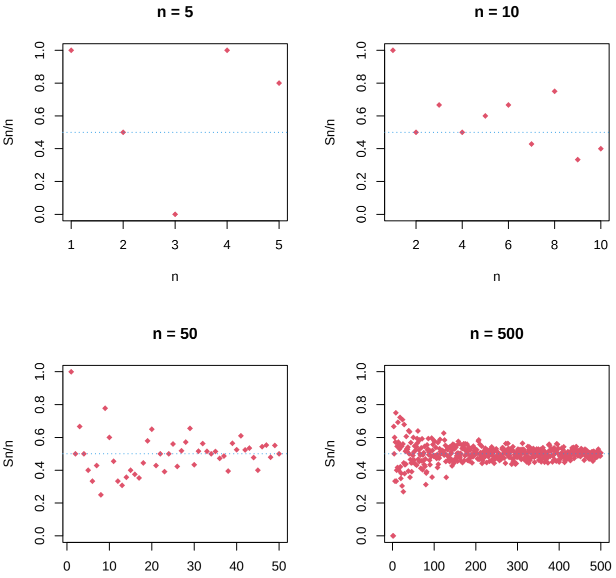
Leyes de los Grandes Números

Teoremas sobre Límites

Experimento: Lanzar una moneda honesta, n veces.
Sea S_n = número "caras" obtenidas en los n lanzamientos.
Intuitivamente (def. frecuentista de la probabilidad):
$\dfrac{S_n}{n} \cong \dfrac{1}{2}$ y $\lim_{n\to\infty} \dfrac{S_n}{n} = \dfrac{1}{2}$.
Este resultado es una consecuencia de una de las Leyes de los Grandes Números.

Estas leyes son una serie de teoremas donde se demuestra cuál es el *límite* o qué propiedades satisfacen en el *límite* combinaciones de v.a.
Las hipótesis planteadas y el tipo de convergencia darán lugar a distintos teoremas.


n = 5
n = 10
n = 50
n = 500
Sn/n
n
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
8
10
0
20
30
40
50
100
200
300
400
500

Tipos de Convergencia

Sean $Y, Y_1, Y_2, \ldots$ v.a. definidas sobre un mismo espacio de probabilidad $(\mathcal{S}, \mathcal{A}, P)$.

- Se dice que Y_n converge a Y en **probabilidad** si $\forall \epsilon > 0$, $P(|Y_n - Y| \geq \epsilon) \to 0$ cuando $n \to \infty$, se denota por $Y_n \xrightarrow{P} Y$.
- Se dice que Y_n converge a Y **casi ciertamente, casi en todo punto** (p.p, a.e) si $P\left(Y_n \to Y\right) = 1$ cuando $n \to \infty$. Esto es, $P(\{\omega : Y_n(\omega) \to Y(\omega)\}) = 1$.
- Sean $Y, Y_1, Y_2, \ldots$ v.a. con funciones de distribución $F, F_1, F_2, \ldots$. Y_n converge en **distribución** a Y cuando $n \to \infty$ si $F_n(x) \to F(x)$ para todo x punto de continuidad de F. Se denota por $Y_n \xrightarrow{D} Y$.

Preliminares

Desigualdad de Chebyshev

Sea X va con $E(X) = \mu$, $Var(X) = \sigma^2$, y sea $a > 0$.
Entonces

$$P\left(|X - \mu| > a\right) \leq \frac{\sigma^2}{a^2}$$

Observaciones

- Esta desigualdad provee una cota para esa probabilidad que sólo depende de la varianza de X.
- Esta cota puede ser grosera, o no informativa (considerar el caso $\sigma^2 > a^2$.

Expresiones equivalentes:

1. $\forall a > 0,\ P(|X - \mu| \leq a) \geq 1 - \dfrac{\sigma^2}{a^2}.$
2. $\forall k \geq 0,\ P(|X - \mu| > k\sigma) \leq \dfrac{1}{k^2}.$
3. $\forall k \geq 1,\ P(|X - \mu| \leq k\sigma) \geq 1 - \dfrac{1}{k^2}.$

Observación: Las dos útimas expresiones nos informan cómo la desviación estandar mide cuan "concentradas" está la distribución alrededor de μ.

Teorema 9.1: Ley Débil de Chebyshev

Sean $X_1, X_2, \ldots$ v.a. independientes, supongamos que $\exists c < \infty$ tal que $\forall n,\ \mathrm{Var(X_n) < c}$. Si $S_n = \sum_{i=1}^{n} X_i$, entonces S_n satisface la **Ley Débil de los Grandes Números**:

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0$$

Ley débil de los grandes números (Teo. de Khintchin)

Teorema 9.2: Teorema de Khintchin

Si $X_1, X_2, \ldots$ son v.a. independientes, idénticamente distribuidas, tales que $E(X_i) = \mu$, $\mathrm{Var(X_i)} = \sigma^2$. Sea $S_n = \sum_{i=1}^{n} X_i$. Entonces

$S_n/n \to \mu$ en probabilidad.

Observación

Este resultado nos dice que el promedio de un número finito de observaciones (realizaciones de las X_i) es una buena aproximación a la estimación de la media poblacional μ.

Corolario 9.1: Teorema de Bernoulli (1713)

Consideremos una sucesión de ensayos Bernoulli independientes, con la misma probabilidad p de éxito en cada ensayo. Sea S_n la variable aleatoria que indica el número de éxitos en los primeros n ensayos. Entonces

$$S_n/n \to p \text{ en probabilidad.}$$

Teorema 9.3: Ley Fuerte de Kolmogorov

Si $X_1, X_2, \ldots$ son v.a. independientes, idénticamente distribuidas, tales que $E(X_i) = \mu$, entonces

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mu \text{ casi ciertamente}$$

(sin demostración)

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 por qué en muchas aplicaciones es válido asumir normalidad en las variables o por qué las distribuciones normales son tan comunes.

Un poco de historia

- La 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.

Teorema 9.4: **TCL de Lindenberg-Levy**

Sean $X_1, X_2, \ldots$ variables aleatorias independientes e indénticamente distribuidas, con $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. Sea $S_n = \sum_{i=1}^{n} X_i$.

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

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: Paul Levy

Sean $X, X_1, \ldots$ v.a y $\phi_X, \phi_{X_1}, \ldots$ sus funciones características. Entonces: $X_n \xrightarrow{D} X \Leftrightarrow \phi_{X_n}(x) \rightarrow_{n\to\infty} \phi_X(x)$

sin demostración

Ejemplo 1:

1. Supongamos que tenemos un libro de 100 páginas. Sea X_i el número de errores de la i-ésima página. Suponga que las variables X_i son i.i.d. con distribución de Poisson con media 1. ¿Cuál es la probabilidad que en el libro se hayan producido menos de 90 errores?
2. La duración T de una bombita de luz, en horas, sigue una distribución normal con $\mu = 1000$ y $\sigma = 100$. Se desea extraer una muestra de manera que la duración promedio permita aproximar a μ con una precisión de al menos 50 horas, con una probabilidad mayor que 0.95.
    1. Halle el tamaño mínimo que debería tener la muestra.
    2. Resuelva el problema, si se desconoce la distribucion de T:
        1. Usando la desigualdad de Chebyshev.
        2. Usando el Teorema Central del Límite.

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

$X \sim \mathcal{B}(n,p) \Rightarrow Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ tiene *aproximadamente* distribución normal estándar.

La aproximación es buena siempre que:

- $np > 5$.
- $n(1-p) > 5$.

Ejemplo 2:

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 $X =$ nro. 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 2:

- $np = 1000 \cdot 0{,}02 = 20 > 5$
- $n(1-p) = 1000 \cdot 0{,}98 = 980 > 5$
- 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:

$$P(X > 25) = P\left(\frac{X-20}{4{,}43} > \frac{25-20}{4{,}43}\right) \approx P(Z > 1{,}13) = 1 - \Phi(1{,}13) = 0{,}13$$

Correción de la aproximación por la Distribución Normal

Continuación Ejemplo 2 ¿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 $X \sim \mathcal{B}(n, p)$, $k \in [0, n]$

$$P(X = k) = P(k - 0,5 < X < k + 0,5) = P\left(\frac{k-0,5-np}{\sqrt{npq}} < \frac{X-np}{\sqrt{npq}} < \frac{k+0,5-np}{\sqrt{npq}}\right)$$
$$\approx \Phi(\frac{k+0,5-np}{\sqrt{npq}}) - \Phi(\frac{k-0,5-np}{\sqrt{npq}})$$

Luego, $P(X = 25) \approx \Phi(\frac{25+0,5-20}{4,43}) - \Phi(\frac{25-0,5-20}{4,43}) = 0,048$
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 2: 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,
   $P(X > 25) \approx P(X_N > 25 + 0{,}5)$
   $= P(Z > \frac{25{,}5-20}{4{,}43}) = 1 - \Phi(\frac{25{,}5-20}{4{,}43}) = 0{,}107$
3. Probabilidad Exacta $P(X > 25) = 0{,}109$

Ejemplo 3: Cierto equipo consta de 30 instrumentos electrónicos $C_1, C_2, \ldots, C_{30}$ 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 $X_1, X_2, \ldots$ variables aleatorias independientes, con $E(X_i) = \mu_i$, $\mathrm{Var}(X_i) = \sigma_i^2$ tales que $\sigma_i < \infty$ y al menos un $\sigma_i > 0$. Sea $S_n = X_1 + X_2 + \ldots + X_n$, $s_n = \sqrt{\mathrm{Var(S_n)}} = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$. Entonces, si se satisfacen las condiciones de Lindeberg,

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

Condiciones de Lindeberg:

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

Bibliografía

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

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