

Probabilidad y Estadística (Borradores, Curso 23)
Normalidad y Teorema central del límite

Sebastian Grynberg

9 de mayo de 2011



*Quién pudiera agarrarte por la cola
magiafantasmanieblapoesía!
Acostarse contigo una vez sola
y después enterrar está manía!
Quién pudiera agarrarte por la cola!
(Juan Gelman)*

Índice

1. La distribución normal	2
1.1. Presentación	2
1.2. Cuentas con normales	5
1.3. Ejemplos	6
1.4. Suma de normales independientes	7
2. Génesis de la distribución normal	8
2.1. Teorema límite de De Moivre - Laplace	8
3. Teorema central del límite	15
3.1. Ejemplos	17
4. Distribuciones relacionadas con la Normal	21
4.1. χ^2 (chi-cuadrado)	21
4.2. t de Student	23
4.3. F de Fisher	24
5. Bibliografía consultada	26

1. La distribución normal

1.1. Presentación

Definición 1.1. *La función definida por*

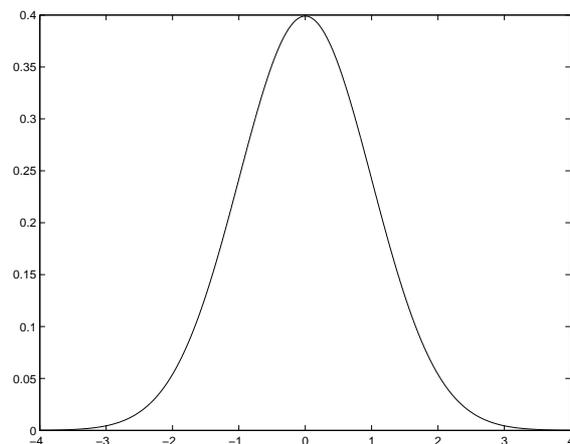
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (1)$$

se llama la función densidad normal; su integral

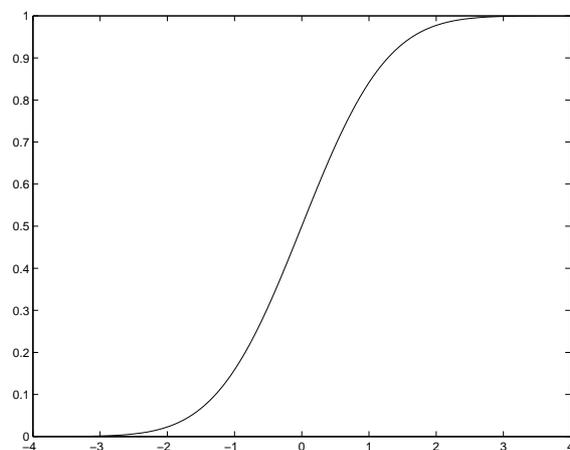
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (2)$$

es la función distribución normal.

Folclore. Se sabe que la función e^{-x^2} no admite una primitiva que pueda expresarse mediante un número finito de funciones elementales: x^ν , $\text{sen}(x)$, $\text{cos}(x)$, a^x , etc.... (Ver Piskunov, N., (1983). *cálculo diferencial e integral*, tomo I, Mir, Moscú). Sin embargo, usando técnicas de cambio de variables bidimensionales se puede demostrar que $\int_{-\infty}^{\infty} \varphi(x) dx = 1$.



(a)



(b)

Figura 1: (a) La función densidad normal $\varphi(x) := \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$; (b) La función distribución normal $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$

La función $\Phi(x)$ crece desde 0 hasta 1. Su gráfico es una curva con forma de S con

$$\Phi(-x) = 1 - \Phi(x). \quad (3)$$

Tablas. La tabla de valores de la función distribución normal se puede consultar en la mayoría de los libros sobre probabilidad y/o estadística. En general se tabulan los valores de $\Phi(x)$ para $x = d_0 + \frac{d_1}{10} + \frac{d_2}{100}$, donde $d_0 \in \{0, 1, 2, 3\}$ y $d_1, d_2 \in \{0, 1, 2, \dots, 9\}$. Las filas de la tabla están indexadas por los números d_0, d_1 y sus columnas por los números $0.0d_2$: en la posición $(d_0, d_1, 0.0d_2)$ de la tabla se encuentra el valor $\Phi(d_0, d_1, d_2)$. Por ejemplo, si se consulta la tabla del libro de Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, en fila 1.2 y columna de 0.08 puede leerse 0.8997, lo que significa que $\Phi(1.28) = 0.8997$.

En el Cuadro 1.1 reproducimos algunos de los valores de la tabla del Feller:

x	1.28	1.64	1.96	2.33	2.58	3.09	3.29
$\Phi(x)$	0.8997	0.9495	0.975	0.9901	0.9951	0.9990	0.9995

Cuadro 1: En la tabla se muestran algunos valores de $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$.

Lema 1.2. *Para cada $x > 0$ valen las siguientes desigualdades:*

$$\varphi(x) \left(\frac{1}{x} - \frac{1}{x^3} \right) < 1 - \Phi(x) < \varphi(x) \left(\frac{1}{x} \right). \quad (4)$$

Demostración. Usando que $\frac{d}{dx}\varphi(x) = -x\varphi(x)$ es fácil ver que las derivadas de los miembros de las desigualdades (4) satisfacen:

$$\begin{aligned} \frac{d}{dx} \left[\varphi(x) \left(\frac{1}{x} - \frac{1}{x^3} \right) \right] &= -\varphi(x) \left(1 - \frac{3}{x^4} \right). \\ \frac{d}{dx} [1 - \Phi(x)] &= -\varphi(x). \\ \frac{d}{dx} \left[\varphi(x) \left(\frac{1}{x} \right) \right] &= -\varphi(x) \left(1 + \frac{1}{x^2} \right). \end{aligned}$$

Por lo tanto,

$$\frac{d}{dx} \left[-\varphi(x) \left(\frac{1}{x} - \frac{1}{x^3} \right) \right] < \frac{d}{dx} [\Phi(x) - 1] < \frac{d}{dx} \left[-\varphi(x) \left(\frac{1}{x} \right) \right] \quad (5)$$

Las desigualdades (4) se obtienen integrando desde x hasta ∞ . \square

Nota Bene. De las desigualdades (4) se infiere un método de cálculo para aproximar los valores de $1 - \Phi(x)$: promediando los valores de los extremos de las desigualdades se obtiene una aproximación cuyo error absoluto es menor que la semi-diferencia entre ambos:

$$\left| 1 - \Phi(x) - \varphi(x) \left(\frac{1}{x} - \frac{1}{2x^3} \right) \right| \leq \frac{\varphi(x)}{2x^3}. \quad (6)$$

De la desigualdad (6) se puede ver que la aproximación

$$\Phi(x) \approx 1 - \varphi(x) \left(\frac{1}{x} - \frac{1}{2x^3} \right) \quad (7)$$

es prácticamente inútil para valores “pequeños” de x (i.e., $x \in (0, 1]$) pero va mejorando a medida que los valores de x “crecen”. Usando la aproximación dada en (7) se obtienen las siguientes aproximaciones

x	1.28	1.64	1.96	2.33	2.58	3.09	3.29
$\Phi(x)$	0.90454	0.94839	0.97406	0.98970	0.99487	0.99896	0.99948
$ \text{error} \leq$	0.04192	0.01178	0.00388	0.00104	0.00041	0.00005	0.00002

Cuadro 2: Algunos valores de $\Phi(x)$ obtenidos mediante la estimación (7).

Nota histórica La distribución normal fue descubierta por De Moivre en 1733 como resultado de analizar la forma límite de la distribución binomial simétrica y redescubierta nuevamente por Gauss (1809) y Laplace (1812) quienes la estudiaron en relación con sus trabajos sobre la teoría de los errores de observación. Laplace dio, además, el primer enunciado (incompleto) del teorema central del límite. (Ver Cramer, H., (1970). *Métodos matemáticos de estadística*, Aguilar, Madrid.)

1.2. Cuentas con normales

Sean $\mu \in \mathbb{R}$ y $\sigma > 0$ arbitrarios, pero fijos. Se dice que la variable aleatoria X tiene distribución normal de parámetros μ y σ^2 y se denota $X \sim \mathcal{N}(\mu, \sigma^2)$ si la función densidad de X es de la forma

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (8)$$

Nota Bene. Un hecho importante sobre las variables aleatorias normales es que si X tiene distribución normal $\mathcal{N}(\mu, \sigma^2)$, entonces

$$Z = \frac{X - \mu}{\sigma} \quad (9)$$

tiene distribución normal $\mathcal{N}(0, 1)$. En efecto,

$$\begin{aligned} \mathbb{P}(Z \leq z) &= \mathbb{P}((X - \mu)/\sigma \leq z) = \mathbb{P}(X \leq z\sigma + \mu) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z\sigma + \mu} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt \quad \text{por sustitución } x = t\sigma + \mu. \end{aligned}$$

Este hecho significa que si trasladamos el origen de las abscisas en μ y cambiamos la escala de manera tal que σ represente la unidad de medida, la distribución normal $\mathcal{N}(\mu, \sigma^2)$ se transforma en la distribución normal $\mathcal{N}(0, 1)$. Su importancia práctica radica en que permite reducir el cálculo de probabilidades de las distribuciones normales $\mathcal{N}(\mu, \sigma^2)$ al de la distribución normal $\mathcal{N}(0, 1)$. Motivo por el cual esta última recibe el nombre de *normal estándar* (o *típica*). Más precisamente, si X tiene distribución normal $\mathcal{N}(\mu, \sigma^2)$, su función de distribución podrá reducirse a la función de distribución normal $\Phi(\cdot)$ definida en (2) de

la siguiente manera:

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (10)$$

La identidad (10) resume toda la información probabilísticamente relevante sobre la variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$ y permite calcular (con ayuda de la tabla de la función de distribución normal $\Phi(\cdot)$) la probabilidad de que la variable X se encuentre en cualquier intervalo prefijado de antemano:

$$\mathbb{P}(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (11)$$

En particular, cuando el intervalo (a, b) es simétrico con respecto a μ , las cantidades a y b se pueden expresar en la forma $a = \mu - \epsilon$, $b = \mu + \epsilon$, donde $\epsilon > 0$, y la fórmula (11) adopta la forma

$$\mathbb{P}(|X - \mu| < \epsilon) = \Phi\left(\frac{\epsilon}{\sigma}\right) - \Phi\left(-\frac{\epsilon}{\sigma}\right) = 2\Phi\left(\frac{\epsilon}{\sigma}\right) - 1. \quad (12)$$

Significado de los parámetros μ y σ^2 . La relación (9) dice que si X es una variable aleatoria con distribución normal de parámetros μ y σ^2 , entonces $X = \sigma Z + \mu$ donde Z es una variable con distribución normal estándar. Cálculos de rutina muestran que $\mathbb{E}[Z] = 0$ y $\mathbb{V}(Z) = 1$, lo que permite deducir que *la media y la varianza de la $\mathcal{N}(\mu, \sigma^2)$ son μ y σ^2 , respectivamente.*

1.3. Ejemplos

Ejemplo 1.3. Una maquina produce ejes cuyos diámetros X tienen distribución normal de media $\mu = 10$ mm y varianza $\sigma^2 = 0.25$ mm. Un eje se considera defectuoso si $X < 9.5$ mm. Cuál es la probabilidad de que un eje elegido al azar resulte defectuoso?

Solución: El problema se resuelve calculando $\mathbb{P}(X < 9.5)$. Poniendo $\mu = 10$ y $\sigma = 0.5$ en la fórmula (10) obtenemos $\mathbb{P}(X < 9.5) = \Phi\left(\frac{9.5-10}{0.5}\right) = \Phi(-1) = 0.1587$. \square

Curva peligrosa. De inmediato podría surgir una objeción al uso de la distribución normal $\mathcal{N}(10, 0.25)$ para modelar el diámetro de los ejes. Al fin y al cabo, los diámetros deben ser positivos y la distribución normal adopta valores positivos y negativos. Sin embargo, el modelo anterior asigna una probabilidad despreciable al evento $X < 0$. En efecto, $\mathbb{P}(X < 0) = \mathbb{P}\left(\frac{X-10}{0.5} < \frac{0-10}{0.5}\right) = \mathbb{P}(Z < -20) = \Phi(-20) = 1 - \Phi(20)$. De acuerdo con la estimación (6) tenemos que $1 - \Phi(20) \approx \varphi(20) \left(\frac{1}{20} - \frac{1}{2 \cdot 20^3}\right) = O(10^{-89})$. Este tipo de situación es habitual en la práctica. Se tiene una variable aleatoria X de la que se sabe que no puede tomar valores negativos (p.ej. una distancia, una longitud, un área, un peso, una temperatura, un precio, etc.) y se la modela utilizando una distribución normal $\mathcal{N}(\mu, \sigma^2)$; motivados, por ejemplo, por cuestiones de simetría. En principio, el modelo podrá ser perfectamente válido siempre y cuando los valores de los parámetros μ y σ^2 sean tales que la probabilidad $\mathbb{P}(X < 0)$ sea prácticamente 0.

Nota Bene sobre grandes desvíos. Sea X una variable aleatoria con distribución normal de media μ y varianza σ^2 . Sea $t > 0$, utilizando la fórmula (12) podemos ver que

$$p_t := \mathbb{P}(|X - \mu| > t\sigma) = 1 - \mathbb{P}(|X - \mu| \leq t\sigma) = 1 - \left(2\Phi\left(\frac{t\sigma}{\sigma}\right) - 1\right) = 2(1 - \Phi(t)).$$

Usando la tabla de la distribución normal $\Phi(\cdot)$ se puede ver que $p_1 = 0.3174$, $p_2 = 0.0454$, $p_3 = 0.0028$. Estas probabilidades admiten la siguiente interpretación: cerca del 32% de los valores de una variable $X \sim \mathcal{N}(\mu, \sigma^2)$ se desvían de su media en más de σ ; solamente cerca de un 5% lo hacen en más de 2σ y solamente cerca de un 3% en más de 3σ . Esto da lugar a que en la mayor parte de los problemas de la práctica se consideren casi imposibles las desviaciones respecto de la media μ que superen 3σ y se consideren limitados por el intervalo $[\mu - 3\sigma, \mu + 3\sigma]$ todos los valores prácticamente posibles de la variable X .

Ejemplo 1.4. Sea X una variable aleatoria con distribución normal de media $\mu = 3$ y varianza $\sigma^2 = 4$. ¿Cuál es la probabilidad de que X sea no menor que 1 y no mayor que 7?

Solución: Poner $\mu = 3$ y $\sigma = 2$ en la fórmula (11) y usar la tabla de la distribución normal $\Phi(\cdot)$: $\mathbb{P}(1 \leq X \leq 7) = \Phi\left(\frac{7-3}{2}\right) - \Phi\left(\frac{1-3}{2}\right) = \Phi(2) - \Phi(-1) = 0.9773 - 0.1587 = 0.8186$. \square

1.4. Suma de normales independientes

Lema 1.5. Sean X_1 y X_2 dos variables aleatorias independientes con distribución normal $\mathcal{N}(\mu_1, \sigma_1^2)$ y $\mathcal{N}(\mu_2, \sigma_2^2)$, respectivamente. Entonces $X_1 + X_2$ tiene distribución normal $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Demostración. Observando que $X_1 + X_2 = (X_1 - \mu_1) + (X_2 - \mu_2) + \mu_1 + \mu_2$ el problema se reduce a considerar el caso $\mu_1 = \mu_2 = 0$. La prueba se obtiene mostrando que la convolución de las densidades $f_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-x_1^2/2\sigma_1^2)$ y $f_2(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-x_2^2/2\sigma_2^2)$ es la densidad normal de media $\mu_1 + \mu_2$ y varianza $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Por definición

$$(f_1 * f_2)(x) = \int_{-\infty}^{\infty} f_1(x-y)f_2(y) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-y)^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}\right) dy \quad (13)$$

El resultado se obtendrá mediante un poco de álgebra, bastante paciencia, y un cambio de variables en la integral del lado derecho de la identidad (13).

$$\begin{aligned} \exp\left(-\frac{(x-y)^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}\right) &= \exp\left(-\frac{1}{2}\left(\frac{\sigma}{\sigma_1\sigma_2}y - \frac{\sigma_2}{\sigma\sigma_1}x\right)^2 - \frac{x^2}{2\sigma^2}\right) \\ &= \exp\left(-\frac{1}{2}\left(\frac{\sigma}{\sigma_1\sigma_2}y - \frac{\sigma_2}{\sigma\sigma_1}x\right)^2\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) \end{aligned}$$

La primera igualdad se obtuvo completando cuadrados respecto de y en la expresión $-\frac{(x-y)^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}$ y reagrupando algunos términos. Mediante el cambio de variables $z = \frac{\sigma_1}{\sigma_1\sigma_2}y - \frac{\sigma_2}{\sigma_1}x$, cuya diferencial es de la forma $dz = \frac{\sigma_1}{\sigma_1\sigma_2}dy$, se puede ver que

$$(f_1 * f_2)(x) = \frac{1}{2\pi\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

□

Este resultado se puede generalizar para una suma de n variables aleatorias independientes: Sean X_1, X_2, \dots, X_n variables aleatorias independientes con distribuciones normales: $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $1 \leq i \leq n$. Entonces,

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

La prueba se obtiene por inducción y utilizando la siguiente propiedad “hereditaria” de familias de variables aleatorias independientes (cuya prueba puede verse en el Capítulo 1 del libro de Durrett, R.(1996): *Probability Theory and Examples*): Si X_1, X_2, \dots, X_n son variables aleatorias independientes, entonces funciones (medibles) de familias disjuntas de las X_i también son independientes.

Nota Bene. Observando que para cada $a \in \mathbb{R}$ y $X \sim \mathcal{N}(\mu, \sigma^2)$ resulta que $aX \sim \mathcal{N}(a\mu, a^2\sigma^2)$ se obtiene el siguiente resultado:

Teorema 1.6. Sean X_1, X_2, \dots, X_n variables aleatorias independientes con distribuciones normales: $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $1 \leq i \leq n$ y sean a_1, a_2, \dots, a_n números reales cualesquiera. Entonces,

$$\sum_{i=1}^n a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

2. Génesis de la distribución normal

2.1. Teorema límite de De Moivre - Laplace

En 1733, De Moivre observó que la distribución binomial correspondiente a la cantidad de éxitos, S_n , en n ensayos de Bernoulli simétricos tiene la forma límite de una campana. Esta observación fue la clave que le permitió descubrir la famosa *campana de Gauss* y allanar el camino que lo condujo a establecer la primera versión del *Teorema Central del Límite*: la convergencia de la distribución Binomial($n, 1/2$) a la distribución normal estándar. En 1801, Laplace refinó y generalizó este resultado al caso de la distribución Binomial(n, p). El Teorema de De Moivre-Laplace, que enunciamos más abajo, mejora sustancialmente la Ley débil de los grandes números porque proporciona una estimación mucho más precisa de las probabilidades $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \epsilon\right)$.

Teorema 2.1 (Teorema límite de De Moivre-Laplace). *Consideramos una sucesión de ensayos de Bernoulli independientes. Sean p la probabilidad de éxito en cada ensayo y S_n la cantidad de éxitos observados en los primeros n ensayos. Para cualquier $x \in \mathbb{R}$ vale que*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x), \quad (14)$$

donde $\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ es la función distribución normal estándar.

Demostración. Ver Capítulo VII de Feller, W., (1971). *An Introduction to Probability Theory and Its Applications*, Vol. I, John Wiley & Sons, New York. \square

¿Qué significa el Teorema Límite de De Moivre-Laplace? Para contestar esta pregunta vamos a reconstruir las ideas principales de su génesis. En otras palabras, vamos a (re)construir el Teorema. La clave de la construcción está “embutida” en la Figura 2. La imagen permite “capturar de inmediato” la existencia de una forma límite para la distribución Binomial en el caso simétrico $p = 1/2$.

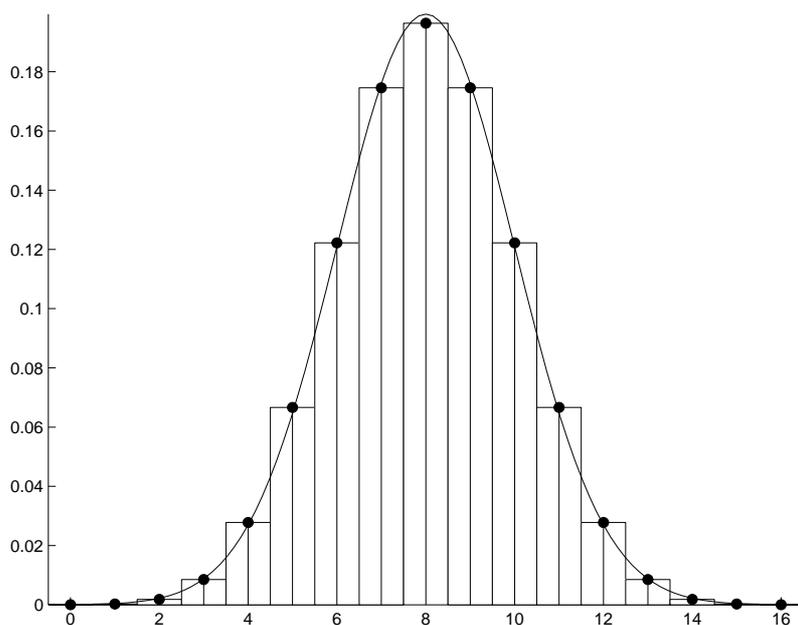


Figura 2: Relación entre la distribución Binomial simétrica y la distribución normal. La probabilidad de que ocurran k éxitos en n ensayos de Bernoulli está representada por un segmento paralelo al eje de las abscisas localizado en la ordenada k de altura igual a $\mathbb{P}(S_n = k)$. La curva continua “aproxima” los valores de $\mathbb{P}(S_n = k)$. Observar que dichas probabilidades también se pueden representar como áreas de rectángulos de altura $\mathbb{P}(S_n = k)$ y de base unitaria centrada en k .

Paso 1. El primer paso en la dirección del Teorema de De Moivre consiste en darse cuenta que la Figura 2 señala la existencia de una forma límite. En una primera fase (completamente abstracta) podemos conjeturar que “*la distribución binomial simétrica tiene una forma asintótica. En otras palabras, cuando la cantidad de ensayos de Bernoulli es suficientemente grande, salvo traslaciones y cambios de escala apropiados, la distribución Binomial se parece a una función continua par, $\varphi(x)$, cuyo gráfico tiene la forma de una campana.*”

Paso 2. El segundo paso consiste en precisar la naturaleza de la traslación y los cambios de escala que permiten “capturar” esa forma límite. Si se reflexiona sobre el significado de la media y la varianza de una variable aleatoria, parece claro que la forma límite se obtendrá centrando la variable S_n en su valor medio, $\mathbb{E}[S_n] = \frac{1}{2}n$, y adoptando como unidad de medida la desviación típica de los valores observados respecto de dicho valor, $\sigma(S_n) = \frac{1}{2}\sqrt{n}$. El significado geométrico de esta transformación consiste en (1) trasladar el origen de las abscisas en $\frac{1}{2}n$ y (2) dividir las por $\frac{1}{2}\sqrt{n}$. Para que las áreas de los rectángulos sigan representando probabilidades, las ordenadas deben multiplicarse por el mismo número. Este paso permite enunciar la siguiente versión mejorada de la conjetura inicial: “*existe una función continua $\varphi(x)$ tal que*

$$\mathbb{P}(S_n = k) = \binom{n}{k} \left(\frac{1}{2}\right)^n \sim \frac{1}{\frac{1}{2}\sqrt{n}} \varphi\left(\frac{k - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right), \quad (15)$$

siempre y cuando n sea suficientemente grande.”

Paso 3. Establecida la conjetura el problema consiste en “descubrir” la expresión de la función $\varphi(x)$ y en precisar cuál es el sentido de la relación aproximada que aparece en (15). En este punto no queda otra que “arremangarse y meter la mano en el barro”. Como resultado se obtiene que la expresión de la función $\varphi(x)$ es

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

y que la relación \sim vale para valores de k del orden de \sqrt{n} y significa que el cociente de los dos lados tiende a 1 cuando $n \rightarrow \infty$.

Nota Bene. La relación (15) expresa matemáticamente un hecho que se observa claramente en la Figura 2: la campana “pasa” por los puntos de base k y altura $\mathbb{P}(S_n = k)$. Conviene observar que la expresión que aparece en el lado derecho de la relación (15) es la función de densidad de la normal $\mathcal{N}\left(\frac{1}{2}n, \frac{1}{4}n\right)$ evaluada en $x = k$. En la práctica, esto significa que para obtener una buena aproximación de la probabilidad de observar k éxitos en n ensayos de Bernoulli independientes, basta con evaluar la densidad de la normal $\mathcal{N}\left(\frac{1}{2}n, \frac{1}{4}n\right)$ en $x = k$. Sin temor a equivocarnos, podemos resumir estas observaciones mediante una expresión de la forma $S_n \sim \mathcal{N}(\mathbb{E}[S_n], \mathbb{V}(S_n))$.

Paso 4. Observar que para cada $x_1 < x_2$ vale que

$$\begin{aligned} \mathbb{P}\left(x_1 \leq \frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq x_2\right) &= \mathbb{P}\left(\frac{1}{2}n + x_1\frac{1}{2}\sqrt{n} \leq S_n \leq \frac{1}{2}n + x_2\frac{1}{2}\sqrt{n}\right) \\ &= \sum_{x_1\frac{1}{2}\sqrt{n} \leq j \leq x_2\frac{1}{2}\sqrt{n}} \mathbb{P}\left(S_n = \frac{1}{2}n + j\right) \\ &\approx \sum_{x_1 \leq jh \leq x_2} h\varphi(jh), \end{aligned} \quad (16)$$

donde $h = \frac{2}{\sqrt{n}}$ y la suma se realiza sobre todos los enteros j tales que $x_1 \leq jh \leq x_2$. Cada uno de los sumandos que aparecen en el lado derecho de la aproximación (16) es el área de un rectángulo de base $[kh, (k+1)h]$ y altura $\varphi(kh)$. Como la función $\varphi(\cdot)$ es continua, para valores pequeños de h la suma total de las áreas de los rectángulo debe estar próxima del área bajo la curva de la densidad normal entre x_1 y x_2 . Por lo tanto, debe valer lo siguiente

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(x_1 \leq \frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq x_2\right) = \int_{x_1}^{x_2} \varphi(t)dt = \Phi(x_2) - \Phi(x_1). \quad (17)$$

Este paso puede hacerse formalmente preciso “arremangandose y metiendo la mano en ...”

Nota Bene. La variable aleatoria que aparece dentro de la probabilidad del lado izquierdo de (17)

$$S_n^* = \frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{S_n - \mathbb{E}[S_n]}{\sigma(S_n)} \quad (18)$$

es una medida de la desviación de S_n respecto de la media $\mathbb{E}[S_n]$ en unidades de la desviación típica $\sigma(S_n)$. El teorema límite de De Moivre-Laplace significa que cuando se considera una cantidad n (suficientemente grande) de ensayos de Bernoulli independientes, la distribución de la variable aleatoria S_n^* es “prácticamente indistinguible” de la distribución normal estándar $\mathcal{N}(0, 1)$.

Comentario sobre prueba del Teorema 2.1. Si se sigue con cuidado la demostración presentada por Feller se puede ver que las herramientas principales de la prueba son el desarrollo de Taylor (1712) de la función $\log(1+t) = t + O(t^2)$ y la fórmula asintótica de Stirling (1730) para los números factoriales $n! \sim \sqrt{2\pi n} n^n e^{-n}$. Partiendo de la función de probabilidad de la Binomial($n, 1/2$) se “deduce” la expresión de la función densidad normal $(\sqrt{2\pi})^{-1}e^{-x^2/2}$: el factor $(\sqrt{2\pi})^{-1}$ proviene de la fórmula de Stirling y el factor $e^{-x^2/2}$ del desarrollo de Taylor. Dejando de lado los recursos técnicos utilizados en la prueba, se observa que las ideas involucradas son simples y “recorren el camino del descubrimiento” de De Moivre (1733).

Ejemplo 2.2. Se lanza 40 veces una moneda honesta. Hallar la probabilidad de que se obtengan exactamente 20 caras. Usar la aproximación normal y compararla con la solución exacta.

Solución: La cantidad de caras en 40 lanzamientos de una moneda honesta, S_{40} , es una variable Binomial de parámetros $n = 40$ y $p = 1/2$. La aproximación normal (15) establece que

$$\mathbb{P}(S_{40} = 20) \sim \frac{1}{\frac{1}{2}\sqrt{40}} \varphi(0) = \frac{1}{\sqrt{20\pi}} = 0.12615\dots$$

El resultado exacto es

$$\mathbb{P}(X = 20) = \binom{40}{20} \left(\frac{1}{2}\right)^{40} = 0.12537\dots$$

□

Ejemplo 2.3. Se dice que los recién nacidos de madres fumadoras tienden a ser más pequeños y propensos a una variedad de dolencias. Se conjetura que además parecen deformes. A un grupo de enfermeras se les mostró una selección de fotografías de bebés, la mitad de los cuales nacieron de madres fumadoras; las enfermeras fueron invitadas a juzgar a partir de la apariencia de cada uno si la madre era fumadora o no. En 1500 ensayos se obtuvieron 910 respuestas correctas. La conjetura es plausible?

Solución: Aunque superficial, un argumento atendible consiste en afirmar que, si todos los bebés parecen iguales, la cantidad de repuestas correctas S_n en n ensayos es una variable aleatoria con distribución Binomial $(n, 1/2)$. Entonces, para n grande

$$\mathbb{P}\left(\frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} > 3\right) = 1 - \mathbb{P}\left(\frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \leq 3\right) \approx 1 - \Phi(3) \approx \frac{1}{1000}$$

por el Teorema límite de De Moivre-Laplace. Para los valores dados de S_n ,

$$\frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{910 - 750}{5\sqrt{15}} \approx 8.$$

Se podría decir que el evento $\{X - \frac{1}{2}n > \frac{3}{2}\sqrt{n}\}$ es tan improbable que su ocurrencia arroja dudas sobre la suposición original de que los bebés parecen iguales. Este argumento otorgaría cierto grado de credibilidad a la conjetura enunciada. □

Comentarios sobre el caso general

1. En el caso general, la probabilidad de éxito en cada ensayo de Bernoulli individual es $p \in (0, 1)$. Si S_n es la cantidad de éxitos observados en los primeros n ensayos, entonces $\mathbb{E}[S_n] = np$ y $\mathbb{V}(S_n) = np(1 - p)$. Por lo tanto, la variable aleatoria

$$S_n^* := \frac{S_n - np}{\sqrt{np(1 - p)}} \tag{19}$$

es una medida de la desviación de S_n respecto de la media $\mathbb{E}[S_n] = np$ en unidades de la desviación típica $\sigma(S_n) = \sqrt{np(1 - p)}$. El teorema límite de De Moivre-Laplace significa

que cuando se considera una cantidad n (suficientemente grande) de ensayos de Bernoulli independientes, la distribución de la variable aleatoria S_n^* es “prácticamente indistinguible” de la distribución normal estándar $\mathcal{N}(0, 1)$.

2. Técnicamente la prueba del teorema se puede hacer recurriendo a las mismas herramientas utilizadas en la prueba del caso simétrico, pero los cálculos involucrados son más complicados. Sin embargo, el resultado también es claro si se observan las gráficas de la distribución Binomial(n, p). En la Figura 3 se ilustra el caso $n = 16$ y $p = 1/4$. Nuevamente es “evidente” que la forma límite de distribución Binomial debe ser la distribución normal.

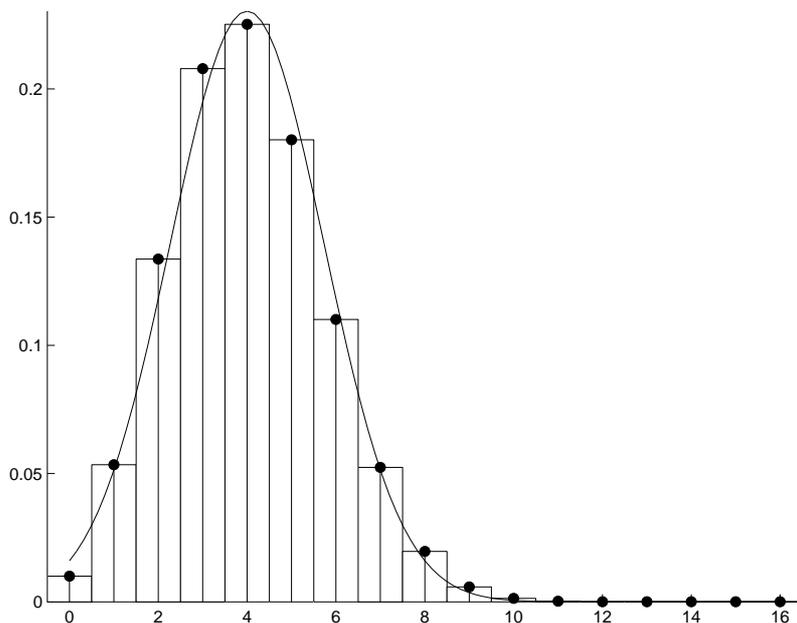


Figura 3: Gráfica de la función de probabilidad binomial con $n = 16$ y $p = 1/4$. Cerca del término central $m = np = 4$, salvo un cambio de escala (cuya unidad de medida es $\sqrt{np(1-p)} = \sqrt{3}$) la gráfica es “indistinguible” de la gráfica de la densidad normal.

3. De la Figura 3 debería estar claro que, para n suficientemente grande, debe valer lo siguiente

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{\sqrt{np(1-p)}} \varphi\left(\frac{k - np}{\sqrt{np(1-p)}}\right). \quad (20)$$

Ejemplo 2.4. Para el caso ilustrado en la Figura 3: $n = 16$ y $p = 1/4$, la aproximación (20) es bastante buena, incluso con un valor de n pequeño. Para $k = 0, \dots, 4$ las probabilidades $\mathbb{P}(S_n = 4 + k)$ son 0.2252, 0.1802, 0.1101, 0.0524, 0.0197. Las aproximaciones correspondientes son 0.2303, 0.1950, 0.1183, 0.0514, 0.0160. \square

Nota Bene. El Teorema límite de De Moivre-Laplace justifica el uso de los métodos de la curva normal para aproximar probabilidades relacionadas con ensayos de Bernoulli con probabilidad de éxito p . La experiencia “indica” que la aproximación es bastante buena siempre que $np > 5$ cuando $p \leq 1/2$, y $n(1 - p) > 5$ cuando $p > 1/2$. Un valor muy pequeño de p junto con un valor de n moderado darán lugar a una media pequeña y con ello se obtendrá una distribución asimétrica. La mayor parte de la distribución se acumulará alrededor de 0, impidiendo con ello que una curva normal se le ajuste bien. Si la media se aparta por lo menos 5 unidades de una y otra extremidad, la distribución tiene suficiente espacio para que resulte bastante simétrica. (Ver la Figura 4).

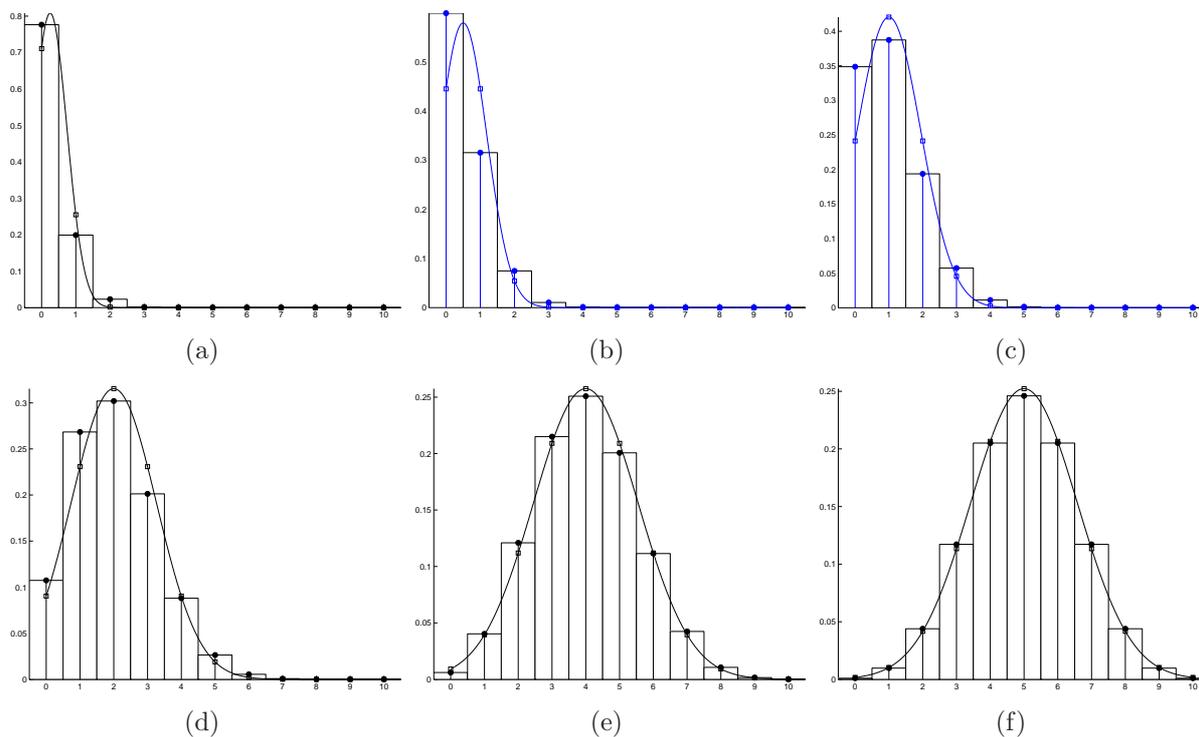


Figura 4: Comparación entre la distribución Binomial($10, p$) y su aproximación por la normal para distintos valores de p (a) $p = 0.025$; (b) $p = 0.05$; (c) $p = 0.1$; (d) $p = 0.2$; (e) $p = 0.4$; (f) $p = 0.5$.

Ejemplo 2.5 (Encuesta electoral). Queremos estimar la proporción del electorado que pretende votar a un cierto candidato. Para ello consideramos que el voto de cada elector tiene una distribución Bernoulli de parámetro p . Concretamente, queremos encontrar un tamaño muestral n suficiente para que con una certeza del 99.99% podamos garantizar un error máximo de 0.02 entre el verdadero valor de p y la proporción muestral S_n/n . En otras palabras, queremos encontrar n tal que

$$\mathbb{P} \left(\left| \frac{S_n}{n} - p \right| \leq 0.02 \right) \geq 0.9999. \quad (21)$$

Para acotar la incerteza usaremos la aproximación por la normal provista por el teorema límite de De Moivre - Laplace. Para ello, en lugar de observar la variable S_n , debemos observar la variable normalizada $S_n^* := (S_n - np)/\sqrt{np(1-p)}$. En primer lugar observamos que, como consecuencia del teorema límite, tenemos la siguiente aproximación

$$\mathbb{P}\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq a\right) \approx \Phi(-a) - \Phi(a) = 2\Phi(a) - 1 \quad (22)$$

o lo que es equivalente

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \frac{a\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 2\Phi(a) - 1. \quad (23)$$

Como el verdadero valor de p es desconocido, la fórmula (23) no puede aplicarse directamente ya que no se conoce el valor de $\sqrt{p(1-p)}$. Sin embargo, es fácil ver que $\sqrt{p(1-p)} \leq 1/2$ y por lo tanto

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \frac{a}{2\sqrt{n}}\right) \geq \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \frac{a\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 2\Phi(a) - 1. \quad (24)$$

Esta última relación es la herramienta con la que podemos resolver nuestro problema.

En primer lugar tenemos que resolver la ecuación $2\Phi(a) - 1 = 0.9999$ o la ecuación equivalente $\Phi(a) = \frac{1.9999}{2} = 0.99995$. La solución de esta ecuación se obtiene consultando una tabla de la distribución normal: $a = 3.9$. Reemplazando este valor de a en (24) obtenemos

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \frac{3.9}{2\sqrt{n}}\right) \geq 0.9999.$$

En segundo lugar tenemos que encontrar los valores de n que satisfacen la desigualdad

$$\frac{3.9}{2\sqrt{n}} \leq 0.02. \quad (25)$$

Es fácil ver que n satisface la desigualdad (25) si y solo si

$$n \geq \left(\frac{3.9}{0.04}\right)^2 = (97.5)^2 = 9506.2$$

El problema está resuelto. □

3. Teorema central del límite

Los teoremas sobre normalidad asintótica de sumas de variables aleatorias se llaman Teoremas Centrales del Límite. El Teorema límite de De Moivre - Laplace es un Teorema Central del Límite para variables aleatorias independientes con distribución Bernoulli(p). Una versión más general es la siguiente:

Teorema 3.1 (Teorema Central del Límite). Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes idénticamente distribuidas, cada una con media μ y varianza σ^2 . Entonces la distribución de

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

tiende a la normal estándar cuando $n \rightarrow \infty$. Esto es,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x \right) = \Phi(x),$$

donde $\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ es la función de distribución de una normal de media 0 y varianza 1.

Demostración. Ver Capítulo XV de Feller, W., (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley & Sons, New York. \square

Corolario 3.2. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes idénticamente distribuidas, cada una con media μ y varianza σ^2 . Si n es suficientemente grande, para cada valor $a > 0$ vale la siguiente aproximación

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \leq a \frac{\sigma}{\sqrt{n}} \right) \approx 2\Phi(a) - 1 \quad (26)$$

Demostración. El teorema central del límite establece que si n es suficientemente grande, entonces para cada $x \in \mathbb{R}$ vale que

$$\mathbb{P} \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x \right) \approx \Phi(x) \quad (27)$$

De la aproximación (27) se deduce que para cada valor $a > 0$

$$\mathbb{P} \left(\left| \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \right| \leq a \right) \approx \Phi(a) - \Phi(-a) = 2\Phi(a) - 1. \quad (28)$$

El resultado se obtiene de (28) observando que

$$\left| \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \right| = \frac{n}{\sigma\sqrt{n}} \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| = \frac{\sqrt{n}}{\sigma} \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right|. \quad (29)$$

\square

Nota Bene. Para los usos prácticos, especialmente en estadística, el resultado límite en sí mismo no es de interés primordial. Lo que interesa es usarlo como una aproximación con valores finitos de n . Aunque no es posible dar un enunciado consiso sobre cuan buena es la aproximación, se pueden dar algunas pautas generales y examinando algunos casos especiales se puede tener alguna idea más precisa del comportamiento de cuan buena es la aproximación. Qué tan rápido la aproximación es buena depende de la distribución de los sumandos. Si la distribución es bastante simétrica y sus colas decaen rápidamente, la aproximación es buena para valores relativamente pequeños de n . Si la distribución es muy asimétrica o si sus colas decaen muy lentamente, se necesitan valores grandes de n para obtener una buena aproximación.

3.1. Ejemplos

Ejemplo 3.3 (Suma de uniformes). Puesto que la distribución uniforme sobre $[-\frac{1}{2}, \frac{1}{2}]$ tiene media 0 y varianza $\frac{1}{12}$, la suma de 12 variables independientes $\mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$ tiene media 0 y varianza 1. La distribución de esa suma está muy cerca de la normal.

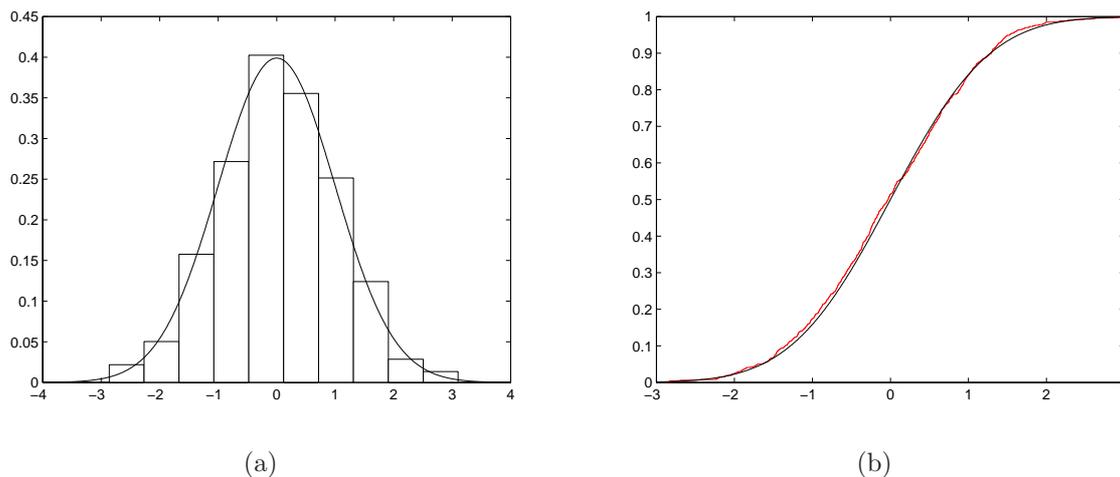


Figura 5: (a) Comparación entre un histograma de 1000 valores, cada uno de los cuales es la suma de 12 variables uniformes $\mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$, y la función densidad normal; (b) Comparación entre la función de distribución empírica correspondiente a 1000 valores de la suma de 12 uniformes $\mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$ y la función de distribución normal. El ajuste es sorprendentemente bueno, especialmente si se tiene en cuenta que 12 no se considera un número muy grande.

Ejemplo 3.4. Para simplificar el cálculo de una suma se redondean todos los números al entero más cercano. Si el error de redondeo se puede representar como una variable aleatoria $\mathcal{U}(-\frac{1}{2}, \frac{1}{2})$ y se suman 12 números, ¿cuál es la probabilidad de que el error de redondeo exceda 1?

Solución: El error de redondeo cometido al sumar 12 números se representa por la suma $\sum_{i=1}^{12} X_i$ de 12 variables aleatorias independientes X_1, \dots, X_{12} cada una con distribución uniforme sobre el intervalo $(-\frac{1}{2}, \frac{1}{2})$. El error de redondeo excede 1 si y solamente si $|\sum_{i=1}^{12} X_i| > 1$. Puesto que $\mathbb{E}[X_i] = 0$ y $\mathbb{V}(X_i) = \frac{1}{12}$ de acuerdo con el teorema central del límite tenemos que la distribución de

$$\frac{\sum_{i=1}^{12} X_i - 12\mathbb{E}[X_i]}{\sqrt{12\mathbb{V}(X_i)}} = \sum_{i=1}^{12} X_i$$

se puede aproximar por la distribución normal estándar. En consecuencia,

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{i=1}^{12} X_i\right| > 1\right) &= 1 - \mathbb{P}\left(\left|\sum_{i=1}^{12} X_i\right| \leq 1\right) \approx 1 - (\Phi(1) - \Phi(-1)) \\ &= 1 - (2\Phi(1) - 1) = 2 - 2\Phi(1) = 0.3173\dots \end{aligned}$$

□

Ejemplo 3.5 (Suma de exponenciales). La suma S_n de n variables aleatorias independientes exponenciales de intensidad $\lambda = 1$ obedece a una distribución gamma, $S_n \sim \Gamma(n, 1)$. En la siguiente figura se comparan, para distintos valores de n , la función de distribución de la suma estandarizada $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{V}(S_n)}}$ con la función de distribución normal estándar.

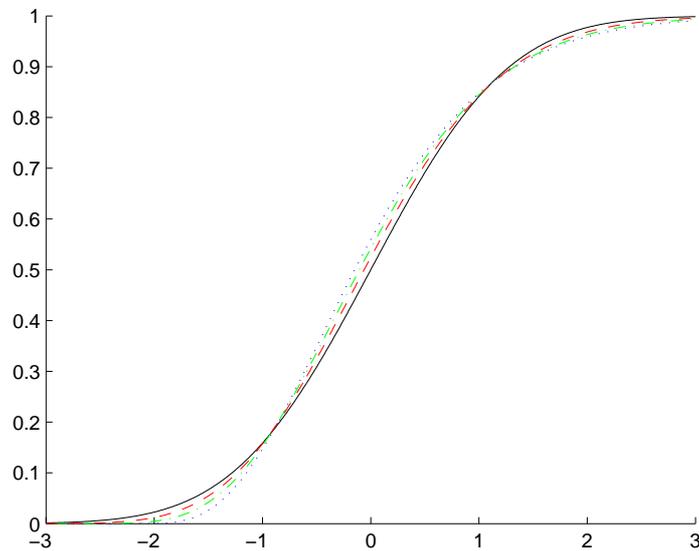


Figura 6: La normal estándar (sólida) y las funciones de distribución de las variables $\Gamma(n, 1)$ estandarizadas para $n = 5$ (punteada), $n = 10$ (quebrada y punteada) y $n = 30$ (quebrada).

Ejemplo 3.6. La distribución de Poisson de media λ se puede aproximar por la normal para valores grandes de λ : si $N \sim \text{Poisson}(\lambda)$, entonces

$$\frac{N - \lambda}{\sqrt{\lambda}} \approx \mathcal{N}(0, 1).$$

□

Ejemplo 3.7. Si la emisión de una cierta clase de partículas obedece a un proceso de Poisson de intensidad 900 por hora, ¿cuál es la probabilidad de que se emitan más de 950 partículas en una hora determinada?

Solución: Sea N una variable Poisson de media 900. Calculamos $\mathbb{P}(N > 950)$ estandarizando

$$\mathbb{P}(N > 950) = \mathbb{P}\left(\frac{N - 900}{\sqrt{900}} > \frac{950 - 900}{\sqrt{900}}\right) \approx 1 - \Phi\left(\frac{5}{3}\right) = 0.04779.$$

□

Ejemplo 3.8. El tiempo de vida de una batería es una variable aleatoria de media 40 horas y desvío 20 horas. Una batería se usa hasta que falla, momento en el cual se la reemplaza por una nueva. Suponiendo que se dispone de un stock de 25 baterías, cuyos tiempos de vida son independientes, aproximar la probabilidad de que pueda obtenerse un uso superior a las 1100 horas.

Solución: Si ponemos X_i para denotar el tiempo de vida de la i -ésima batería puesta en uso, lo que buscamos es el valor de $p = \mathbb{P}(X_1 + \dots + X_{25} > 1000)$, que puede aproximarse de la siguiente manera:

$$p = \mathbb{P}\left(\frac{\sum_{i=1}^{25} X_i - 1000}{20\sqrt{25}} > \frac{1100 - 1000}{20\sqrt{25}}\right) \approx 1 - \Phi(1) = 0.1587.$$

□

Ejemplo 3.9. El peso W (en toneladas) que puede resistir un puente sin sufrir daños estructurales es una variable aleatoria con distribución normal de media 1400 y desvío 100. El peso (en toneladas) de cada camión de arena es una variable aleatoria de media 22 y desvío 0.25. Calcular la probabilidad de que ocurran daños estructurales cuando hay 64 camiones de arena sobre el tablero del puente.

Solución: Ocurren daños estructurales cuando la suma de los pesos de los 64 camiones, X_1, \dots, X_{64} , supera al peso W . Por el teorema central del límite, la distribución de la suma $\sum_{i=1}^{64} X_i$ es aproximadamente una normal de media 1408 y desvío 2. En consecuencia,

$W - \sum_{i=1}^{64} X_i$ se distribuye (aproximadamente) como una normal de media $1400 - 1408 = -8$ y varianza $10000 + 4 = 10004$. Por lo tanto,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{64} X_i > W\right) &= \mathbb{P}\left(W - \sum_{i=1}^{64} X_i < 0\right) = \mathbb{P}\left(\frac{W - \sum_{i=1}^{64} X_i + 8}{\sqrt{10004}} < \frac{8}{\sqrt{10004}}\right) \\ &\approx \Phi(0.07998\dots) = 0.5318\dots \end{aligned}$$

□

Ejercicios adicionales

1. Un astronauta deberá permanecer 435 días en el espacio y tiene que optar entre dos alternativas. Utilizar 36 tanques de oxígeno de tipo A o 49 tanques de oxígeno de tipo B . Cada tanque de oxígeno de tipo A tiene un rendimiento de media 12 días y desvío $1/4$. Cada tanque de oxígeno de tipo B tiene un rendimiento de media de 8,75 días y desvío $25/28$. ¿Qué alternativa es la más conveniente?
 2. 432 números se redondean al entero más cercano y se suman. Suponiendo que los errores individuales de redondeo se distribuyen uniformemente sobre el intervalo $(-0.5, 0.5)$, aproximar la probabilidad de que la suma de los números redondeados difiera de la suma exacta en más de 6.
 3. Dos aerolíneas A y B que ofrecen idéntico servicio para viajar de Buenos Aires a San Pablo compiten por la misma población de 400 clientes, cada uno de los cuales elige una aerolínea al azar. ¿Cuál es la probabilidad de que la línea A tenga más clientes que sus 210 asientos?
 4. Completamente borracho, Lucas, ejecuta un “paseo aleatorio” de la siguiente manera: cada minuto da un paso hacia el norte con probabilidad $1/4$ o un paso hacia el sur con probabilidad $3/4$. Las direcciones de los pasos son independientes y en cada paso Lucas avanza 50 cm. en la dirección elegida. Usar el teorema central del límite para aproximar la distribución de probabilidades de la ubicación de Lucas luego de una hora. ¿Dónde es más probable encontrarlo?
-

4. Distribuciones relacionadas con la Normal

En esta sección se presentan tres distribuciones de probabilidad relacionadas con la distribución normal: las distribuciones χ^2 , t y F . Esas distribuciones aparecen en muchos problemas estadísticos.

4.1. χ^2 (chi-cuadrado)

Definición 4.1 (Distribución chi-cuadrado con un grado de libertad). Si Z es una variable aleatoria con distribución normal estándar, la distribución de $U = Z^2$ se llama la distribución *chi-cuadrado con 1 grado de libertad*.

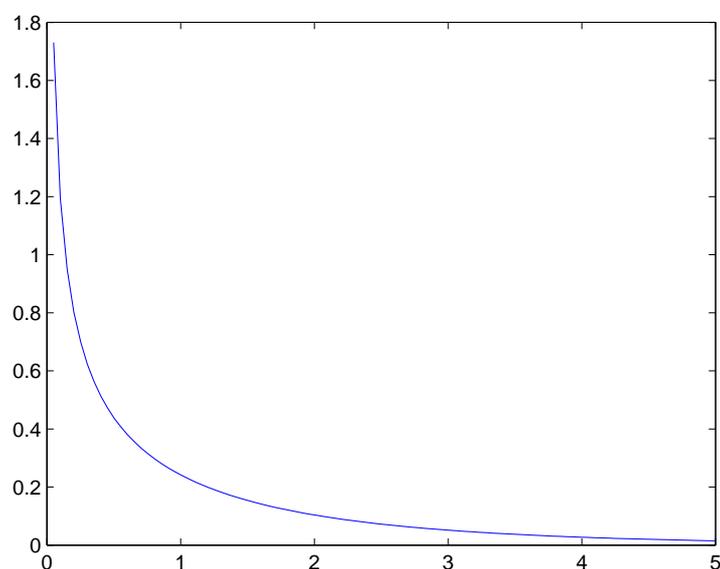


Figura 7: Gráfico de la función densidad de probabilidad de la distribución χ_1^2 .

Caracterización de la distribución χ_1^2 . La función de distribución de la variable $U = Z^2$ es $F_U(u) = \mathbb{P}(Z^2 \leq u)$, donde Z es $\mathcal{N}(0, 1)$. Para cada $u > 0$, vale que

$$F(x) = \mathbb{P}(Z^2 \leq u) = \mathbb{P}(|Z| \leq \sqrt{u}) = \mathbb{P}(-\sqrt{u} \leq Z \leq \sqrt{u}) = \int_{-\sqrt{u}}^{\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

Usando el teorema fundamental del cálculo integral y la regla de la cadena obtenemos que para cada $u > 0$ vale que

$$\begin{aligned}
 f_U(u) &= \frac{d}{du} F_U(u) = \frac{d}{du} \int_{-\sqrt{u}}^{\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
 &= \frac{1}{\sqrt{2\pi}} \left(e^{-(\sqrt{u})^2/2} \frac{d}{du}(\sqrt{u}) - e^{-(-\sqrt{u})^2/2} \frac{d}{du}(-\sqrt{u}) \right) \\
 &= \frac{1}{\sqrt{2\pi}} \left(e^{-u/2} \frac{1}{2\sqrt{u}} + e^{-u/2} \frac{1}{2\sqrt{u}} \right) = \frac{1}{\sqrt{2\pi}} \left(e^{-u/2} \frac{1}{\sqrt{u}} \right) \\
 &= \frac{(1/2)^{\frac{1}{2}}}{\sqrt{\pi}} (u^{-1/2} e^{-(1/2)u}) = \frac{(1/2)^{\frac{1}{2}}}{\sqrt{\pi}} u^{\frac{1}{2}-1} e^{-(1/2)u}. \tag{30}
 \end{aligned}$$

La última expresión que aparece en el lado derecho de la identidad (30) es la expresión de la densidad de la distribución $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Por lo tanto,

$$\chi_1^2 = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right).$$

Nota Bene. Notar que si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, y por lo tanto $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$.

Definición 4.2 (Distribución chi-cuadrado). Si U_1, U_2, \dots, U_n son variables aleatorias independientes, cada una con distribución χ_1^2 , la distribución de $V = \sum_{i=1}^n U_i$ se llama distribución *chi-cuadrado con n grados de libertad* y se denota χ_n^2 .

Caracterización de la distribución chi-cuadrado. La distribución χ_n^2 es un caso particular de la distribución Gamma. Más precisamente,

$$\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

Basta recordar que la suma de variables Γ i.i.d. también es Γ . En particular, la función densidad de V es

$$f_V(v) = \frac{(1/2)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} v^{\frac{n}{2}-1} e^{-\frac{1}{2}v} \mathbf{1}\{v > 0\}.$$

□

Nota Bene. La distribución χ_n^2 no es simétrica.

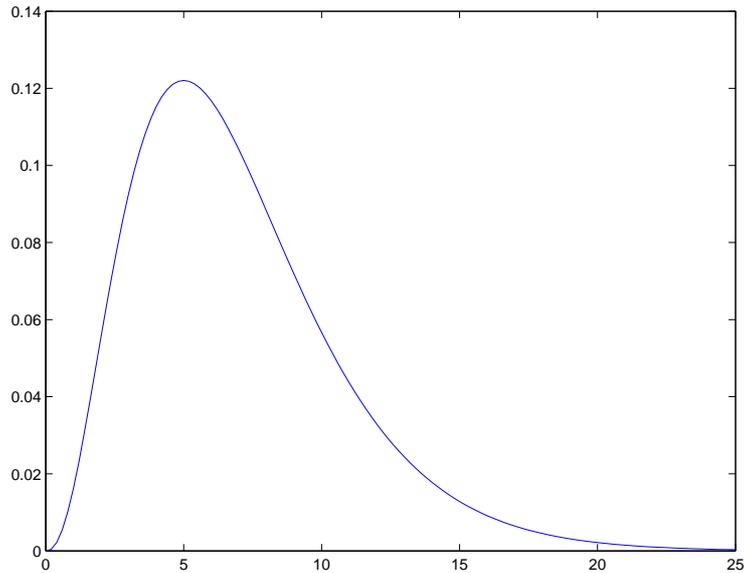


Figura 8: Gráfico de la función densidad de probabilidad de la distribución χ_7^2 .

4.2. t de Student

Definición 4.3 (La distribución t de Student). Sean Z y U variables aleatorias independientes con distribuciones $\mathcal{N}(0, 1)$ y χ_n^2 , respectivamente. La distribución de la variable

$$T = \frac{Z}{\sqrt{U/n}}$$

se llama *distribución t de Student con n grados de libertad* y se denota mediante t_n .

La función densidad de la t de Student con n grados de libertad es

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

La fórmula de la densidad se obtiene por los métodos estándar desarrollados en las notas sobre transformaciones de variables.

Observación 4.4. Notar que la densidad de t_n es simétrica respecto del origen. Cuando la cantidad de grados de libertad, n , es grande la distribución t_n se aproxima a la distribución $\mathcal{N}(0, 1)$; de hecho para más de 20 o 30 grados de libertad, las distribuciones son muy cercanas. .

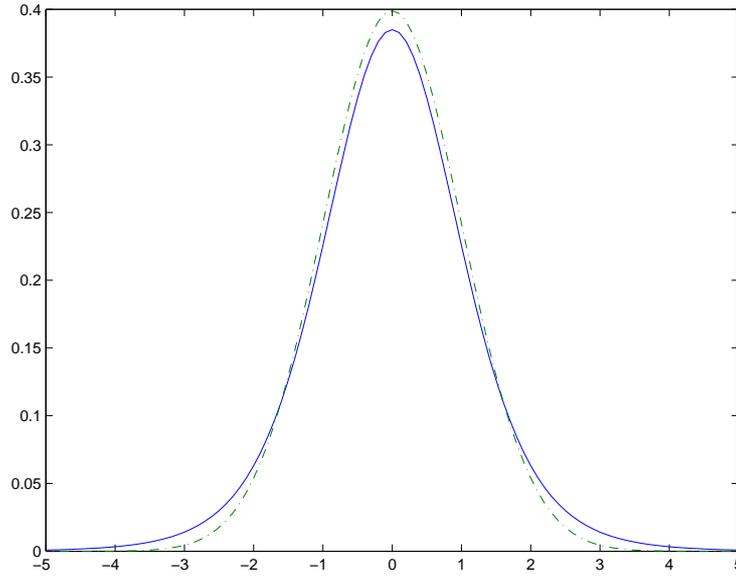


Figura 9: Comparación de la función densidad de probabilidad de una distribución t_7 (línea sólida) con la de la distribución $\mathcal{N}(0, 1)$ (línea punteada).

4.3. F de Fisher

Definición 4.5 (Distribución F). Sean U y V variables aleatorias independientes con distribuciones χ_m^2 y χ_n^2 , respectivamente. La distribución de la variable

$$W = \frac{U/m}{V/n}$$

se llama distribución F con m y n grados de libertad y se denota por $F_{m,n}$.

La función densidad de W es

$$f_W(w) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} w^{\frac{m}{2}-1} \left(1 + \frac{m}{n}w\right)^{-\frac{m+n}{2}} \mathbf{1}\{w \geq 0\}.$$

W es el cociente de dos variables aleatorias independientes, y su densidad se obtiene usando los métodos estándar desarrollados en las notas sobre transformaciones de variables.

Nota Bene. Se puede mostrar que, para $n > 2$, $\mathbb{E}[W] = n/(n-2)$. De las definiciones de las distribuciones t y F , se deduce que el cuadrado de una variable aleatoria t_n se distribuye como una $F_{1,n}$.

¿Cómo usar las tablas de las distribuciones F ? Para cada $\alpha \in (0, 1)$, sea $F_{\alpha,m,n}$ el punto del semieje positivo de las abscisas a cuya derecha la distribución $F_{m,n}$ acumula una probabilidad α :

$$\mathbb{P}(F_{m,n} > F_{\alpha,m,n}) = \alpha.$$

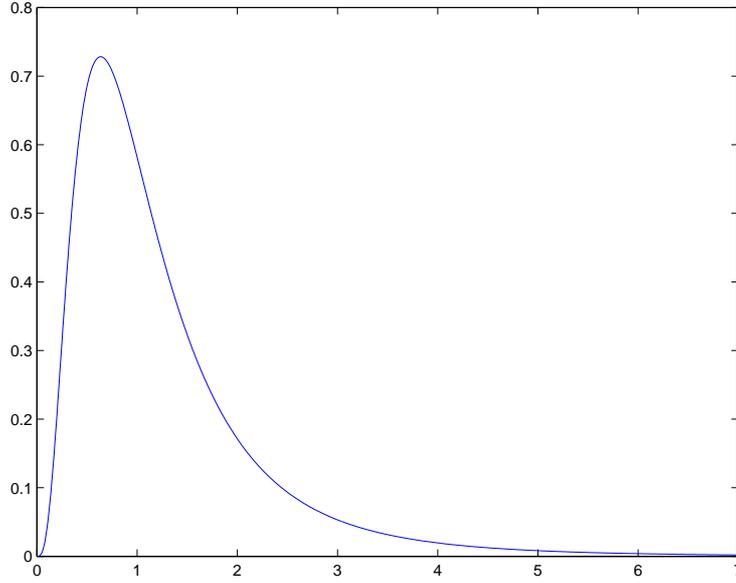


Figura 10: Gráfico típico de la función densidad de probabilidad de una distribución F .

Observación 4.6. Notar que de las igualdades

$$\alpha = \mathbb{P}\left(\frac{U/m}{V/n} > F_{\alpha,m,n}\right) = \mathbb{P}\left(\frac{V/n}{U/m} < \frac{1}{F_{\alpha,m,n}}\right) = 1 - \mathbb{P}\left(\frac{V/n}{U/m} \geq \frac{1}{F_{\alpha,m,n}}\right)$$

se deduce que

$$F_{1-\alpha,n,m} = \frac{1}{F_{\alpha,m,n}}. \quad (31)$$

En los manuales de estadística se pueden consultar las tablas de los valores $F_{\alpha,m,n}$ para diferentes valores de m, n y $\alpha \in \{0.01, 0.05\}$. Por ejemplo, según la tabla que tengo a mi disposición¹

$$\mathbb{P}(F_{9,9} > 3.18) = 0.05 \quad \text{y} \quad \mathbb{P}(F_{9,9} > 5.35) = 0.01$$

Usando esa información queremos hallar valores ϕ_1 y ϕ_2 tales que

$$\mathbb{P}(F_{9,9} > \phi_2) = 0.025 \quad \text{y} \quad \mathbb{P}(F_{9,9} < \phi_1) = 0.025.$$

El valor de ϕ_2 se obtiene por interpolación lineal entre los dos puntos dados en la tabla: $A = (3.18, 0.05)$ y $B = (5.35, 0.01)$. La ecuación de la recta que pasa por ellos es $y - 0.01 = -\frac{0.04}{2.17}(x - 5.35)$. En consecuencia, ϕ_2 será la solución de la ecuación $0.025 - 0.01 = -\frac{0.04}{2.17}(\phi_2 - 5.35)$. Esto es, $\phi_2 = 4.5362$.

El valor de ϕ_1 se obtiene observando que la ecuación $\mathbb{P}(F_{9,9} < \phi_1) = 0.025$ es equivalente a la ecuación $\mathbb{P}(1/F_{9,9} > 1/\phi_1) = 0.025$. Por definición, la distribución de $1/F_{9,9}$ coincide

¹Introducción a la estadística matemática. Ariel, Barcelona. (1980).

con la de $F_{9,9}$. En consecuencia, ϕ_1 debe satisfacer la ecuación $\mathbb{P}(F_{9,9} > 1/\phi_1) = 0.025$. Por lo tanto, $\phi_1 = 1/4.5362 = 0.2204$.

5. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Cramer, H.: Métodos matemáticos de estadística. Aguilar, Madrid. (1970)
2. Durrett R.: Probability. Theory and Examples. Duxbury Press, Belmont. (1996)
3. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 1. John Wiley & Sons, New York. (1968)
4. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 2. John Wiley & Sons, New York. (1971)
5. Hoel P. G.: Introducción a la estadística matemática. Ariel, Barcelona. (1980)
6. Piskunov, N.: Cálculo diferencial e integral, tomo I. Mir, Moscú (1983)
7. Rice, J. A.: Mathematical Statistics and Data Analysis. Duxbury Press, Belmont. (1995)
8. Ross, S. M: Introduction to Probability and Statistics for Engineers and Scientists. Elsevier Academic Press, San Diego. (2004)
9. Ross, S.: Introduction to Probability Models. Academic Press, San Diego. (2007)