SDS 387
Linear Models

Fall 2025

Lecture 8 - Thu, Sept 25, 2025

Instructor: Prof. Ale Rinaldo

- Last time: CMT

$\{X_n\}$ sequence of r.v.'s in $\mathbb{R}^d$ s.t. $X_n \xrightarrow[d]{\overset{\text{wp } 1}{p}} X$

and $f: \mathbb{R}^d \to \mathbb{R}^p$ s.t. $\mathbb{P}(X \in C) = 1$

where C is the set of continuity points of f,

then $f(X_n) \xrightarrow[d]{\overset{\text{wp } 1}{p}} f(X)$

Example

$X_1, \dots, X_n \overset{iid}{\sim} (\mu, \sigma^2)$ then

we will see that

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

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

$$T_n^2 \xrightarrow{d} \chi^2_1$$

Characteristic functions

- Powerful analytic approach to characterize - convergence and more.

- See Ferguson, chapter 3, and Section 2.3 of Van der Vaart's book

  see lecture notes on Asymptotic Statistics by David Hunter at Penn State

- For a r.v. X in $\mathbb{R}^d$, the characteristic function of X (actually, of the distribution of X) is the function

$$t \in \mathbb{R}^d \to \varphi_X(t) = \mathbb{E}\left[\exp\{\overset{\sqrt{-1}}{i}\, \underbrace{t^T X}_{\sum_{i=1}^d t(i) X(i)}\}\right]$$

Results:

1) $X_n \xrightarrow{d} X$ iff $\varphi_{X_n}(t) \to \varphi_X(t)$ $\forall t \in \mathbb{R}^d$

Moreover, if $\varphi_{X_n}(t)$ converges pointwise (i.e. for each t separately) to a function φ continuous at 0 then φ is the ch. f. of the r.v. X s.t. $X_n \xrightarrow{d} X$

2) Uniqueness

$X \overset{d}{=} Y$ iff $\varphi_X(t) = \varphi_Y(t)$ $\forall t \in \mathbb{R}^d$

If X has mean μ then $\nabla \varphi_X(0) = i\mu$

φ_X is diff at zero

- $Z \sim N(\mu, \underset{d\times d}{\Sigma})$

$\hookrightarrow$ $\varphi_Z(t) = \exp\left\{ i t^T \mu - \dfrac{t^T \Sigma t}{2} \right\}$ $\forall t \in \mathbb{R}^d$

- Application: WLLN

$X_1, X_2, \ldots \overset{iid}{\sim} \mu$ and chf φ_X. Then

$\bar{X}_n \xrightarrow{P} \mu$

To show this, we will just do Taylor expansion.

Let $f: \mathbb{R}^d \to \mathbb{R}$ have $(k+1)$ continuous partial derivatives in an open set U.

Let $x, x_0 \in U$ s.t. the line segment connecting them is also in U. Then

kth order Taylor series expansion

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

where:

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

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

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

and

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

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

See also Folland's book on advanced calculus

- Mean value theorem :

$$f(x) - f(x_0) = \nabla f(z)^T (x - x_0)$$

some z between x and x_0

$$\nabla f(z) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(z) \\ \vdots \\ \frac{\partial f}{\partial x_d}(z) \end{bmatrix} \in \mathbb{R}^d$$

$$= \left[\int_0^1 \nabla f(x_0 + u(x - x_0)) \, du \right]^T (x - x_0)$$

- Now suppose $f : \mathbb{R}^d \to \mathbb{R}^p$ and let

$Df(x)$: $p \times d$ matrix whose (j,i) entry is $\dfrac{\partial f_j}{\partial x_i}(x)$

Jacobian of f at x

$i = 1, \dots, d$

$j = 1, \dots, p$

Then,

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \underbrace{o(\|x - x_0\|)}$$

p-dim vector
Rem s.t.
$\|\text{Rem}\| = o(\|x - x_0\|)$

In a 2013 paper on The American Statistician titled "The mean value theorem and Taylor series expansion in statistics", the authors remind statisticians that it is not true that

$$f(x) = f(x_0) + Df(z)(x - x_0) \quad \text{some } z$$

Wrong!

- Instead you can write:

$$f(x) - f(x_0) = \left[\int_0^1 Df\left(x_0 + u(x - x_0)\right) du\right](x - x_0)$$

or use p mean value theorems

$$f_j(x) = f_j(x_0) + \nabla f_j^T(z_j)(x - x_0)$$

$j = 1, \dots, p$

z_j is between x and x_0

Back to WLLN's $X_1, X_2, \dots \overset{iid}{\sim} \mu$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \xrightarrow{P} \mu. \quad \text{Equivalently}$$

$$\bar{X}_n \xrightarrow{d} \mu$$

The ch.f. of $\bar{X}_n$ at t is

$$\varphi_{\bar{X}_n}(t) = \varphi_{X_1+X_2+\dots+X_n}(t/n) = \prod_{i=1}^{n} \varphi_{X_i}(t/n)$$

$$\mathbb{E}\exp\left\{ i \frac{t^T}{n} \sum X_i \right\}$$

$$= \left(\varphi(t/n)\right)^n \qquad \varphi = \varphi_{X_i} \text{ all } i$$

Taylor series expansion around 0

$k=1$

$(1-u)^{k-1} = 1$

$x_0 = 0$

$x = t/n$

$$= \left(\varphi(0) + \int_0^1 \frac{t^T}{n} \nabla\varphi(u\, t/n)\, du\right)^n$$

differentiable at zero because mean exists

Now use the fact $(1+a_n)^n \to \exp\{\lim_n n a_n\}$ if limit exists

$(1+\frac{1}{n})^n \to e$

Dominated convergence Thm: $\to$

$$\exp\left\{\lim_{n\to\infty} \cancel{n}\, \frac{t^T}{\cancel{n}} \int_0^1 \nabla\varphi(v \cdot t/n)\, dv\right\}$$

$X_n \to X$ wp 1

$|X_n| \leq Y$ $\mathbb{E}[Y]$ exists

$\lim_n \mathbb{E}[X_n] = \mathbb{E}[X]$ ↳ $\lim X_n$

I am going the limit inside the integral

6

$$= \exp\left\{ t^T \int_0^1 \underbrace{\lim_{n\to\infty} \nabla \varphi(t_{r_n} v)}_{i\mu}\, dv \right\}$$

$$= \exp\{ i\, t^T \mu \}$$

$\Downarrow$

ch. f. of a point mass at μ (the mean)

$X_n \xrightarrow{p} c$

$\Updownarrow$

$X_n \xrightarrow{d} c$

$\hookrightarrow \bar{X}_n \xrightarrow{d} \mu$

$\hookrightarrow \bar{X}_n \xrightarrow{p} \mu$

- (Cramer – Wold device)

Let $\{X_n\}$ be a sequence of r.v.'s in $\mathbb{R}^d$.

Then $X_n \xrightarrow{d} X$ iff

$t^T X_n \xrightarrow{d} t^T X \quad \forall t \in \mathbb{R}^d$

$\Downarrow$

univariate

- Remember the example $U \sim \text{Uniform}(0,1)$

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} = \begin{cases} \begin{bmatrix} U \\ U \end{bmatrix} & n \text{ even} \\ \begin{bmatrix} U \\ 1-U \end{bmatrix} & n \text{ odd} \end{cases}$$

Then $X_n \xrightarrow{d} U \quad Y_n \xrightarrow{d} U$ but $\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \not\xrightarrow{d}$

Let $t = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ then

$$t^T \begin{bmatrix} X_n \\ Y_n \end{bmatrix} = \begin{cases} 2V & n \text{ even} \\ 1 & n \text{ odd} \end{cases}$$

This example shows also that if X_n and Y_n converge in distribution marginally then $X_n + Y_n$ needs <u>not</u> to converge in distribution.