n)$, $x_n = \Omega(r_n)$

$x_n = \Theta(1)$?

STOCHASTIC CONVERGENCE

- Suppose we have a sequence $\{X_n\}$ of random vectors in $\mathbb{R}^d$.

Q: How do we define convergence ?

Example $(d=1)$ $Z \sim N(0,1)$ $X_n = (-1)^n Z$

Does $\{X_n\}$ converge ?

$(-1)^n Z \sim N(0,1)$ by rotational invariance

Notation: for an event A (a collection of possible outcomes)

$\mathbb{P}(A)$ is the probability that A occurs.

Example $Z \sim N(0,1)$ $A = \{|Z| > 1.96\}$

Almost sure / Almost everywhere } Convergence (AKA. convergence with probability 1)

Let $\{X_n\}$ be a sequence of r.v.'s and X another r.v. (possibly degenerate). Then

$$X_n \xrightarrow{a.s.} X \quad \text{or} \quad X_n \xrightarrow{a.e.} X \quad \text{or} \quad X_n \xrightarrow{wp\ 1} X$$

when

$$\mathbb{P}\left(\left\{\lim_{n\to\infty} d(X_n, X) = 0\right\}\right) = 1$$

distance, e.g. $\|X_n - X\|$

The probability of a realization of all $\{X_n\}$ and of X s.t. $\lim_n d(X_n, X)$ does not exist is 0.

- Very <u>strong</u> notion of stochastic convergence that requires you to have a handle on the joint distribution of $\{X_n\}$ and X.

- $X_n \xrightarrow{wp\,1} X$ means that $\forall \varepsilon > 0$ small

$$\mathbb{P}\left(\limsup_n d(X_n, X) < \varepsilon\right) = 1$$

$\hookrightarrow$ $d(X_n, X) < \varepsilon$ eventually with prob 1

$\exists N$ (random) s.t. $d(X_n, X) < \varepsilon$ $\forall n > N$

equivalently

$$\mathbb{P}\left(\limsup_n d(X_n, X) > \varepsilon\right) = 0$$

$\hookrightarrow$ $d(X_n, X) > \varepsilon$ infinitely often with prob. zero.

- liminf and limsup of events

Let $A_{\varepsilon,n}$ where $\varepsilon > 0$ be the event

$$\{ d(X_n, X) < \varepsilon \}$$

Then $X_n \xrightarrow{wp\,1} X$ iff $\forall \varepsilon > 0$

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

or

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