

Teorema Central del Límite

Teorema Central del Límite

- ▶ *Teorema Central de Límite* o (Teorema del Límite Central): conjunto de teoremas con variaciones acerca del comportamiento de la distribución de la suma (o promedio) de variables aleatorias. En ellos se afirma que, bajo ciertas condiciones, la distribución de probabilidad de la suma de un número “grande” de variables aleatorias es, aproximadamente, una distribución normal.
- ▶ Pólya (1920) lo denominó *Teorema “Central” del Límite* por el rol fundamental (central) que cumple este teorema en la Teoría de Probabilidades, ya que, entre otras cosas, justifica porqué en muchas aplicaciones es válido asumir normalidad en las variables o porqué las distribuciones normales son tan comunes.

Un poco de historia

- ▶ Primera versión impresa se debe a De Moivre (ppos siglo 18). En su libro “The Doctrine of Chances” aproxima a la distribución binomial (para el caso especial $p = 1/2$) por una curva “suave” que hoy se conoce como “normal”
- ▶ Científicos de la época, observaron que muchos fenómenos naturales tenían una distribución aproximadamente normal. En 1809 Gauss desarrolló la fórmula de la distribución normal y mostró que ajustaba perfectamente a la distribución de los errores cometidos en las observaciones astronómicas.
- ▶ Laplace generaliza el resultado de De Moivre al caso p arbitrario en lo que hoy se conoce como *Teorema de De Moivre - Laplace*.

- ▶ Teorema Central del Límite de Laplace (s. 19): bajo ciertas condiciones la suma de un número considerable de variables aleatorias mutuamente independientes e idénticamente distribuidas puede aproximarse por una normal. Sólo lo demostró para distribuciones discretas y para ciertas distribuciones continuas.
- ▶ Primeras demostraciones rigurosas: Tshebyshev (1887), Markov(1898) (momentos) y Liapunov (1901) (funciones características). Liapunov estableció además condiciones de suficiencia (*Condiciones de Lyapunov*).
- ▶ Lindeberg (1922), propuso condiciones que son hasta cierto punto, necesarias. Esta demostración se simplifica mediante el uso de funciones características, idea propuesta por Lévy (*Teorema Central del Límite de Lindeberg-Lévy*).
- ▶ En 1937 Feller demuestra el *Teorema Central del Límite de Lindeberg-Feller* donde establece las condiciones necesarias para la validez de este resultado.

Distribuciones Muestrales

Sea X_1, X_2, \dots, X_n muestra aleatoria de v.a.i.i.d. con $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $\forall i = 1, \dots, n$.

1. $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es la **media muestral** de la muestra X_1, X_2, \dots, X_n .
2. $S_n = \sum_{i=1}^n X_i$ es el **total muestral** de la muestra X_1, X_2, \dots, X_n .

- $\mu_{\bar{X}} = E(\bar{X}) = \mu$
- $\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$
- $\mu_{S_n} = E(S_n) = n\mu$
- $\sigma_{S_n}^2 = Var(S_n) = n\sigma^2$

Teorema 9.4: TCL de Lindenberg-Levy

Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas, con $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$. Sea $S_n = X_1 + X_2 + \dots + X_n$. Entonces

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

Demostración Apunte

Para probar el teorema de Lindenberg - Levy necesitaremos el siguiente resultado que es consecuencias de los teoremas de Helly Bray(necesidad) y de Continuidad de Levy (suficiencia)):

Teorema 9.5

Sean X, X_1, \dots v.a y $\phi_X, \phi_{X_1}, \dots$ sus funciones características.
Entonces

$$X_n \xrightarrow{D} X \Leftrightarrow \phi_{X_n}(x) \rightarrow_{n \rightarrow \infty} \phi_X(x)$$

sin demostración

Teorema 9.6 (DeMoivre- Laplace):

Sea S_n el número de éxitos en n ensayos Bernoulli independientes, con probabilidad p de éxito en cada ensayo. Entonces

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0, 1)$$

El teorema de DeMoivre- Laplace dice que si

$$X \sim \mathcal{B}(n, p) \Rightarrow Z = \frac{X - np}{\sqrt{np(1-p)}}$$

tiene *aproximadamente* distribución normal estándar.

La aproximación es buena siempre que:

- ▶ $np > 5$
- ▶ $n(1 - p) > 5$

Ejemplo 1

En la fabricación de ciertos chips semiconductores se produce un 2% de chips defectuosos. Suponga que los chips son independientes entre sí y que se examina un lote que contiene 1000 chips.

¿Cuál es la probabilidad que más de 25 chips sean defectuosos?

Respuesta

Si $X = n^o$ de chips defectuosos en ese lote $\Rightarrow X \sim \mathcal{B}(1000, 0.02)$

$$P(X > 25) = \sum_{k=25}^{1000} \binom{1000}{k} 0.02^k (0.8)^{1000-k}$$

Aproximaciones de la Distribución Normal

Continuación Ejemplo 1

- ▶ $np = 1000 \cdot 0.02 = 20 > 5$
- ▶ $n(1 - p) = 1000 \cdot 0.98 = 980 > 5$
- ▶ Además $\sqrt{np(1 - p)} = 4.43$

$$\Rightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X - 20}{4.43}$$

tiene aproximadamente
distribución $N(0, 1)$

Entonces:

$$P(X > 25) = P\left(\frac{X - 20}{4.43} > \frac{25 - 20}{4.43}\right) \approx P(Z > 1.13) = 1 - \Phi(1.13) = 0.13$$

Corrección de la aproximación por la Distribución Normal

Continuación Ejemplo 1 ¿ Cuál es la probabilidad que se encuentren exactamente 25 chips defectuosos? Es decir,
 $P(X = 25) = ?$

Si quisiéramos aproximar la distribución de una va Binomial (X , discreta) por una va Normal (Z , continua), debemos tener en cuenta que Z es continua, entonces $P(Z = k) = 0$.

Supongamos $X \sim \mathcal{B}(n, p)$, $k \in [0, n]$

$$\begin{aligned} P(X = k) &= P(k - 0.5 < X < k + 0.5) \\ &= P\left(\frac{k-0.5-np}{\sqrt{npq}} < \frac{X-np}{\sqrt{npq}} < \frac{k+0.5-np}{\sqrt{npq}}\right) \\ &\approx \Phi\left(\frac{k+0.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{k-0.5-np}{\sqrt{npq}}\right) \end{aligned}$$

$$\text{Luego, } P(X = 25) \approx \Phi\left(\frac{25+0.5-20}{4.43}\right) - \Phi\left(\frac{25-0.5-20}{4.43}\right) = 0.048$$

La probabilidad exacta, es $P(X = 25) = 0.044$

En general, si se va a aproximar una va X_D discreta, por una va X_C continua debe hacerse una **corrección por continuidad**

Corrección por continuidad

Dados $a, b \in \mathbb{R}$,

| X_D | X_C |
|---------------------|---------------------------|
| $X_D = a$ | $a - 0.5 < X_C < a + 0.5$ |
| $X_D < a$ | $X_C < a - 0.5$ |
| $X_D \leq a$ | $X_C < a + 0.5$ |
| $X_D > a$ | $X_C > a + 0.5$ |
| $X_D \geq a$ | $X_C > a - 0.5$ |
| $a < X_D < b$ | $a + 0.5 < X_C < b - 0.5$ |
| $a \leq X_D < b$ | $a - 0.5 < X_C < b - 0.5$ |
| $a < X_D \leq b$ | $a + 0.5 < X_C < b + 0.5$ |
| $a \leq X_D \leq b$ | $a - 0.5 < X_C < b + 0.5$ |

Continuación Ejercicio 1:

Usar la corrección a la aproximación por una normal y calcular nuevamente $P(X > 25)$.

1. SIN corrección por continuidad,

$$P(X > 25) \approx P(X_N > 25) = 0.13$$

2. CON corrección por continuidad,

$$\begin{aligned} P(X > 25) &\approx P(X_N > 25 + 0.5) \\ &= P(Z > \frac{25.5 - 20}{4.43}) = 1 - \Phi\left(\frac{25.5 - 20}{4.43}\right) = 0.107 \end{aligned}$$

3. Probabilidad Exacta $P(X > 25) = 0.109$

Ejemplo 2:

Cierto equipo consta de 30 instrumentos electrónicos C_1, C_2, \dots, C_{30} que se usan de la siguiente manera: si C_1 falla, entonces comienza a trabajar C_2 , cuando éste falla, recién comienza a trabajar C_3 y así sucesivamente hasta C_{30} . Supongamos que el tiempo de falla $C_i \sim \mathcal{E}(\lambda)$, con $\lambda = 0.1$ por hora. (exponencial de parámetro $\lambda = 0.1$)

¿Cuál es la probabilidad que el equipo dure más de 350h?

Teorema Central del Límite de Lindeberg Sean X_1, X_2, \dots

variables aleatorias independientes e idénticamente distribuidas, con $E(X_i) = \mu_i$, $\text{Var}(X_i) = \sigma_i^2$ tales que $\sigma_i < \infty$ y al menos un $\sigma_i > 0$. Sea $S_n = X_1 + X_2 + \dots + X_n$,

$s_n = \sqrt{\text{Var}(S_n)} = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$. Entonces, si se satisfacen las condiciones de Lindeberg,

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

Condiciones de Lindeberg:

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \frac{1}{s_n} \sum_{k=1}^n \int_{|x-\mu_k| > \epsilon s_n} (x - \mu_k)^2 dF_{X_k}(x) = 0$$

Bibliografía

James B. *Probabilidade: um curso em nível intermediário* IMPA Rio de Janeiro (1983)

Chung, K.L, *A course in Probability Theory*, Academic Press, 2nd Ed.,(1974)