Probabilidad II
Tema 6: Convergencia en distribución y teorema central del límite

José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

Estructura de este tema

- Convergencia débil y convergencia en distribución
- Teorema de continuidad de Lévy
- TCL para v.a. independientes e idénticamente distribuidas
- TCL de Lindeberg-Feller. Arreglos triangulares.

Convergencia débil

Definición
Sea E un espacio métrico y sean $\mathrm{P}, \mathrm{P}_1, \mathrm{P}_2, \ldots$ medidas de probabilidad sobre $(E, \mathcal{B}(E))$. Se dice que P_n **converge débilmente** a P si $\mathrm{P}_n(A) \to \mathrm{P}(A)$, para todo $A \in \mathcal{B}(E)$ con $\mathrm{P}(\partial A) = 0$, donde ∂A denota la frontera de A.

Teorema (portmanteau)
Las siguientes afirmaciones son equivalentes:

(1) $\int g dP_n \to \int g dP$, para toda función $g : E \to \mathbb{R}$ continua y acotada.

(2) $\limsup P_n(C) \leq P(C)$, para todo $C \in \mathcal{B}(E)$ cerrado.

(3) $\liminf P_n(A) \geq P(A)$, para todo $A \in \mathcal{B}(E)$ abierto.

(4) P_n converge débilmente a P.

Convergencia débil

Proposición
Si $E = \mathbb{R}$, la convergencia débil y la convergencia en distribución son equivalentes.

Teorema de la aplicación continua
Sean $X, X_1, X_2, \ldots$ variables aleatorias tales que $X_n \xrightarrow{d} X$ y sea $g : \mathbb{R} \to \mathbb{R}$ una función continua. Entonces, $g(X_n) \xrightarrow{d} g(X)$.

Teorema de Slutsky
Sean $X, X_1, X_2, \ldots$ variables aleatorias tales que $X_n \xrightarrow{d} X$ y sean $Y_1, Y_2, \ldots$ variables aleatorias tales que $Y_n \xrightarrow{P} a$, donde $a \in \mathbb{R}$. Entonces $X_n + Y_n \xrightarrow{d} X + a$, $X_n Y_n \xrightarrow{d} aX$ y $X_n / Y_n \xrightarrow{d} X/a$, si $a \neq 0$.

Observación: Si el límite de Y_n es una v.a. no degenerada el resultado no es cierto en general.

Notación $O_p(a_n)$ y $o_p(a_n)$

Definición

Una sucesión de v.a. X_n es **acotada en probabilidad** si para todo $\epsilon > 0$ existen $M > 0$ y $N \geq 1$ tales que $\mathrm{P}(|X_n| \leq M) > 1 - \epsilon$ para todo $n \geq N$.

Proposición

Si $X_n \xrightarrow{\mathrm{d}} X$ para alguna v.a. X, entonces la sucesión X_n es acotada en probabilidad.

Notación: Sea a_n una sucesión de números reales.

$X_n = O_p(a_n)$ significa que X_n/a_n es acotada en probabilidad.
$X_n = o_p(a_n)$ significa que $X_n/a_n \xrightarrow{\mathrm{P}} 0$.

Proposición

Si $X_n = O_p(1)$ e $Y_n = o_p(1)$, entonces $X_n Y_n = o_p(1)$.

Teorema de continuidad de Lévy

Teorema
Sea F_n una sucesión de funciones de distribución sobre $\mathbb{R}$ y sea φ_n la correspondiente sucesión de funciones características.

(1) Si $F_n \xrightarrow{d} F$, entonces $\varphi_n(t) \to \varphi(t)$, para todo $t \in \mathbb{R}$, donde φ es la función característica de F.

(2) Si existe $\lim_{n\to\infty} \varphi_n(t)$, para todo $t \in \mathbb{R}$ y $\varphi(t) = \lim_{n\to\infty} \varphi_n(t)$ es una función continua en cero, entonces φ es la función característica de una función de distribución F y $F_n \xrightarrow{d} F$.

Para estudiar la convergencia en distribución de una sucesión de variables aleatorias basta estudiar la convergencia de la sucesión de sus funciones características.

La demostración, por ejemplo, en Resnick, pags. 307-312.

Ejemplos

(a) Si X tiene distribución de Poisson de parámetro λ, estudia el comportamiento asintótico de X/λ cuando $\lambda \to \infty$.

(b) **LDGN**: Si $X_1, X_2, \ldots$ son v.a.i.i.d. con $\mathrm{E}|X| < \infty$. Entonces,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\mathrm{P}} \mu,$$

donde $\mu = \mathrm{E}(X_1)$.

Recordamos: Sea X una v.a. tal que $\mathrm{E}|X|^n < \infty$. Sea φ la f.c. de X. Entonces,

$$\left| \varphi(t) - \sum_{k=0}^{n} \frac{(it)^k \mathrm{E}(X^k)}{k!} \right| \leq \mathrm{E}\left[\min\left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right].$$

TCL para v.a. independientes e idénticamente distribuidas

Teorema
Sea $X_1, X_2, \ldots$ una sucesión de v.a. independientes e idénticamente distribuidas con $0 < \sigma^2 < \infty$, donde $\sigma^2 = \mathrm{Var}(X_1)$, y $S_n = X_1 + \cdots + X_n$. Entonces,

$$\frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Observación: Si $\bar{X}_n = S_n/n$ y $\mu = \mathrm{E}(X_1)$, entonces

$$\frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Además, el teorema implica

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2).$$

Método delta

Se conoce como **método delta** un resultado que permite determinar la distribución límite de funciones suaves de sucesiones de v.a. cuyo comportamiento asintótico es conocido.

Proposición

Sea X_n una sucesión de v.a. tales que $a_n(X_n - c) \stackrel{d}{\to} X$, donde $0 \leq a_n \uparrow \infty$, $c \in \mathbb{R}$ y X es una v.a. Sea $g : \mathbb{R} \to \mathbb{R}$ una función derivable con derivada continua. Entonces $a_n(g(X_n) - g(c)) \stackrel{d}{\to} g'(c)X$.

Ejemplo

Sea $X_1, X_2, \ldots$ una sucesión de v.a.i.i.d. con distribución de Poisson de parámetro λ.

- ¿Cuál es la distribución asintótica de $\sqrt{n}(\overline{X}_n - \lambda)$?
- Determina una función g tal que la varianza de la distribución asintótica de $\sqrt{n}(g(\overline{X}_n) - g(\lambda))$ no depende de λ. Se dice que g es una **transformación que estabiliza la varianza**.

Arreglos triangulares

Consideramos un **arreglo triangular** de v.a. $(X_{n,k})$. Es decir,

$$\begin{array}{cccc} X_{1,1} & X_{1,2} & \cdots & X_{1,r_1} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,r_2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,r_n} \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

Teorema de Lindeberg-Feller: *Una variable que es el resultado de la suma de muchos efectos independientes entre sí sin que ninguno domine al total es aproximadamente normal.*

El teorema de Lindeberg-Feller

Sea $(X_{n,k})$ un arreglo triangular de v.a. cumpliendo:

(1) Para todo n, $X_{n,1}, X_{n,2}, \ldots, X_{n,r_n}$ son independientes y con varianzas finitas no todas nulas.

Si $\mathrm{E}X_{n,k} = \mu_{n,k}$, $\mathrm{Var}(X_{n,k}) = \sigma^2_{n,k}$, $S_n = X_{n,1} + \cdots + X_{n,r_n}$,
$s_n^2 = \mathrm{Var}(S_n) = \sigma^2_{n,1} + \cdots + \sigma^2_{n,r_n}$.

(2) **Condición de Lindeberg (CL)**: Para todo $\epsilon > 0$, se tiene

$$\frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{\{|X_{n,k}-\mu_{n,k}| \geq \epsilon s_n\}} (X_{n,k} - \mu_{n,k})^2 \, d\mathrm{P} \xrightarrow{n\to\infty} 0.$$

Entonces,

$$\frac{S_n - \mathrm{E}(S_n)}{s_n} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1).$$

Una aplicación: estimadores de mínimos cuadrados

Modelo de regresión lineal simple

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, n.$$

con $\epsilon_1, \epsilon_2, \ldots$ v.a.i.i.d. con $\mathrm{E}(\epsilon_1) = 0$ y $\mathrm{Var}(\epsilon_1) = \sigma^2$.

Estimador de mínimos cuadrados de la pendiente

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n c_{n,i} Y_i}{\sum_{i=1}^n c_{n,i}^2}.$$

Pendiente

$$\beta = \frac{\sum_{i=1}^n c_{n,i} \mu_i}{\sum_{i=1}^n c_{n,i}^2}, \quad \mu_i = \mathrm{E}(Y_i) = \beta_0 + \beta_1 x_i.$$

Diferencia entre parámetro y estimador

$$\hat{\beta} - \beta = \frac{\sum_{i=1}^n c_{n,i} \epsilon_i}{\sum_{i=1}^n c_{n,i}^2} = \sum_{i=1}^n X_{n,i}, \quad X_{n,i} := \frac{c_{n,i} \epsilon_i}{\sum_{i=1}^n c_{n,i}^2}.$$