

Teoremas límite

Consideremos el experimento aleatorio que consiste en arrojar una moneda equilibrada n veces. Supongamos que se registra la proporción de caras. Un resultado conocido es que esta proporción estará cerca de $1/2$.

Formalizando un poco tenemos:

ε : 'se arroja una moneda equilibrada n veces' A = salió cara

X : nº de caras que se observan, $X \sim \text{Bi}(n, 1/2)$

$$X_i = \begin{cases} 0 & \text{si sale ceca} \\ 1 & \text{si sale cara} \end{cases} \quad 1 \leq i \leq n, \text{ luego } X/n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

X/n : proporción de caras (= frecuencia relativa del suceso $A = f_A = n_A/n$)

Se espera que a medida que n crece la frecuencia relativa de cara $\frac{1}{n} \sum_{i=1}^n X_i = X/n$

-----se acerque a-----> $0.5 = E(X_i)$ (media poblacional)

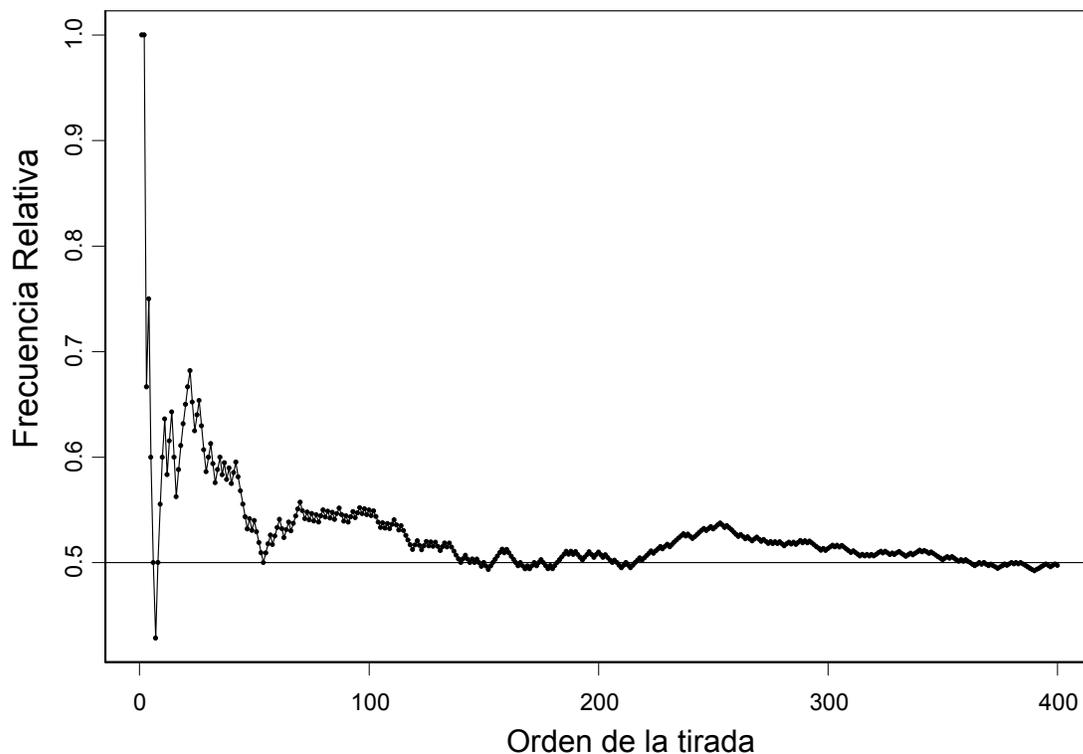


Figura 2: Frecuencia relativa de cara para una sucesión de 400 tiradas

La figura 2 muestra que las frecuencias relativas de cara obtenidas en 400 tiradas sucesivas de una moneda se acercan a 0.5 a medida que aumenta la cantidad de tiradas (los valores de X_i para las 20 primeras y 20 últimas tiradas fueron: 1 1 0 1 0 0 0 1 1 1 1 0 1 1 0 0 1 1 1 1 y 0 1 0 1 0 0 0 0 0 0 0 1 1 1 1 1 0 0 1 1 0 respectivamente)

Veremos a continuación la ley de los grandes números. Establece que la probabilidad de que la media muestral (calculada a partir de una muestra aleatoria de una distribución) difiera de la media poblacional en más de un valor fijo (que puede ser arbitrariamente pequeño) tiende a cero a medida que el tamaño de la muestra tiende a infinito. Por eso promediamos!!

Ley (débil) de los Grandes Números

Sea X_1, X_2, \dots, X_n , una sucesión de v.a. independientes tales que $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2 < \infty$, entonces $\forall \varepsilon > 0$

$$P(|\bar{X}_n - \mu| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \quad (1)$$

siendo $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.

Dem: Sabemos que $E(\bar{X}_n) = \mu$ y $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$, entonces aplicando la desigualdad de Chebishev a \bar{X}_n ,

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \quad \forall \varepsilon > 0$$

y, por lo tanto

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0 \quad \forall \varepsilon > 0 \quad \text{cqd.}$$

Cuando vale (1) decimos que la media muestral converge en probabilidad a μ , lo indicamos por

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

También decimos que \bar{X}_n es un estimador consistente de μ

Observación 1: La frecuencia relativa de un suceso A tiende a su probabilidad. Consideremos n repeticiones independientes de un experimento aleatorio y sea A un suceso con probabilidad $P(A) = p$, constante en las n repeticiones. Si llamamos n_A a la frecuencia absoluta de A (número de veces que ocurre A en las n repeticiones) y $f_A = n_A / n$ a la frecuencia relativa, entonces

$$f_A \xrightarrow{p} p$$

Esto resulta de la ley de los grandes números con $X_i \sim \text{Bi}(1,p)$, $p = P(A)$

$$f_A = n_A / n = \sum_{i=1}^n X_i / n \quad E(f_A) = p \quad \text{y} \quad \text{Var}(f_A) = p(1-p) / n.$$

Observación 2 La desigualdad de Chebishev también nos permite acotar la cantidad de repeticiones necesarias para que la frecuencia relativa del suceso A difiera de su probabilidad ($P(A) = p$) como máximo en 0.01 con probabilidad mayor o igual que 0.90.

En efecto, queremos encontrar n tal que

$$P(|f_A - p| \leq 0.01) \geq 0.90, \quad (1)$$

por Chebishev con $\varepsilon = 0.01$ tenemos que

$$P(|f_A - p| \leq 0.01) \geq 1 - \frac{p(1-p)}{n(0.01)^2}$$

Luego la desigualdad (1) se cumplirá siempre que

$$1 - \frac{p(1-p)}{n(0.01)^2} \geq 0.90 \Leftrightarrow \frac{p(1-p)}{n(0.01)^2} \leq 0.10 \Leftrightarrow n \geq \frac{p(1-p)}{(0.01)^2(0.10)}$$

El valor mínimo de n depende de p . Es máximo cuando $p = 0.50$. Por ejemplo,

$$p = 0.50 \Rightarrow n \geq 25000$$

$$p = 0.10 \Rightarrow n \geq 9250$$

$$p = 0.01 \Rightarrow n \geq 990$$

Observación 3: La existencia de la varianza no es necesaria para la validez de la ley de los grandes números, mientras que la de la esperanza es indispensable.

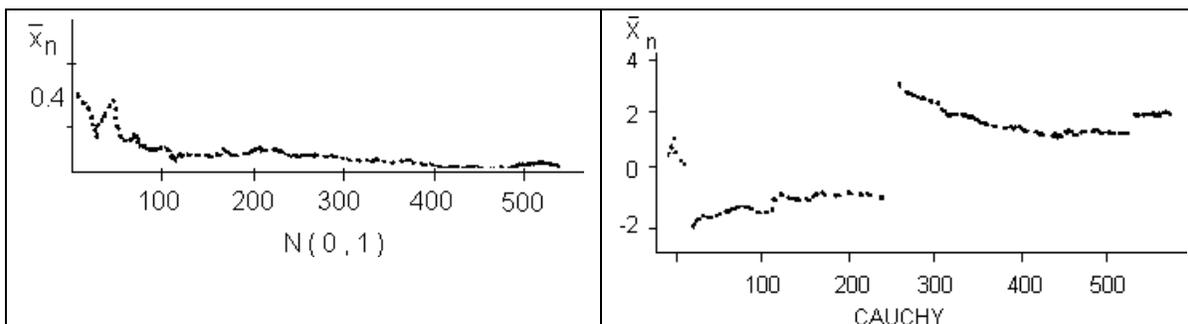


Figura 3: comportamiento de la media muestral para una sucesión de observaciones de una variable aleatoria $N(0,1)$ y una variable aleatoria Cauchy

La figura 3 muestra que la sucesión correspondiente a la $N(0,1)$ se acerca a cero, mientras que la sucesión correspondiente a la distribución Cauchy muestra un comportamiento errático.

¿Cómo es la distribución Cauchy?

La función de densidad de probabilidad de una variable aleatoria Z llamada Cauchy es

$$f_Z(z) = \frac{1}{\pi(1+z^2)} \quad -\infty < z < \infty$$

y su función de distribución acumulada

$$F_Z(z) = \frac{1}{2} + \frac{1}{\pi} \arctan(z) \quad -\infty < z < \infty$$

¿Qué forma tiene la función de densidad de la distribución Cauchy?

- Es simétrica con respecto al cero
- Tiene forma de campana
- Las colas tienden a cero más lentamente que la Normal

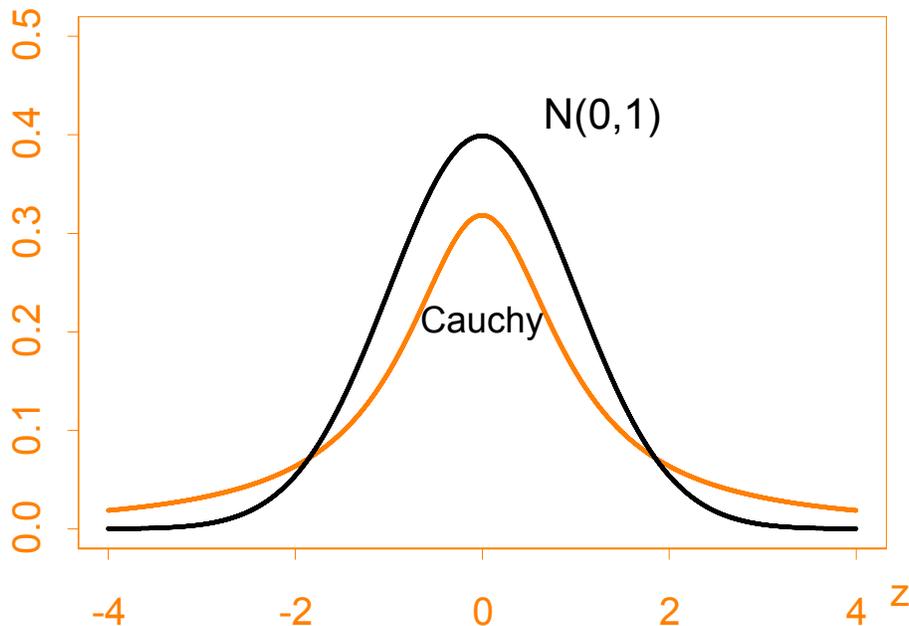


Figura 4: funciones de densidad de probabilidad $N(0,1)$ y Cauchy

Debido a la simetría alrededor de cero de la función de densidad la esperanza de una variable Cauchy debería ser cero. Sin embargo

$$E(|Z|) = \int_{-\infty}^{\infty} \frac{|z|}{\pi(1+z^2)} dz = \frac{2}{\pi} \int_0^{\infty} \frac{z}{(1+z^2)} dz = \frac{2}{\pi} \lim_{a \rightarrow \infty} \log(1+a^2) = \infty$$

La razón por la cual falla la existencia de la esperanza es debido a que la densidad decrece tan lentamente que valores grandes de la variable pueden ocurrir con una probabilidad sustancial

¿ Cómo surge la distribución Cauchy?

Se puede demostrar (vea por ejemplo John Rice “Mathematical Statistics and Data Analysis” 1988-1995) que el cociente de dos variables aleatorias $N(0,1)$ independientes tiene distribución Cauchy.

Corolario de la Ley de los Grandes Números

Sea X_1, \dots, X_n una muestra aleatoria de una distribución con $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2 < \infty$ entonces

- a) $\bar{X} \xrightarrow{p} \mu$ La media muestral converge en probabilidad a la media poblacional
- b) $S_X^2 \xrightarrow{p} \sigma^2$ La varianza muestral converge en probabilidad a la varianza poblacional

Dem.

a) Por la Ley de los Grandes Números.

b) Demostraremos que la varianza muestral S_X^2 es un estimador consistente de la varianza poblacional.

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = \frac{n}{n-1} \left(\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 \right)$$

Por la Ley de los Grandes Números $\bar{X} \xrightarrow{p} \mu$, entonces $\bar{X}^2 \xrightarrow{p} \mu^2$.

Por otra parte, aplicando nuevamente la Ley de los Grandes Números a X_i^2

$$\frac{\sum_{i=1}^n X_i^2}{n} \xrightarrow{p} E(X^2) = V(X) + [E(X)]^2 = \sigma^2 + \mu^2$$

Como además $\frac{n}{n-1} \rightarrow 1$, se obtiene

$$S_X^2 = \frac{n}{n-1} \left(\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 \right) \xrightarrow{p} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

y por lo tanto la varianza muestral es un estimador consistente de σ^2 .

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

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

con $T = \sum X_i$

Dicho de otro modo,

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

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

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

es decir, que las funciones de distribución convergen a la función de distribución normal standard.

Observación: ¿Qué significa n suficientemente grande? ¿Cómo sabemos si la aproximación es buena? El tamaño de muestra requerido para que la aproximación sea razonable depende de la forma de la distribución de las X_i . Mientras más simétrica y acampanada sea, más rápidamente se obtiene una buena aproximación.

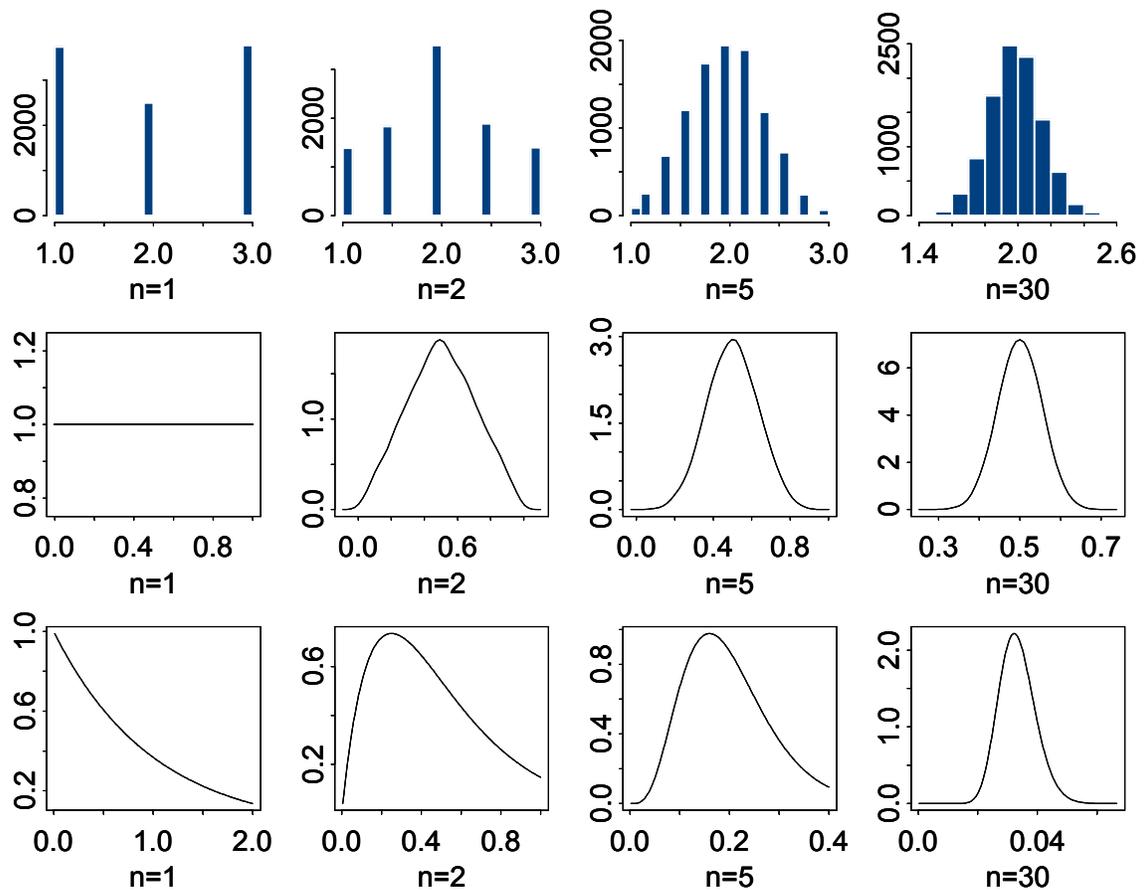


Figura 5: Distribución de \bar{X} para distintas distribuciones: a) discreta, b) Uniforme, c) Exponencial

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?

Sea X_i el error correspondiente al i -ésimo sumando, $E(X_i) = 0$ y $Var(X_i) = \frac{1}{12}$.

El error total es $T_{1500} = \sum_{i=1}^{1500} X_i$ y por lo tanto $E(T_{1500}) = 0$ y $Var(T_{1500}) = \frac{1500}{12}$.

$$P(|T_{1500}| > 15) = 1 - P(|T_{1500}| \leq 15) = 1 - P(-15 \leq T_{1500} \leq 15) =$$

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

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

b) ¿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?

Buscamos el valor de n tal que

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

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

Aplicando la aproximación normal, debemos hallar n tal que

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

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

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 $X \sim \text{Bi}(n, p)$, entonces X es el número de éxitos en n repeticiones de un experimento binomial con probabilidad de éxito igual a p , y X/n es la proporción muestral de éxitos.

Definamos las siguientes variables aleatorias

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

para $i = 1, \dots, n$. Estas v.a. son independientes, $X_i \sim \text{Bi}(1, p) \forall i$ y $X = \sum_{i=1}^n X_i$.

Aplicando el Teorema Central del Límite, si n es suficientemente grande,

$$X \sim N(np, np(1-p)) \quad \frac{X}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

La aproximación es buena si $np \geq 5$ y $n(1-p) \geq 5$. Si $p = 1/2$ la aproximación es buena aún para n pequeño pues en ese caso la distribución binomial es simétrica.

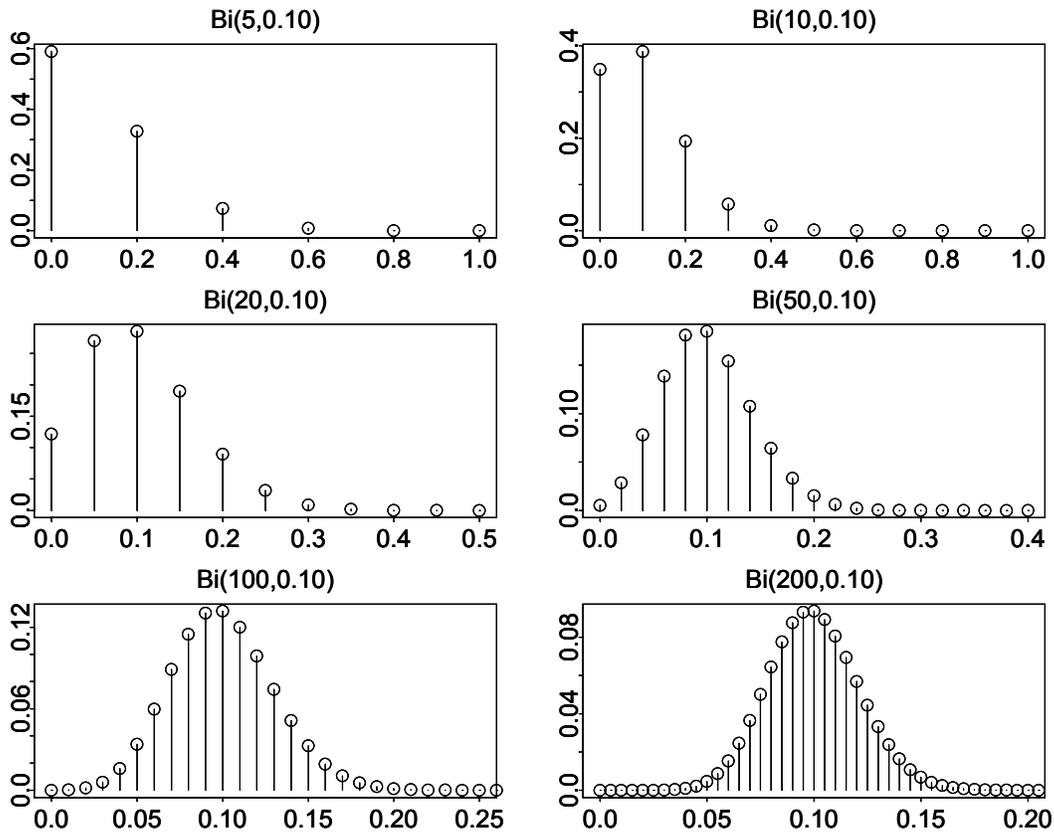


Figura 6: Distribución probabilidades de $\frac{X}{n}$

Corrección por continuidad

Cuando se aproxima una distribución discreta por una continua, como es el caso de la aproximación de la distribución binomial por la normal, es necesario efectuar una corrección. Consideremos el siguiente ejemplo:

Sea $X \sim Bi(100, 0.6)$ y calculemos en forma aproximada $P(X \leq 50)$ y $P(X \geq 51)$.

Si aplicamos directamente el TCL, obtenemos:

$$P(X \leq 50) = P\left(\frac{X - 60}{\sqrt{24}} \leq \frac{50 - 60}{\sqrt{24}}\right) \cong \Phi(-2.04) = 0.021$$

$$P(X \geq 51) = P\left(\frac{X - 60}{\sqrt{24}} \geq \frac{51 - 60}{\sqrt{24}}\right) \cong 1 - \Phi(-1.84) = 0.967$$

Si bien, $P(X \leq 50) + P(X \geq 51) = 1$, los valores aproximados no satisfacen esta restricción. Para evitar este problema, se efectúa la siguiente corrección, denominada *corrección por continuidad*,

$$P(X \leq 50) = P(X \leq 50.5) = P\left(\frac{X - 60}{\sqrt{24}} \leq \frac{50.5 - 60}{\sqrt{24}}\right) \cong \Phi(-1.94) = 0.026$$

$$P(X \geq 51) = P(X \geq 50.5) = P\left(\frac{X - 60}{\sqrt{24}} \geq \frac{50.5 - 60}{\sqrt{24}}\right) \cong 1 - \Phi(-1.94) = 0.974$$

En general, cuando la v.a. es discreta y $x_i - x_{i-1} = 1$, la corrección se realiza en la forma:

$$P(X \leq a) = P(X \leq a + 0.5)$$

$$P(X \geq a) = P(X \geq a - 0.5)$$

Si la distancia entre dos valores sucesivos de X es $k > 1$, ¿cómo aplicaría la corrección por continuidad?

Ejemplo: Sea $X \sim Bi(60, 1/3)$. Calcular en forma aproximada la probabilidad de que X sea mayor o igual que 25.

$$P(X \geq 25) = P(X \geq 24.5) = P\left(\frac{X - 60 \cdot \frac{1}{3}}{\sqrt{60 \cdot \frac{1}{3} \cdot \frac{2}{3}}} \geq \frac{24.5 - 60 \cdot \frac{1}{3}}{\sqrt{60 \cdot \frac{1}{3} \cdot \frac{2}{3}}}\right) \cong 1 - \Phi(1.23) = 0.11$$

Otras aplicaciones del Teorema Central del Límite:

a) Sean X_1, X_2, \dots, X_n v.a. i.i.d. con distribución Poisson de parámetro λ , entonces

$$\sum_{i=1}^n X_i \sim P(n\lambda)$$

Por lo tanto, cualquier v.a. con distribución de Poisson con parámetro suficientemente grande puede ser aproximada por la distribución normal.

b) Sean X_1, X_2, \dots, X_n v.a. independientes con distribución Gamma de parámetros n_i y λ , o sea $X_i \sim \Gamma(n_i, \lambda)$ entonces

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n n_i, \lambda\right)$$

Por lo tanto, cualquier v.a. con distribución $\Gamma(m, \lambda)$ con parámetro m suficientemente grande puede ser aproximada por la distribución normal.