SDS 387
Linear Models

Fall 2025

Lecture 9 - Tue, Sept 30, 2025

Instructor: Prof. Ale Rinaldo

- Last time: Taylor series expansion of multivariate function f:

kth order Taylor series expansion

$$\boxed{f(x) = f(x_0) + \sum_{j=1}^{k} \frac{1}{j!} D^{(j)} f(x_0, x - x_0) + \text{Rem}}$$

where

$$D^{(j)} f(x_0, x - x_0) = \sum_{i_1, \ldots, i_j} \frac{\partial^j}{\partial x_{i_1} \cdots \partial x_{i_j}} f(x_0) h_{i_1} h_{i_2} \cdots h_{i_j}$$

where $h = x - x_0$ and $h_i = h(i)$ $i = 1, \ldots, d$

and Rem is such that $\text{Rem} = o(\|x - x_0\|^k)$

and

Lagrangean i) $\text{Rem} = \frac{1}{(k+1)!} D^{(k+1)} f(z, x - x_0)$ for some z on the line connecting x and x_0

Integral

iii) $$\text{Rem} = \frac{1}{k!} \int_0^1 (1-u)^k \, D^{(k)} f\big(\underbrace{x_0 + u(x - x_0)}_{ux + (1-u)x_0}, x - x_0\big) \, du$$

- Last time: Cramer Wald device: $\{X_n\}$ sequence of r.v.'s in $\mathbb{R}^d$ and X a r.v. in $\mathbb{R}^d$ then

$$X_n \xrightarrow{d} X \quad \text{iff} \quad \underbrace{t^T X_n}_{\langle t, X_n \rangle = \sum_{j=1}^{d} X_n(j)\, t(j)} \xrightarrow{d} t^T X \quad \text{for all choices of } t \in \mathbb{R}^d$$

HW

random variable in $\mathbb{R}$

- We also saw that if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ then $\begin{bmatrix} X_n \\ Y_n \end{bmatrix}$ needs not to converge. And as a result $f(X_n, Y_n)$ needs not to converge, even if f is well-behaved (e.g. continuous).

That is, marginal convergence in distribution does not imply joint convergence!

Exception if $X_n \perp\!\!\!\perp Y_n$ all n then $\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ Y \end{bmatrix}$

independent

HW
use ch. functions

Result: If $X_n \xrightarrow{d} X$ and $Y_n - X_n \xrightarrow{P} 0$ then

$$Y_n \xrightarrow{d} X$$

Pf/ Let C be any closed set. We want to show that

$$\limsup_n \mathbb{P}(Y_n \in C) \leq \mathbb{P}(X \in C)$$

[Portmanteau Thm ...]

For any $\varepsilon > 0$ ↦ small

$$\{Y_n \in C\} = \left(\{Y_n \in C\} \cap \{\overbrace{d(X_n, Y_n)}^{\|X_n - Y_n\|} \leq \varepsilon\}\right) \cup \left(\{Y_n \in C\} \cap \{d(X_n, Y_n) > \varepsilon\}\right)$$

$$A = (A \cap B) \cup (A \cap B^c)$$

any A and B

$$\subseteq \{d(X_n, C) \leq \varepsilon\} \cup \{d(X_n, Y_n) > \varepsilon\}$$

$$d(x, C) = \inf_{y \in C} d(x, y)$$

point in $\mathbb{R}^d$ closed set in $\mathbb{R}^d$

Therefore:

$$\mathbb{P}(Y_n \in C) \leq \mathbb{P}(X_n \in C_\varepsilon) + \underbrace{\mathbb{P}(d(X_n, Y_n) > \varepsilon)}$$

also closed

$\to 0$ as $n \to \infty$ because $X_n - Y_n \xrightarrow{P} 0$

where $C_\varepsilon = \{x \in \mathbb{R}^d : d(x, C) \leq \varepsilon\}$

$$\limsup_n \mathbb{P}(Y_n \in C) \leq \limsup_n \mathbb{P}(X_n \in C_\varepsilon)$$

$$\leq \mathbb{P}(X \in C_\varepsilon)$$

by Portmanteau Thm because $X_n \xrightarrow{d} X$

$\Rightarrow \quad \limsup_n \mathbb{P}(Y_n \in C) \leq \mathbb{P}(X \in C_\varepsilon)$

Now let $\varepsilon \downarrow 0 \quad \mathbb{P}(X \in C_\varepsilon) \downarrow \mathbb{P}(X \in C)$

So by letting $\varepsilon \downarrow 0$

$$\limsup_n \mathbb{P}(Y_n \in C) \leq \mathbb{P}(X \in C)$$

$$\Updownarrow$$

by Portmanteau $\quad Y_n \xrightarrow{d} X \quad \blacksquare$

<u>Corollary</u> $\quad X_n \xrightarrow{d} X \quad Y_n \xrightarrow{p} c$, c constant. Then

HW!

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ c \end{bmatrix}$$

- <u>Slutsky Theorem</u> $\quad X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, c a constant

Then:

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} cX$$

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c} \qquad c \neq 0$$

Analogous results hold for random vectors and random matrices

If $\underset{d \times d}{Y_n} \xrightarrow{p} \underset{d \times d}{C}$ and $X_n \xrightarrow{d} X$ in $\mathbb{R}^d$

Then

$$Y_n X_n \xrightarrow{d} CX$$

Example $X_1, X_2, \ldots \overset{iid}{\sim} (\mu, \sigma^2)$. Then

$$\underbrace{\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma}}_{\sim (0,1)} \xrightarrow{d} N(0,1) \quad \text{by CLT}$$

more refined result than $\bar{X}_n \xrightarrow{P} \mu$ by WLLN

You can argue that

$\bar{X}_n \pm z_{1-\alpha/2}\, \sigma/\sqrt{n}$ is an asymptotic $1-\alpha$ CI for μ

$1-\alpha/2$ upper quantile of a $N(0,1)$

- What if we don't know σ? We can estimate it using sample variance

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

To show that $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$ [i.e. $\hat{\sigma}_n^2$ is a consistent estimator of σ^2] let's do the following quick calculation:

some algebra

$$\hat{\sigma}_n^2 = \underbrace{\frac{n}{n-1} \underbrace{\left[\underbrace{\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2}_{\xrightarrow{P} E[(X-\mu)^2] = \sigma^2 \text{ by WLLN}} - \underbrace{(\bar{X}_n - \mu)^2}_{\xrightarrow{P} 0 \text{ by CMT}} \right]}_{\xrightarrow{P} \sigma^2 \text{ by Slutsky}}}_{\xrightarrow{P} \sigma^2 \text{ by Slutsky}}$$

$$\sqrt{n}\,\frac{(\bar{X}_n-\mu)}{\hat{\sigma}_n} = \sqrt{n}\,\frac{(\bar{X}_n-\mu)}{\sigma\ \underbrace{\hat{\sigma}_n/\sigma}_{\xrightarrow{p} 1}} \xrightarrow{d} N(0,1)$$

by Slutsky again

t-statistic

O_p and o_p

= "big oh p" = "little oh-p"

see e.g. van der Vaart Chapter 2 or Dave's Hunter notes on large sample theory

Let $\{X_n\}$ be a sequence of r.v.'s.

stochastic smaller order

$X_n = o_p(1) \iff X_n \xrightarrow{p} 0$

$X_n = o_p(R_n)$ means $X_n = Y_n R_n$ where $Y_n = o_p(1)$

$\{R_n\}$ deterministic or random positive numbers

stochastically bounded

$X_n = O_p(1)$ means that $\{X_n\}$ is bounded in probability

$\forall \varepsilon > 0$ (small) $\exists M = M(\varepsilon)$ and an $N = N(\varepsilon)$

s.t. $\forall n \geq N$

$\mathbb{P}(\|X_n\| > M) \leq \varepsilon$

$\Updownarrow$

$\forall \varepsilon > 0$ small $\exists M = M(\varepsilon)$ s.t.

$\mathbb{P}(\|X_n\| > M) \leq \varepsilon$

for all n!

6

$X_n = O_p(R_n)$ means $X_n = Y_n R_n$ where $Y_n = O_p(1)$

R_n is deterministic or random sequence > 0

Example $X_1, X_2, \ldots \overset{iid}{\sim} (\mu, \sigma^2)$ then

$$\bar{X}_n = \mu + o_p(1) \quad \text{by WLLN}$$

$$\bar{X}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{by CLT}$$

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{d} N(0,1)$$

$\downarrow$ more informative

because $O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1)$

O_p/o_p calculus:

$$o_p(1) \pm o_p(1) = o_p(1)$$

$$O_p(1) \pm o_p(1) = O_p(1)$$

$$O_p(1)\, o_p(1) = O_p(o_p(1)) = o_p(1)$$

$$(1 + o_p(1))^{-1} = O_p(1)$$

$$\frac{1}{O_p(1)} \quad \text{who knows ?}$$

Of course $o_p(1)$ is $O_p(1)$

Remark: If $X_n = O_p(1)$ does it mean $X_n \xrightarrow{d}$?

No ! It only means that it is bounded in probability. $\leftrightarrow$ Tightness

Prokhorov's Thm

i) if $X_n \xrightarrow{d} X$ then $X_n = O_p(1)$

ii) if $X_n = O_p(1)$ then $\exists \{n_k\}$ s.t. $X_{n_k} \xrightarrow{d} X$ some r.v. X

$X_n = O_p(1)$ iff $\exists \{n_k\}$ s.t. $X_{n_k} \xrightarrow{d} X$ some r.v. X.

SDS 387
Linear Models

Fall 2025

Lecture 10 - Thu, Oct 2, 2025

Instructor: Prof. Ale Rinaldo

CLT (Central Limit Theorem)

$\mathbb{E}[(X-\mu)(X-\mu)^T]$
$= \mathbb{E}[XX^T] - \mu\mu^T$

infinite sequence

<u>Basic form</u> Let $X_1, X_2, \ldots \overset{iid}{\sim} (\mu, \Sigma)$. Then

$d \times d$

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{d} N(0, \Sigma)$$

$$\Updownarrow$$

$$\frac{\Sigma^{-1/2}}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow{d} N(0, I_d)$$

↳ identity matrix.

$\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i =$

$Y_i = \Sigma^{-1/2}(X_i - \mu)$
$\sim (0, I_d)$

Normalized sum of iid $(0, I_d)$ variables

- Another way to think about this is the following:

let $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0, I_d)$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \sim N(0, I_d)$$

The CLT says that