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

Basic form Let $X_1, X_2, \ldots \overset{iid}{\sim} (\mu, \Sigma)$. Then
$d \times d$

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

$\Updownarrow$

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

↳ identity matrix.

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

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Sigma^{-1/2}(X_i - \mu)$$

behaves just like $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$ for n large enough

- Basic idea of universality: replace the X_i's with $\Sigma^{1/2}(Z_i + \mu)$ $\sim N(\mu, \Sigma)$

(Assume wlog that $\mu = 0$)

Pf/ We use characteristic functions. Let $\varphi(t) = \mathbb{E}\left[\exp\{i\, t^T X\}\right]$, $t \in \mathbb{R}^d$, be the ch.f. of $X - \mu \sim (0, \Sigma)$. Then

ch.f. of $\sqrt{n}(\bar{X}_n - \mu)$

$$\varphi_{\sqrt{n}(\bar{X}_n - \mu)}(t) = \varphi_{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)}(t) = \varphi_{\sum_{i=1}^{n} (X_i - \mu)}(t/\sqrt{n})$$

$$= \prod_{i=1}^{n} \varphi_{X_i - \mu}(t/\sqrt{n})$$

$$= \left(\varphi(t/\sqrt{n})\right)^n$$

Next, we will do a Taylor series expansion of $\varphi(t/\sqrt{n})$ around $\varphi(0) = 1$. Because the first 2 moments of X exist

$$\nabla \varphi(0) = i\, \mathbb{E}[X - \mu] = 0$$

Hessian

dxd of second order partial derivatives

$$\nabla^2 \varphi(0) = i^2 \Sigma = -\Sigma$$

So, by Taylor series expansion

$$\left(\varphi(t/\sqrt{n})\right)^n = \Big(\underbrace{1}_{\varphi(0)} + \underbrace{\frac{1}{\sqrt{n}} t^T \nabla \varphi(0)}_{=0} + \frac{1}{n} t^T \Big(\int_0^1 (1-\tau) \nabla^2 \varphi(\tau\, t/\sqrt{n})\, d\tau \Big) t \Big)^n$$

Recall that $(1 + a_n)^n \to \exp\left\{\lim_n n a_n\right\}$ if $n a_n$ has a limit

For us

$$a_n = \frac{1}{n} t^T \left(\int_0^1 (1-\tau) \nabla^2 \varphi\left(\tau\, t/\sqrt{n}\right) d\tau \right) t$$

Next

$$\left(\varphi\left(t/\sqrt{n}\right)\right)^n \to \exp\left\{ \lim a_n n \right\}$$

$$= \exp\left\{ - \frac{t^T \Sigma_1 t}{2} \right\}$$

because

$$\lim_{n\to\infty} \int_0^1 (1-\tau) \nabla^2 \varphi\left(\tau\, t/\sqrt{n}\right) d\tau =$$

$$\Sigma_1 \int_0^1 (1-\tau)\, d\tau = - \frac{\Sigma_1}{2}$$

$\downarrow$

We have shown that, for every $t \in \mathbb{R}^d$,

$$\varphi_{\sqrt{n}(\bar{X}-\mu)}(t) \to \exp\left\{ - \frac{t^T \Sigma t}{2} \right\} \quad \text{as } n \to \infty$$

which is the ch. f. of $N(0, \Sigma)$

$\downarrow$

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} N(0, \Sigma) \quad \blacksquare$$

- The same CLT guarantee holds if X_i's are only independent. In this case we need to consider a triangular array:

$$\begin{array}{ccc} X_{1,1} & & \\ X_{2,1} & X_{2,2} & \\ X_{3,1} & X_{3,2} & X_{3,3} \\ \vdots & & \end{array}$$

<u>Assumption</u>: the rows of this triangular array contains independent r.v.

$\vdots$

$X_{n,1} \quad X_{n,2} \quad X_{n,3} \quad \cdots \quad X_{n,n}$

$\vdots$

The Lindeberg–Feller CLT (Univariate case)

Let $\{X_{n,j}\}$ be an infinite triangular array s.t.

Not a limitation if $E[X_{n,j}] = \mu_{n,j}$ replace $X_{n,j}$ with $X_{n,j} - \mu_{n,j}$

$E[X_{n,j}] = 0$ and $Var[X_{n,j}] = \sigma^2_{n,j}$.

Let $$S_n = \sum_{j=1}^{n} X_{n,j}$$

$$B_n^2 = \sum_{j=1}^{n} \sigma^2_{n,j}$$

Then
$$\frac{S_n}{B_n} \xrightarrow{d} N(0,1)$$

provided that

(LF condition) $$\forall \varepsilon > 0 \quad \frac{1}{B_n^2} \sum_{j=1}^{n} E\left[X_{n,j}^2 \, 1\{|X_{n,j}| > \varepsilon B_n\}\right] \to 0$$
as $n \to \infty$

- Conversely if $\frac{S_n}{B_n} \xrightarrow{d} N(0,1)$ and if

uniform asymptotic negligibility

$$\frac{\max_{j=1}^{n} \sigma^2_{n,j}}{B_n^2} \to 0 \quad \text{as } n \to \infty$$

Then the LF condition holds.

- Often, instead of checking the LF condition, it may be easier to check the following stronger condition:

Lyapunov's condition $\Leftarrow$ $\frac{1}{B_n^{2+\delta}} \sum_{j=1}^{n} \mathbb{E}\left[|X_{n,j}|^{2+\delta} \right] \to 0$ some $\delta > 0$

requires existence of moments of order $2+\delta$

The multivariate case of LF - CLT.

Consider an infinite triangular array of centered d-dimensional random vectors $X_{n,j}$ $j \le n$ s.t. $\mathrm{Var}[X_{n,j}]$ exists

Let

$$Y_{n,j} = \left(\sum_{\ell=1}^{n} \mathrm{Var}[X_{n,\ell}] \right)^{-1/2} X_{n,j}$$

If

(LF) $$\lim_{n \to \infty} \sum_{j=1}^{n} \mathbb{E}\left[\| Y_{n,j} \|^2 \mathbb{1}\{ \|Y_{n,j}\| \ge \varepsilon \} \right] = 0 \quad \forall \varepsilon > 0$$

Then $$\sum_{j=1}^{n} Y_{n,j} \xrightarrow{d} N(0, I_d)$$