Teorema central del límit

Diplomatura d'Estadística

Estadística Matemàtica I

Jordi Ocaña Rebull

Departament d'Estadística

Divisió de Ciències Experimentals i Matemàtiques

Punts que tractarem:

- Concepte de teorema central del límit.
- Teorema de Lindeberg i Levy. Casos

Idea general del Teorema Central del Límit (T.C.L.)

- Direm que una successió de v.a. $\{X_n\}$ verifica el T.C.L. sii existeixen successions de constants $\{a_n\}$ i $\{b_n\}$ tals que la v.a. suma

$$S_n = \sum_{i=1}^{n} X_i$$

verifica

$$\frac{S_n - a_n}{b_n} \xrightarrow{d} Z \sim N(0,1)$$

UNIVERSITAT DE BARCELONA

Teorema de Lindeberg i Lévy

- Si les X_n són iid, amb esperança i variància finites, μ i σ^2 respectivament, tenim $E(S_n) = n\mu, \quad \mathrm{var}(S_n) = n\sigma^2$ i aleshores

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

- A la pràctica podrem fer l'aproximació

$$S_n \approx N\left(n\mu, \sigma\sqrt{n}\right)$$

UNIVERSITAT DE BARCELONA

Teorema de Lindeberg i Lévy

casos particulars

- Teorema de Laplace-De Moivre:

$$X_n \sim b(n,p) \Rightarrow \frac{X_n - np}{\sqrt{npq}} \xrightarrow{d} Z \sim N(0,1)$$

$$X_n = Y_1 + \ldots + Y_n,\ \ Y_i \sim b(1,p), iid$$

- Mitjana aritmètica de n observacions X_i independents d'una mateixa v.a. X:

$$\bar{X}_n = \sum_{i=1}^{n} Y_i,\ \ Y_i = \frac{1}{n} X_i,\ E\left(\bar{X}_n\right) = \mu,\ \ \mathrm{var}\left(\bar{X}_n\right) = \frac{\sigma^2}{n}$$

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

UNIVERSITAT DE BARCELONA

Teorema de Lindeberg i Feller

- Si les X_n són solament independents, amb esperances i variàncies finites, μ_n i σ_n^2 respectivament:

$$\frac{\sum_{i=1}^{n}(X_i - \mu_i)}{b_n} \xrightarrow{d} Z \sim N(0,1), \quad b_n^2 = \sum_{i=1}^{n}\sigma_i^2$$

- Condició: $\lim_{n\to\infty} \frac{1}{b_n^2}\sum_{i=1}^{n} \int_{|x-\mu_i|>\varepsilon b_i} (x-\mu_i)^2\, dF_i(x) = 0$

$$\left(\Rightarrow \lim_{n\to\infty} \max_{1\le i\le n} \frac{\sigma_i}{b_n} = 0\right)$$

Altres casos i generalitzacions del T.C.L.

- T. Liapunov: existència de moments de tercer ordre i condició basada en ells.
- Funcions qualsevol (no forçosament suma)

$$g(X_1,\dots,X_n) = g(\mu_1,\dots,\mu_n) + \sum_{i=1}^{n} c_i (X_i - \mu_i) + R_n$$

$$g(\mu_1,\dots,\mu_n) \quad \text{constant}, \qquad c_i = \frac{\partial g}{\partial X_i}\Big|_{X_i=\mu_i} \quad \text{constant}$$

R_n v.a., generalment $R_n \xrightarrow{P} 0$

- Sumes de vectors aleatoris: quan estudiem la normal multivariant.

UNIVERSITAT DE BARCELONA

Altres casos de convergència en distribució importants

- Convergència de la binomial a la Poisson.
- Convergència de la hipergeomètrica a la binomial: $H(N,M,n) \xrightarrow[\substack{N\to\infty \\ \frac{M}{N}\to p}]{} b(n,p)$
- Convergència de la khi-quadrat amb n graus de llibertat a la normal

$$\sqrt{2\chi_n^2} - \sqrt{2n-1} \xrightarrow{d} Z \sim N(0,1)$$

UNIVERSITAT DE BARCELONA

Versió multivariant del teorema de Laplace-de Moivre

a la pràctica podem aproximar una multinomial mitjançant una normal

- Sigui $\mathbf{X}_n \sim M(n, \mathbf{p})$
- Definim la versió estandarditzada del vector multinomial anterior

$$\mathbf{Z}_n = \begin{pmatrix} \dfrac{X_1 - np_1}{\sqrt{np_1 q_1}} \\ \vdots \\ \dfrac{X_k - np_k}{\sqrt{np_k q_k}} \end{pmatrix}$$

- Tenim que: $\mathbf{Z}_n \xrightarrow{d} \mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathrm{P})$

UNIVERSITAT DE BARCELONA

Teorema de Lindeberg i Lévy
plantejament

- Tenim una successió de vectors aleatoris *iid*

$$\{\mathbf{X}_n : n \geq 1\} \qquad \mathbf{X}_n = \begin{pmatrix} X_{1n} \\ X_{2n} \\ \vdots \\ X_{kn} \end{pmatrix}$$

- amb esperança i variància finites

$$E(\mathbf{X}_n) = \mu \qquad \operatorname{cov}(\mathbf{X}_n) = \Sigma \qquad \operatorname{corr}(\mathbf{X}_n) = \mathrm{P}$$

UNIVERSITAT DE BARCELONA

Teorema de Lindeberg i Lévy
plantejament

- Definim els vectors aleatoris "suma"

$$\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n = \begin{pmatrix} X_{11} + \cdots + X_{1n} \\ \vdots \\ X_{k1} + \cdots + X_{kn} \end{pmatrix} = \begin{pmatrix} S_1 \\ \vdots \\ S_k \end{pmatrix}$$

- i atès que

$$E(\mathbf{S}_n) = n\mu \qquad \operatorname{cov}(\mathbf{S}_n) = n\Sigma$$

- definim la "suma estandaritzada"

$$\mathbf{W}_n = \begin{pmatrix} \dfrac{S_{1n} - n\mu_1}{\sigma_1\sqrt{n}} \\ \\ \dfrac{S_{kn} - n\mu_\kappa}{\sigma_k\sqrt{n}} \end{pmatrix}$$

UNIVERSITAT DE BARCELONA

Teorema de Lindeberg i Lévy

- Aleshores la suma estandaritzada convergeix en llei a una normal multivariant

$$\mathbf{W}_n \xrightarrow{d} \mathbf{W} \sim \mathrm{N}(\mathbf{0}, \mathrm{P})$$

- A la pràctica podem aproximar la distribució d'un vector de sumes S_n mitjançant una distribució normal multivariant

$$\mathbf{S}_n \approx \mathrm{N}(n\mu, n\Sigma)$$