Ley de Los Grandes Números
y
Teorema Central del Límite

Desigualdad de Chebyshev

- Para calcular la probabilidad de un evento descripto en términos de una v.a. X es necesario conocer la distribución de la v.a. La desigualdad de Chebyshev provee una cota que no depende de la distribución sino sólo de la varianza de X.

- Desigualdad de Chebyshev:

Sea X una v.a. con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$, entonces

$$\forall\, \varepsilon > 0, \qquad P\left(\left|X - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

Otras formas de la desigualdad de Chebyshev

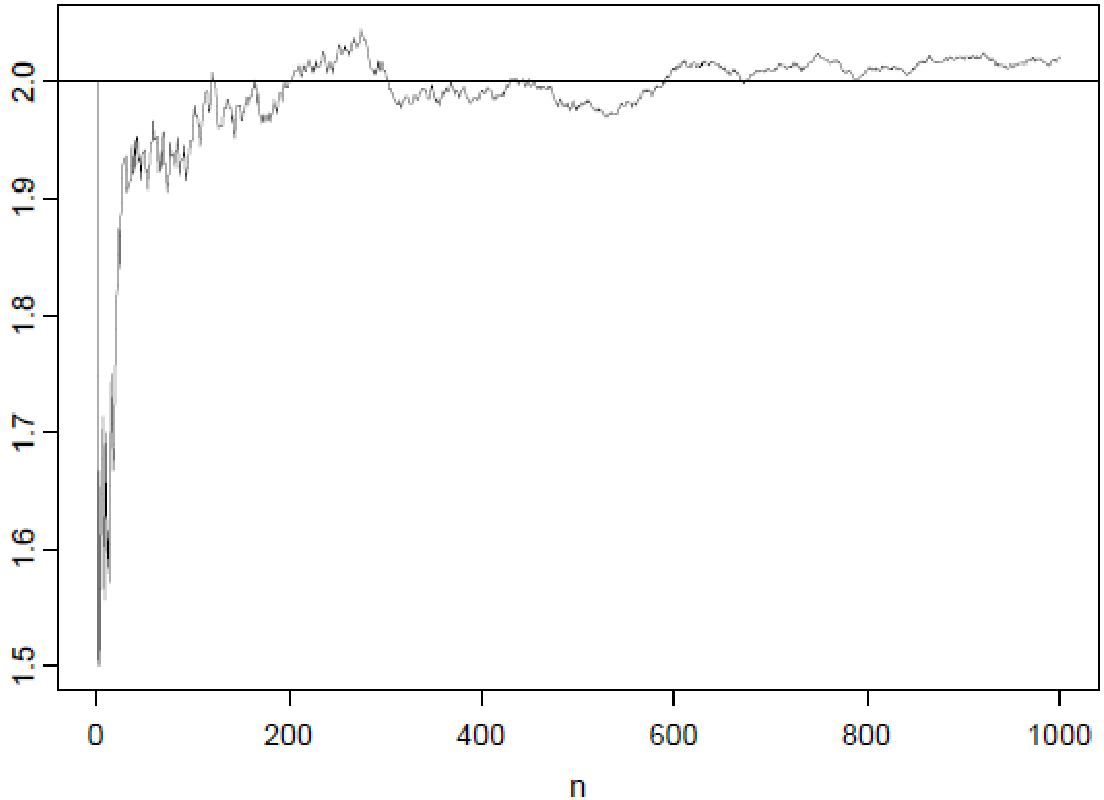
a) $\forall\, \varepsilon > 0, \quad P(|X - \mu| \leq \varepsilon) \geq 1 - \dfrac{\sigma^2}{\varepsilon^2}$

b) $\forall\, k \geq 1, \quad P(|X - \mu| > k\sigma) \leq \dfrac{1}{k^2}$

c) $\forall\, k \geq 1, \quad P(|X - \mu| \leq k\sigma) \geq 1 - \dfrac{1}{k^2}$

Esta cota es muy general y por lo tanto, puede ser poco precisa. Veamos un ejemplo.

Ley de los Grandes Números



Comportamiento asintótico del promedio muestral. El promedio del múmero observado de caras, $\overline{x}$, cuando 4 monedas equilibradas son arrojadas se aproxima al valor medio μ=2 de la distribución.

¿En qué sentido converge $\overline{X}$ a μ?

Sea (X_n) $(n \geq 1)$ una sucesión de variables aleatorias, diremos que X_n *converge en probabilidad* a la v.a. X y lo notaremos $X_n \xrightarrow{p} X$, si

$$\boxed{\lim_{n\to\infty} P\left(\left|X_n - X\right| > \varepsilon\right) = 0 \qquad \forall\, \varepsilon > 0}$$

Ley de los Grandes Números:

Sean X_1, X_2, ... v.a. independientes e idénticamente distribuidas (muestra aleatoria) con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$, entonces

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

siendo $\bar{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$ el denominado promedio muestral.

Teorema Central del Límite

- Veremos ahora que aún cuando las variables no son normales, la distribución normal resulta una buena aproximación para la distribución del promedio y la suma de variables aleatorias i.i.d, a condición de que estemos promediando o sumando una cantidad suficientemente grande de variables.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad T = \sum_{i=1}^{n} X_i$$

Recordemos que

$$\boxed{\begin{array}{ll} E(\overline{X}) = \mu & E(T) = n\mu \\ V(\overline{X}) = \sigma^2/n & V(T) = n\sigma^2 \end{array}}$$

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

Recordemos que

$$E(\bar{X}) = \mu \qquad E(T) = n\mu$$

$$V(\bar{X}) = \sigma^2/n \qquad V(T) = n\sigma^2$$

Estandarizados resultarían

$$\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \qquad \frac{T - n\mu}{\sqrt{n}\,\sigma}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad\qquad T = \sum_{i=1}^{n} X_i$$

Teorema Central del Límite: Sean $X_1, X_2, \ldots.$ v.a. i.i.d con $E(X_i) = \mu$ y $V(X_i) = \sigma^2 < \infty$, entonces si n es suficientemente grande,

$$\frac{T - n\mu}{\sqrt{n}\,\sigma} \overset{(a)}{\sim} N(0,1)$$

$$\frac{\sqrt{n}\left(\overline{X} - \mu\right)}{\sigma} \overset{(a)}{\sim} N(0,1)$$

- También podemos escribir

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

donde la convergencia en distribución ($\xrightarrow{d}$) se interpreta en el siguiente sentido:

$$P\left(\frac{T - n\mu}{\sqrt{n}\,\sigma} \leq a\right) \cong \Phi(a) \qquad P\left(\frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma} \leq a\right) \cong \Phi(a)$$

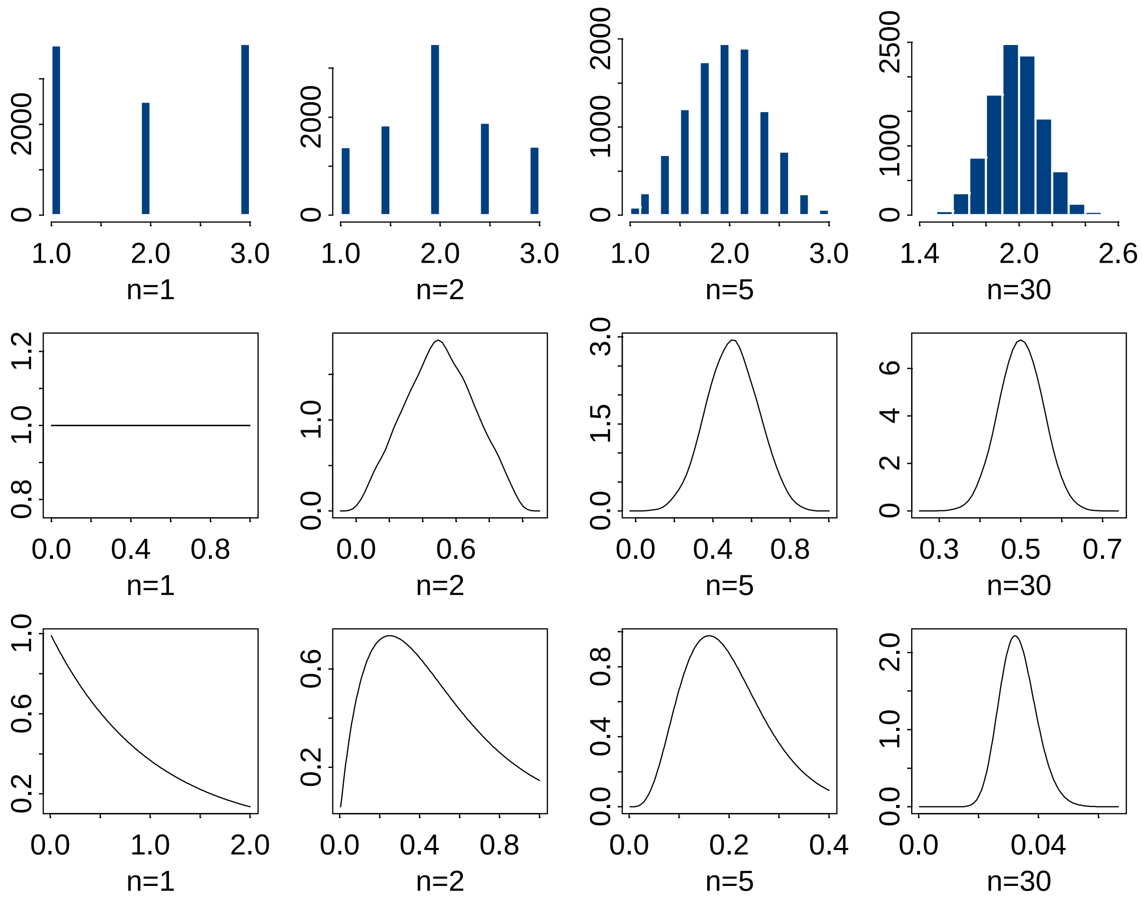


Figura 2: Distribución de $\bar{x}$ para distintas distribuciones cuando n=2, 5 y 30.

a) Distribución discreta, b) Distribución Uniforme, c) Distribución Exponencial

Ejemplo: Al sumar números, una calculadora aproxima cada número al entero más próximo. Los errores de aproximación se suponen independientes y con distribución $U(-0.5,0.5)$.

a) Si se suman 1500 números, ¿cuál es la probabilidad de que el valor absoluto del error total exceda 15?

b) ¿Cuántos números pueden sumarse a fin de que el valor absoluto del error total sea menor o igual que 10 con probabilidad mayor o igual que 0.90?

a) Si llamamos X_i al error correspondiente al i-ésimo sumando, el error total es

$$T_{1500} = \sum_{i=1}^{1500} X_i \ .$$

$E(X_i) = 0$ y $V(X_i) = \dfrac{1}{12}$ y por lo tanto $E(T_{1500}) = 0$ y $V(T_{1500}) = \dfrac{1500}{12}$.

Entonces,

$$P\left(\left|T_{1500}\right| > 15\right) = 1 - P\left(\left|T_{1500}\right| \le 15\right) = 1 - P(-15 \le T_{1500} \le 15) =$$

$$= 1 - P\left(\frac{-15}{\sqrt{1500/12}} \le \frac{T_{1500}}{\sqrt{1500/12}} \le \frac{15}{\sqrt{1500/12}}\right) \cong 1 - \Phi\left(\frac{15}{\sqrt{1500/12}}\right) + \Phi\left(\frac{-15}{\sqrt{1500/12}}\right) =$$

$$= 1 - \Phi(1.34) + \Phi(-1.34) = 0.18$$

b) Buscamos el valor de *n* tal que

$$P(|T_n| \leq 10) \geq 0.90$$

$$P(|T_n| \leq 10) \geq 0.90 \Leftrightarrow P(-10 \leq T_n \leq 10) \geq 0.90 \Leftrightarrow P\left(\frac{-10}{\sqrt{n/12}} \leq \frac{T_n}{\sqrt{n/12}} \leq \frac{10}{\sqrt{n/12}}\right) \geq 0.90$$

Aplicando la aproximación normal, debemos hallar *n* tal que

$$\Phi\left(\frac{10}{\sqrt{n/12}}\right) - \Phi\left(\frac{-10}{\sqrt{n/12}}\right) \geq 0.90 \Leftrightarrow 2\Phi\left(\frac{10}{\sqrt{n/12}}\right) - 1 \geq 0.90 \Leftrightarrow \Phi\left(\frac{10}{\sqrt{n/12}}\right) \geq 0.95$$

$$\Leftrightarrow \frac{10}{\sqrt{n/12}} \geq 1.64 \Leftrightarrow \sqrt{n} \leq 21.12 \Leftrightarrow \boxed{n \leq 446}$$

es decir, que se pueden sumar a lo sumo 446 números para que el valor absoluto del error total sea menor o igual que 10 con probabilidad mayor o igual que 0.90.

Aproximación de la distribución binomial por la normal

Sea $T \sim Bi\,(n,p)$ ⟹

T: el nro. de éxitos en *n* repeticiones de un experimento binomial con probabilidad de éxito igual a *p*

T / n: es la proporción muestral de éxitos.

Definamos las variables aleatorias para i = 1, ..., n

$$X_i = \begin{cases} 1 & \text{si se obtuvo Éxito en la repetición } i \\ 0 & \text{si se obtuvo Fracaso en la repetición } i \end{cases}$$

⟹ $X_i \sim Bi\,(1, p)$ $\forall$ i y

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

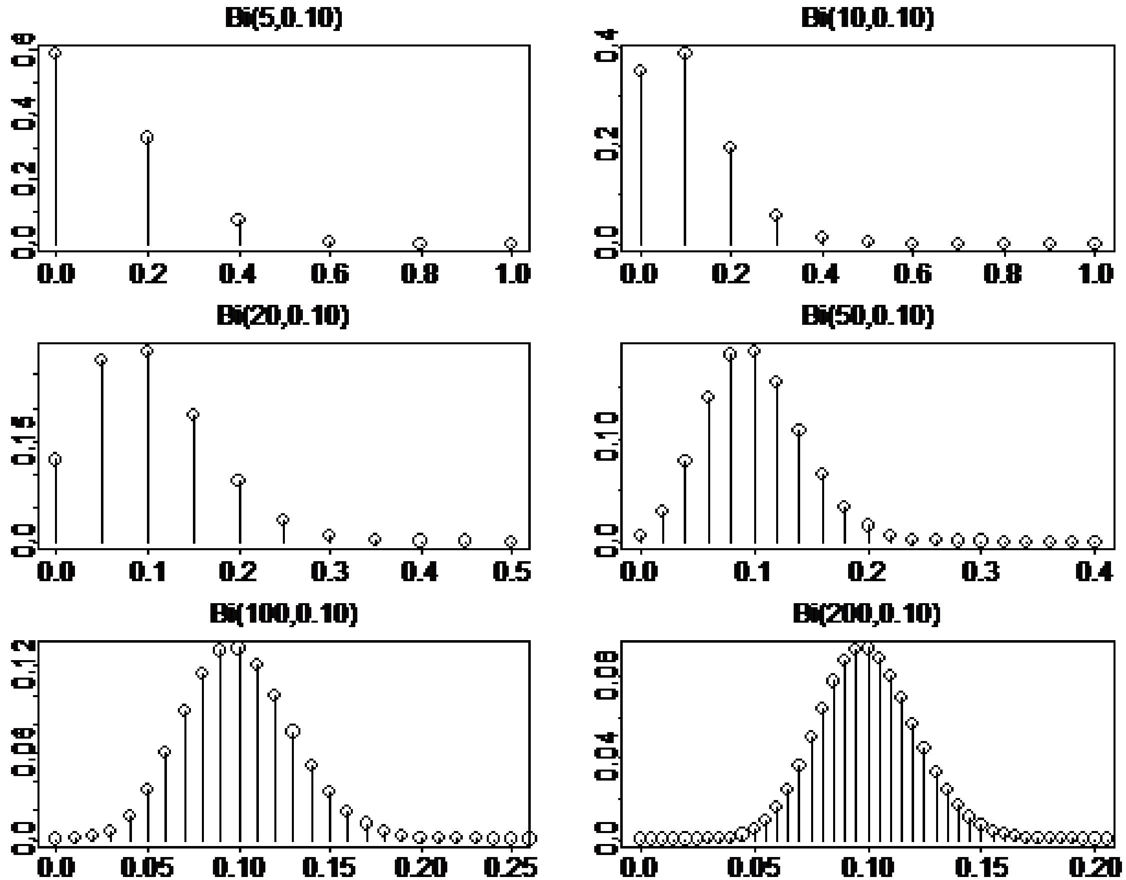
$$T = \sum_{i=1}^{n} X_i$$

Aplicando el TCL, si n es suficientemente grande,

$$\boxed{T \overset{(a)}{\sim} N\left(np, np(1-p)\right) \qquad \frac{T}{n} \overset{(a)}{\sim} N\left(p, \frac{p(1-p)}{n}\right)}$$

considera que la aproximación es buena si

n p $\geq$ 5 y n (1-p) $\geq$ 5.


Bi(5,0.10)
Bi(10,0.10)
Bi(20,0.10)
Bi(50,0.10)
Bi(100,0.10)
Bi(200,0.10)