

## Ensayos de Bernoulli

### 2.1. Distribuciones Asociadas a Ensayos de Bernoulli

#### 2.1.1. La Distribución de Bernoulli

Esta es la distribución más sencilla y corresponde a una variable que toma sólo dos valores: 1 con probabilidad  $p$  y 0 con probabilidad  $q = 1 - p$ .

Si  $A$  es un evento y definimos la *función indicadora* de  $A$  por

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Esta variable aleatoria vale 1 cuando  $A$  ocurre y 0 cuando no, es decir, nos indica cuándo ocurre  $A$ . Por lo tanto  $\mathbf{1}_A$  tiene distribución de Bernoulli con  $p = P(A)$ .

Es sencillo calcular los momentos de esta distribución:

$$E(X) = 1 \times p + 0 \times q = p. \quad (2.1)$$

Como  $X = X^2$ , porque la variable solo toma los valores 0 y 1, tenemos que  $E(X^2) = p$  y en consecuencia:

$$\text{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p) \quad (2.2)$$

#### 2.1.2. La Distribución Binomial.

Para  $n \in \mathbb{N}$  realizamos una serie de experimentos de Bernoulli independientes con igual probabilidad de éxito  $p$  y llamemos  $S_n$  al número total de éxitos en los  $n$  ensayos. Cada ensayo corresponde a una variable de Bernoulli  $X_i, 1 \leq i \leq n$  y

$$S_n = X_1 + X_2 + \cdots + X_n. \quad (2.3)$$

El resultado de una serie de experimentos lo podemos representar por un vector  $n$ -dimensional  $(X_1, X_2, \dots, X_n)$  cuyas entradas valen 0 ó 1 según el resultado del experimento haya sido éxito o fracaso. Por ejemplo el vector  $(1, 1, 0)$  corresponde a una serie de tres experimentos en los cuales los dos primeros resultaron en éxito y el tercero en fracaso.

Para hallar la distribución de  $S_n$  comenzamos por responder la siguiente pregunta: ¿Cuál es la probabilidad de obtener el vector  $(i_1, i_2, \dots, i_n)$  en una sucesión de  $n$  ensayos? Como las variables son independientes

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = P(X_1 = i_1)P(X_2 = i_2) \cdots P(X_n = i_n)$$

Ahora bien, las probabilidades que multiplicamos a la derecha solo pueden valer  $p$  ó  $q$ , según correspondan a un éxito o a un fracaso. Por lo tanto, si hay  $S_n$  éxitos en los  $n$  ensayos, sin importar dónde estén ubicados estos éxitos, esta probabilidad será  $p^{S_n}q^{n-S_n}$ :

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = p^{S_n}q^{n-S_n}.$$

Queremos ahora hallar  $P(S_n = k)$  para  $0 \leq k \leq n$ . Para esto tenemos que multiplicar la probabilidad  $p^k q^{n-k}$  por el número de resultados que tienen exactamente  $k$  éxitos. Este número es igual a las maneras de seleccionar  $k$  lugares en  $n$ , que es igual a  $\binom{n}{k}$ . En conclusión

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = p_{k,n} \quad (k = 0, 1, \dots, n). \quad (2.4)$$

En general, si una variable aleatoria discreta  $X$  tiene esta función de probabilidad decimos que  $X$  tiene una *distribución binomial* con parámetros  $n$  y  $p$ . En este caso usaremos la notación  $X \sim b(n, p)$

Si  $p = 1/2$  la función de probabilidad es simétrica con respecto a  $n/2$ , ya que en este caso  $P(d_n = k) = P(d_n = n - k)$ .

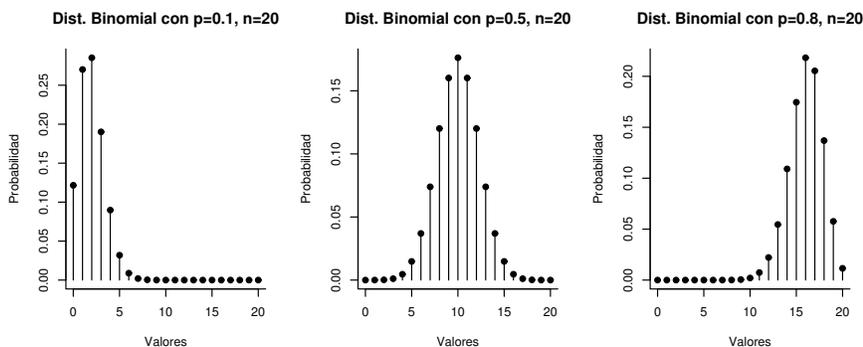


Figura 2.1 Distribución binomial para  $n = 20$  y tres valores de  $p$ .

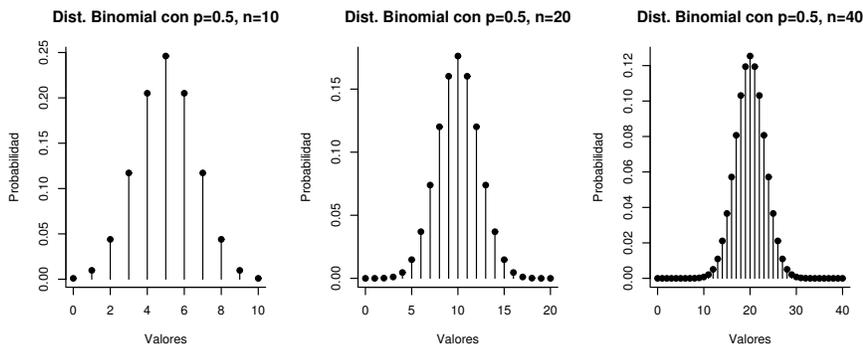


Figura 2.2 Distribución binomial para  $p = 0.5$  y tres valores de  $n$ .

### Ejemplo.

Se extraen con reposición cinco cartas de un juego de barajas. Sea  $X$  el número de diamantes en la muestra. ¿Cuál es la probabilidad de que haya exactamente dos diamantes entre las cinco cartas? ¿Cuál es la probabilidad de que haya a lo sumo dos diamantes?

- Para responder la primera pregunta queremos calcular  $P(X = 2)$ , y como la probabilidad de obtener un diamante en cada extracción es  $1/4$  tenemos:

$$P(X = 2) = \binom{5}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 = 0.264$$

Para la segunda pregunta tenemos que:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{5}{0} \left(\frac{3}{4}\right)^5 + \binom{5}{1} \left(\frac{1}{4}\right) \left(\frac{3}{4}\right)^4 + \binom{5}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 \\ &= 0.237 + 0.396 + 0.264 \\ &= 0.897 \end{aligned}$$

▲

### Relación recursiva

Podemos obtener una relación recursiva entre los términos de la distribución. Si  $X \sim b(n, p)$  tenemos

$$\begin{aligned} P(X = k + 1) &= \binom{n}{k + 1} p^{k+1} (1 - p)^{n-k-1} \\ &= \frac{n!}{(k + 1)!(n - k - 1)!} p^{k+1} (1 - p)^{n-k-1} \\ &= \frac{n - k}{k + 1} \frac{n!}{k!(n - k)!} \left(\frac{p}{1 - p}\right) p^k (1 - p)^{n-k} \\ &= \frac{n - k}{k + 1} \left(\frac{p}{1 - p}\right) P(X = k). \end{aligned} \tag{2.5}$$

Podemos usar esta relación comenzando en  $P(X = 0) = (1 - p)^n$  o en  $P(X = n) = p^n$  para calcular los valores de la distribución.

### Momentos

A partir de la expresión (2.3) Se deduce que

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = np.$$

Además, en virtud de que  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes

$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

y

$$\text{Var}(X_i) = E(X_i^2) - (E(X_i))^2 = p - p^2 = p(1 - p).$$

Reemplazando en la igualdad anterior, resulta

$$\text{Var}(S_n) = np(1 - p)$$

### 2.1.3. La Distribución de Poisson.

Decimos que la variable aleatoria  $X$  tiene *distribución de Poisson* con parámetro  $\lambda$ , ( $\lambda > 0$ ) si

$$P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda} \quad (n = 0, 1, 2, \dots).$$

Esta relación define efectivamente una función de probabilidad ya que, usando el desarrollo en serie de Taylor de la función exponencial,

$$\sum_{n=0}^{\infty} P(X = n) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1,$$

y es un ejemplo de una variable aleatoria que toma valores en un conjunto numerable. Usaremos la notación  $X \sim \mathcal{P}(\lambda)$

Esta distribución tiene numerosas aplicaciones y gran interés en sí misma, pero además es útil como aproximación a la distribución binomial para  $n$  grande y  $p$  pequeño, hecho que estudiaremos a continuación.

Consideremos la distribución binomial cuando  $n$  crece y  $p_n$  tiende a cero de manera tal que el producto  $np_n$  permanece fijo. Sea  $np_n = \mu > 0$ . La distribución binomial es

$$p_{k,n} = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} p_n^k (1 - p_n)^{n-k}$$

multiplicando numerador y denominador por  $n^k$  obtenemos

$$\begin{aligned} p_{k,n} &= \frac{n(n-1) \cdots (n-k+1)}{n^k k!} (np_n)^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{n^k} \frac{\mu^k}{k!} (1 - p_n)^{n-k} \\ &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\mu^k}{k!} (1 - p_n)^{n-k} \\ &= \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{(1 - p_n)^k} \frac{\mu^k}{k!} (1 - p_n)^n \end{aligned} \quad (2.6)$$

Ahora bien, podemos escribir

$$(1 - p_n)^n = [(1 - p_n)^{-1/p_n}]^{-np_n} = [(1 - p_n)^{-1/p_n}]^{-\mu}$$

pero por la definición de  $e$  sabemos que

$$\lim_{z \rightarrow 0} (1 + z)^{1/z} = e.$$

Por lo tanto, si ponemos  $z = -p_n$  obtenemos

$$\lim_{p \rightarrow 0} (1 - p_n)^n = \lim_{p \rightarrow 0} [(1 - p_n)^{-1/p_n}]^{-\mu} = e^{-\mu}.$$

Además

$$\lim_{n \rightarrow \infty} \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{(1 - p_n)^k} = 1$$

ya que hemos supuesto que  $p_n \rightarrow 0$  cuando  $n \rightarrow \infty$  y  $np_n = \mu$  permanece constante. Usando estos dos resultados en (2.6) obtenemos

$$\lim_{n \rightarrow \infty} p_{k,n} = \frac{e^{-\mu} \mu^k}{k!}.$$

Hemos demostrado el siguiente resultado.

**Teorema 2.1 (de Aproximación de Poisson)** *Sea  $X_n \sim b(n, p_n)$  y supongamos que cuando  $n \rightarrow \infty$ ,  $p_n \rightarrow 0$  de modo que  $np_n$  permanece constante y es igual a  $\mu$ . Entonces, cuando  $n \rightarrow \infty$*

$$p_{k,n} = P(X_n = k) \rightarrow e^{-\mu} \frac{\mu^k}{k!}.$$

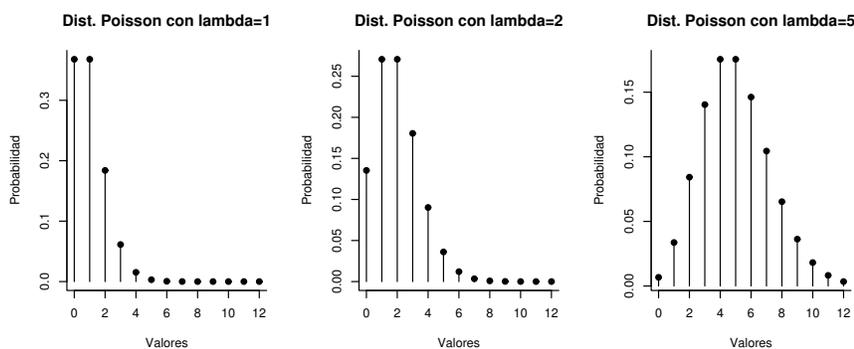


Figura 2.3 Distribución de Poisson para tres valores de  $\lambda$ .

### Ejemplos.

1. Las llamadas que se reciben en una central telefónica por minuto tienen distribución de Poisson con parámetro  $\mu = 4$ . Si la central puede manejar un máximo de seis llamadas por minuto, ¿cuál es la probabilidad de que la central sea insuficiente para atender las llamadas que llegan en un minuto?

► Sea  $X$  el número de llamadas que se reciben en un período de un minuto. Calculemos primero

$$P(X \leq 6) = \sum_{i=0}^6 P(X = i) = \sum_{i=0}^6 \frac{e^{-4} 4^i}{i!} = 0.889,$$

por lo tanto

$$P(X > 6) = 1 - P(X \leq 6) = 1 - 0.889 = 0.11. \quad \blacktriangle$$

2. Se toma una muestra de 400 fusibles fabricados usando un procedimiento que, en promedio, produce 1% de fusibles defectuosos. ¿Cuál es la probabilidad de que, a lo sumo, haya 5 fusibles defectuosos en la muestra?

► Sea  $X$  el número de fusibles defectuosos en la muestra. Sabemos que  $X$  tiene una distribución binomial con  $n = 400$ ,  $p = 0.01$  y deseamos calcular

$$P(X \leq 5) = \sum_{i=0}^5 P(X = i) = \sum_{i=0}^5 \binom{400}{i} (0.01)^i (0.99)^{400-i}$$

y el cálculo de esta suma es trabajoso. Por lo tanto utilizaremos la distribución de Poisson con parámetro

$$\mu = np = 400 \times 0.01 = 4,$$

para aproximar la distribución binomial.

$$\begin{aligned} P(X \leq 5) &= \sum_{i=0}^5 \frac{e^{-4} 4^i}{i!} = e^{-4} \left( 1 + 4 + \frac{4^2}{2} + \frac{4^3}{6} + \frac{4^4}{24} + \frac{4^5}{120} \right) \\ &= 0.785. \end{aligned}$$

▲

### Relación recursiva

Para la distribución de Poisson también hay una relación recursiva que permite calcular sus valores. Si  $X \sim \mathcal{P}(\lambda)$  tenemos

$$\frac{P(X = i + 1)}{P(X = i)} = \frac{e^{-\lambda} \lambda^{i+1} / (i + 1)!}{e^{-\lambda} \lambda^i / i!} = \frac{\lambda}{i + 1}$$

es decir,

$$P(X = i + 1) = \frac{\lambda}{i + 1} P(X = i), \quad i \geq 0. \quad (2.7)$$

### Momentos

Tenemos

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda, \end{aligned}$$

donde hemos empleado el desarrollo de Taylor de la función exponencial  $e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$ . Por lo tanto

$$E(X) = \lambda.$$

Además

$$\begin{aligned} E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2 \end{aligned}$$

y en consecuencia

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 = E(X(X-1)) + E(X) - (E(X))^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

Es decir

$$\text{Var}(X) = \lambda.$$

### 2.1.4. La Distribución Geométrica

Consideremos un fusible eléctrico que no se deteriora con el paso del tiempo pero que se quema debido a fallas en la corriente eléctrica que ocurren al azar pero en forma homogénea en el tiempo. El fusible es observado cada día y llamaremos  $X$  al número de días que transcurren hasta que el fusible falla, suponiendo que el día cero el fusible es nuevo. Queremos hallar la función de probabilidades de  $X$ .

La idea de que el fusible no se deteriora con el paso del tiempo se puede expresar con mayor precisión de la manera siguiente: si sabemos que el fusible no ha fallado antes o durante el día  $n$ , es decir,  $X > n$ , entonces la probabilidad de que no falle hasta después del día  $n + m$ ,  $P(X > n + m | X > n)$  debe ser igual a la probabilidad de que un fusible nuevo el día  $n$  no falle hasta después del día  $n + m$ .

Como las fallas eléctricas que hacen que el fusible falle ocurren en forma homogénea en el tiempo, esta probabilidad debe depender solamente del número de días transcurridos, que es  $m$ , pero no de  $n$ . Por lo tanto tenemos la ecuación

$$P(X > n + m | X > n) = P(X > m)$$

y usando la definición de probabilidad condicional podemos reescribir esta identidad como

$$P(X > n + m) = P(X > n)P(X > m) \quad n, m = 0, 1, 2, \dots \quad (2.8)$$

Si hacemos  $n = m = 0$  obtenemos

$$P(X > 0) = (P(X > 0))^2$$

y por lo tanto  $P(X > 0) = 0$  ó  $1$ . Si  $P(X > 0) = 0$  entonces  $P(X = 0) = 1$ , lo cual corresponde a un fusible defectuoso y no es interesante. Por lo tanto,  $P(X > 0) = 1$ .

Llamemos  $p = P(X = 1)$ , entonces

$$P(X > 1) = 1 - p$$

y usando (2.8) con  $m = 1$  obtenemos

$$P(X > n + 1) = (1 - p)P(X > n).$$

Iterando en  $n$  obtenemos que

$$P(X > n) = (1 - p)^n$$

y por lo tanto

$$\begin{aligned} P(X = n) &= P(X > n - 1) - P(X > n) \\ &= (1 - p)^{n-1} - (1 - p)^n \\ &= p(1 - p)^{n-1} \end{aligned}$$

para  $n \geq 1$ .

**Definición 2.1** Decimos que la variable aleatoria  $Y$  tiene *distribución geométrica* si su función de probabilidad es

$$P(Y = n) = \begin{cases} p(1 - p)^{n-1} & n = 0, 1, 2, \dots \\ 0 & \text{para cualquier otro } n \end{cases}$$

donde  $0 < p < 1$ . Usaremos la notación  $X \sim \mathcal{G}(p)$  en este caso.

Observamos que en el ejemplo anterior la variable  $X$  tiene distribución geométrica.

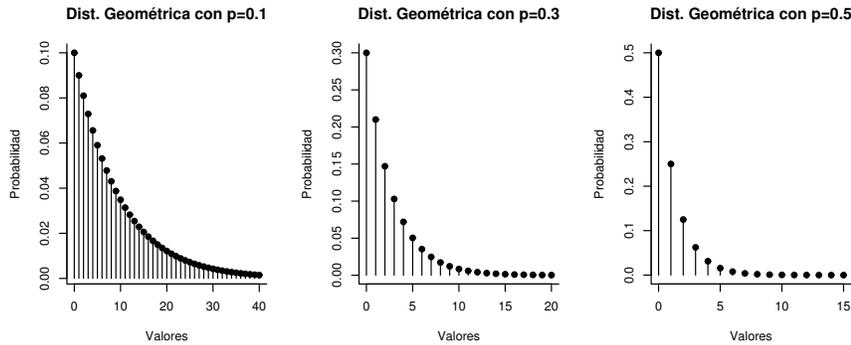


Figura 2.4 Distribución Geométrica para tres valores de  $p$ .

### Momentos

Tenemos

$$E(X) = \sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} np(1-p)^{n-1}. \quad (2.9)$$

Para calcular esta suma, recordemos que si

$$\sum_{n=0}^{\infty} c_n z^n$$

es una serie de potencias y  $R$  es su radio de convergencia (lo cual significa que si  $|z| < R$  la serie converge y que, cuando  $R < \infty$ , si  $|z| > R$  la serie no converge), si denotamos por  $f(z)$  la suma de la serie para  $|z| < R$ :

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \quad (|z| < R),$$

entonces  $f$  es derivable y

$$f'(z) = \sum_{n=0}^{\infty} n c_n z^{n-1} \quad (|z| < R).$$

Lo cual significa que, para  $|z| < R$ , podemos calcular la derivada de la función  $f(z)$  como si fuera una suma finita, es decir, sumando las derivadas de cada uno de los términos.

Apliquemos este resultado a la serie de potencias

$$\sum_{n=0}^{\infty} z^n.$$

Primero, se verifica que el radio de convergencia de esta serie es igual a 1. Esto es directo: si  $|z| > 1$  entonces  $|z|^n \rightarrow \infty$ , es decir que el término general no tiende a cero y la serie no converge, y si  $|z| < 1$ , la suma parcial de orden  $N$  es

$$\sum_{n=0}^N z^n = 1 + z + z^2 + \dots + z^N = \frac{1 - z^{N+1}}{1 - z} \xrightarrow{N \rightarrow \infty} \frac{1}{1 - z}$$

es decir que la serie converge cuando  $|z| < 1$  y su suma es  $\frac{1}{1-z}$ , o sea

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1 - z} \quad |z| < 1.$$

Derivando término a término como hemos indicado anteriormente, se tiene

$$\frac{1}{(1-z)^2} = \sum_{n=0}^{\infty} nz^{n-1} = \sum_{n=1}^{\infty} nz^{n-1} \quad |z| < 1. \quad (2.10)$$

Volviendo ahora al problema que estábamos considerando, reemplazando  $z$  por  $1-p$  en la ecuación (2.10) resulta

$$E(X) = p \sum_{n=1}^{\infty} n(1-p)^{n-1} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Para calcular la varianza de esta distribución comenzamos por la relación

$$\text{Var}(X) = E(X(X-1)) + E(X) - (E(X))^2, \quad (2.11)$$

y calculamos

$$\begin{aligned} E(X(X-1)) &= \sum_{n=1}^{\infty} n(n-1)P(X=n) = \sum_{n=1}^{\infty} n(n-1)p(1-p)^{n-1} \\ &= p(1-p) \sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2}. \end{aligned}$$

Ahora bien, para calcular  $f''(z)$  procedemos como antes, derivando término a término la serie que define  $f'(z)$ . Es decir, que para  $|z| < 1$ ,

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \Rightarrow \frac{1}{(1-z)^2} = \sum_{n=1}^{\infty} nz^{n-1} \Rightarrow \frac{2}{(1-z)^3} = \sum_{n=2}^{\infty} n(n-1)z^{n-2}.$$

En consecuencia, reemplazando  $z$  por  $1-p$ , resulta

$$E(X(X-1)) = p(1-p) \frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}$$

y volviendo a (6.13)

$$\text{Var}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

### 2.1.5. La Distribución Binomial Negativa.

Esta distribución también se conoce como la Distribución de Pascal y aparece en el contexto de una sucesión de ensayos de Bernoulli con probabilidad de éxito  $p$ , cuando nos hacemos una pregunta similar a la realizada para la distribución geométrica, pero en lugar de preguntar por el número de ensayos necesarios para lograr el primer éxito, preguntamos por el número de ensayos necesarios para lograr  $k$  éxitos.

Sea  $X$  la variable descrita anteriormente.  $X$  vale  $n$  si y sólo si el  $k$ -ésimo éxito ocurre en el  $n$ -ésimo ensayo, esto es, en los primeros  $n-1$  ensayos hay  $k-1$  éxitos y en el  $n$ -ésimo ensayo hay un éxito. La probabilidad de esto último es  $p$ , mientras que la probabilidad de tener  $k-1$  éxitos en  $n-1$  ensayos es una distribución binomial:

$$\binom{n-1}{k-1} p^{k-1} q^{n-k}.$$

Como los ensayos son independientes, tenemos que la probabilidad  $P(X=n)$  es el producto de las dos expresiones anteriores, es decir,

$$P(X=n) = \binom{n-1}{k-1} p^k q^{n-k}$$

para  $n \geq 1, 1 \leq k \leq n$ .

**Ejemplo.**

Un pescador va todos los días al muelle y se queda pescando hasta que hayan pasado dos horas o hasta que logre pescar un pez. Si la probabilidad de que no pesque nada es 0.6, ¿cuál es la probabilidad de que tenga que esperar cinco días para pescar tres peces?

- Sea  $X$  el número de días necesarios para pescar tres peces. Esta variable tiene distribución binomial negativa con parámetros 3 y 0.4, por lo tanto

$$P(X = 5) = \binom{4}{2} (0.4)^3 (0.6)^2 = 0.138$$

▲

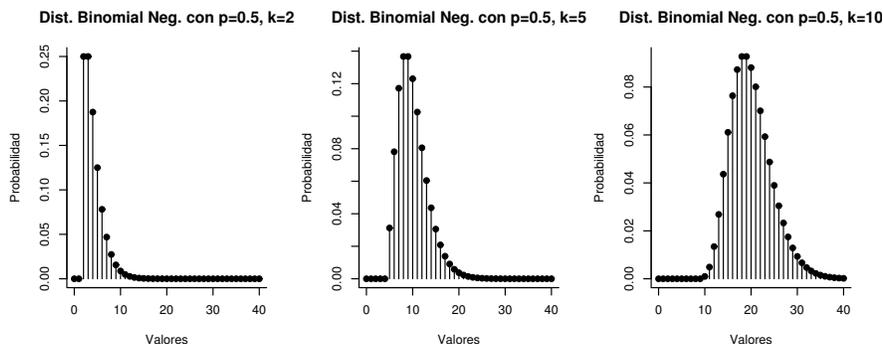


Figura 2.5 Distribución Binomial Negativa con  $p = 0.5$  para tres valores de  $k$ .

## 2.2. La Ley Débil de los Grandes Números.

**Lema 2.1 (Desigualdad de Markov)** Sea  $X \geq 0$  una variable aleatoria y a un número positivo, entonces

$$P(X \geq a) \leq \frac{1}{a} E(X). \quad (2.12)$$

*Demostración.* Si  $E(X) = +\infty$ , no hay nada que probar. En caso contrario, llamando  $A$  al evento  $A = \{X \geq a\}$ , se tienen las desigualdades siguientes,

$$X(\omega) \geq X(\omega)\mathbf{1}_A(\omega) \geq a\mathbf{1}_A(\omega).$$

La primera desigualdad se debe simplemente a que  $X(\omega) \geq 0$  y  $\mathbf{1}_A(\omega)$  es igual a 0 ó 1. En cuanto a la segunda, si  $\omega \notin A$  entonces  $\mathbf{1}_A(\omega) = 0$  y ambos miembros son iguales y si  $\omega \in A$ , por un lado  $\mathbf{1}_A(\omega) = 1$  y por otro, dada la definición del evento  $A$ ,  $X(\omega) \geq a$ . Se deduce que

$$E(X) \geq E(a\mathbf{1}_A) = a E(\mathbf{1}_A) = aP(A).$$

Dividiendo por  $a$  se obtiene (2.12). ■

**Lema 2.2 (Desigualdad de Chebychef)** Sea  $X$  una variable aleatoria con varianza finita y sea  $\varepsilon > 0$ . Entonces

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var}(X). \quad (2.13)$$

*Demostración.* El evento  $\{|X - E(X)| \geq \varepsilon\}$  es igual a  $\{(X - E(X))^2 \geq \varepsilon^2\}$ . Apliquemos la desigualdad de Markov poniendo  $\varepsilon^2$  en lugar de  $a$  y  $(X - E(X))^2$  en lugar de  $X$ . Resulta

$$P(|X - E(X)| \geq \varepsilon) = P((X - E(X))^2 \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} E((X - E(X))^2) = \frac{1}{\varepsilon^2} \text{Var}(X)$$

que es (2.13). ■

**Teorema 2.2 ( Ley Débil de los Grandes Números)** *Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias independientes,*

$$E(X_n) = \mu, \quad \text{Var}(X_n) = \sigma^2 \quad n = 1, 2, \dots$$

*Entonces, cualquiera que sea  $\varepsilon > 0$ , se tiene*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0. \quad (2.14)$$

*Demostración.* Pongamos  $S_n = X_1 + X_2 + \dots + X_n$  y apliquemos la desigualdad de Chebychef a la variable aleatoria  $S_n/n$ . Tenemos

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu,$$

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Aplicando (2.13) con  $\varepsilon > 0$

$$p\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \frac{\sigma^2}{n} \rightarrow 0$$

cuando  $n \rightarrow \infty$  (para  $\varepsilon > 0$  fijo). Esto prueba la ley débil de los grandes números. ■

### 2.2.1. Ejemplos y Comentarios.

1. En el caso de la distribución binomial, en el cual consideramos la variable aleatoria  $S_n$  que representa el número de veces que ocurre un suceso  $A$  con probabilidad  $p = P(A)$  en  $n$  observaciones independientes del mismo, hemos considerado oportunamente la estimación del parámetro  $p$  – en general desconocido – mediante la frecuencia relativa

$$\hat{p}_n = \frac{S_n}{n}$$

del número de veces que ocurre  $A$  en las  $n$  observaciones, con relación al total de observaciones realizadas. Se verifica entonces que

$$P(|\hat{p}_n - p| \geq \varepsilon) \rightarrow 0 \quad \text{para todo } \varepsilon > 0 \text{ cuando } n \rightarrow \infty. \quad (2.15)$$

La propiedad (2.15), de aproximación entre  $\hat{p}_n$  (que es función de las observaciones empíricas que podemos efectuar), y el número  $p$  (que es un parámetro del problema), nos indica que si el número  $n$  de observaciones es bastante grande, entonces es pequeña la probabilidad de que la distancia de  $\hat{p}_n$  a  $p$  sea mayor que un número dado.

La desigualdad de Chebychef nos dá, además, una acotación de la velocidad con la que el primer miembro de (2.15) tiende a cero cuando  $n \rightarrow \infty$ . En el ejemplo que estamos considerando, como ya hemos visto en la sección 2.1.2,

$$E(\hat{p}_n) = p \quad \text{Var}(\hat{p}_n) = \frac{p(1-p)}{n}.$$

y entonces

$$P(|\hat{p}_n - p| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \frac{p(1-p)}{n} \quad (2.16)$$

Esta desigualdad puede ser utilizada para resolver problemas como el siguiente:

Supongamos que  $p$  es la probabilidad – desconocida – de que resulte defectuoso un objeto producido en una fábrica y  $S_n$  es el número de objetos defectuosos en una muestra de  $n$  objetos observados en un muestreo al azar con reposición. Calculamos

$$\hat{p}_n = \frac{S_n}{n}$$

Con frecuencia se plantea el problema de hallar el tamaño  $n$  que debe tener la muestra para que se satisfagan ciertos márgenes de tolerancia. Por ejemplo, supongamos que desconocemos el verdadero valor de  $p$  y queremos estimarlo por  $\hat{p}_n$  pero de modo que la probabilidad de que  $p$  difiera de  $\hat{p}_n$  en a lo sumo 0.01, no sea inferior a 0.95. Es decir, queremos elegir  $n$  de modo que

$$P(|\hat{p}_n - p| \leq 0.01) \geq 0.95$$

o equivalentemente

$$P(|\hat{p}_n - p| \geq 0.01) \leq 0.05. \quad (2.17)$$

Usando (2.16) tenemos

$$P(|\hat{p}_n - p| \geq 0.01) \leq \frac{1}{(0.01)^2} \frac{p(1-p)}{n}. \quad (2.18)$$

Si no tenemos ninguna información adicional sobre  $p$  teniendo en cuenta que  $p(1-p) \leq 1/4$  para  $0 \leq p \leq 1$ , si elegimos  $n$  de modo que

$$\frac{1}{(0.01)^2} \frac{1}{4n} \leq 0.05, \quad (2.19)$$

resultará

$$P(|\hat{p}_n - p| \geq 0.01) \leq 0.05.$$

Es sencillo ver que (2.19) se verifica si  $n \geq 50.000$ .

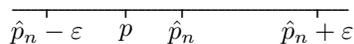


Figura 7.1

En principio, la acotación dada por la desigualdad de Chebychef es grosera. En el ejemplo numérico que estamos analizando, se podrían dar acotaciones mucho más precisas del primer miembro que figura en la desigualdad (2.18), lo cual permitiría reducir considerablemente el valor de  $n$  a partir del cual se verifica (2.17). En particular, si aproximamos por la distribución normal – cosa que veremos más adelante – se puede ver que si el verdadero valor de  $p$  es  $1/2$  (para dar algún valor) y  $n = 50.000$ , entonces el primer miembro de (2.17) está acotado por  $0.00001$ , es decir, es mucho menor que el segundo miembro. Retomaremos este punto en la próxima sección.

2. En el caso de la distribución de Poisson, que aparece en el estudio de fallas de aparatos que no envejecen, hemos visto que los tiempos de vida de los aparatos siguen una distribución exponencial del tipo

$$P(T > t) = e^{-\lambda t}, \quad t \geq 0,$$

donde  $\lambda$  es un parámetro positivo. El número (aleatorio) de fallas en el intervalo  $[0, t]$  tiene la función de probabilidad

$$P(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots$$

como veremos más adelante en el curso al estudiar el proceso de Poisson. Entonces, si

$$\hat{\lambda}_t = \frac{N_t}{t}$$

que representa el número de fallas por unidad de tiempo que efectivamente han ocurrido durante nuestro período de observación, se tiene

$$\begin{aligned} E(\hat{\lambda}_t) &= \frac{1}{t} E(N_t) = \frac{1}{t} \lambda t = \lambda \\ \text{Var}(\hat{\lambda}_t) &= \frac{1}{t^2} \text{Var}(N_t) = \frac{1}{t^2} \lambda t = \frac{\lambda}{t} \rightarrow 0 \quad (t \rightarrow +\infty) \end{aligned}$$

En consecuencia

$$P(|\hat{\lambda}_t - \lambda| \geq \varepsilon) < \frac{1}{\varepsilon^2} \frac{\lambda}{t},$$

para cada  $\varepsilon > 0$ ,

$$P(|\hat{\lambda}_t - \lambda| \geq \varepsilon) \rightarrow 0 \quad (t \rightarrow +\infty)$$

y podemos aplicar consideraciones análogas a las del ejemplo anterior.

3. En la demostración de la ley débil de los grandes números, hemos supuesto que las variables aleatorias  $X_1, X_2, \dots$  son independientes y tienen igual varianza. Se puede verificar, sin gran esfuerzo, que vale una demostración análoga si las variables no están correlacionadas y

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \rightarrow 0 \quad (n \rightarrow +\infty)$$

donde  $\sigma_i = \text{Var}(X_i)$ . En realidad, hay un buen número de generalizaciones de la ley débil de los grandes números, algunas de las cuales requieren complicaciones técnicas para su demostración, que están más allá del alcance de este curso. El lector interesado puede consultar, por ejemplo, los textos clásicos de M. Loève y de W. Feller, citados en la bibliografía.

4. Vale la pena tener en cuenta que la ley débil expresa, en algún sentido, la idea comúnmente aceptada de que se puede reemplazar promedios teóricos por promedios experimentales. El promedio teórico del fenómeno bajo consideración es  $\mu$ , es decir, la esperanza matemática, y el promedio experimental es

$$\frac{X_1 + X_2 + \dots + X_n}{n},$$

es decir, la media aritmética de las observaciones. La ley débil dice que si las observaciones se repiten *bajo ciertas condiciones* (independencia, etc.) en el sentido de la expresión (2.14), la media experimental está cerca del promedio teórico. Dicho de otro modo, si el número de observaciones es grande, resulta poco probable que uno y otro estén alejados.

### 2.2.2. El Teorema de Aproximación de Weierstrass

El teorema de aproximación de Weierstrass dice que cualquier función continua  $f$  definida en un intervalo acotado  $[a, b]$  puede ser aproximada de manera uniforme por polinomios: Dado  $\varepsilon > 0$  existe un polinomio  $g$  tal que

$$\sup_{x \in [a, b]} |f(x) - g(x)| < \varepsilon.$$

Serge Bernstein dió una demostración de este resultado usando la LDGN, que presentamos a continuación para funciones definidas en el intervalo  $[0, 1]$ . La extensión al caso general es sencilla.

**Proposición 2.1 (Polinomios de Bernstein)** Sea  $f : [0, 1] \rightarrow \mathbb{R}$  una función continua, entonces

$$\sup_{x \in [0,1]} \left| f(x) - \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \right| \rightarrow 0$$

cuando  $n \rightarrow \infty$ .

*Demostración.* Sea  $\varepsilon > 0$ . Como  $f$  es uniformemente continua, existe  $\delta > 0$  tal que si  $0 \leq x, y \leq 1$  satisfacen  $|x - y| < \delta$  entonces se tiene  $|f(x) - f(y)| < \varepsilon$ . Por otro lado consideramos variables binomiales  $S_n$  con parámetros  $n$  y  $x$ .

Consideremos ahora la variable aleatorias  $f(S_n/n)$  cuyo valor esperado es

$$\begin{aligned} \mathbb{E} \left[ f\left(\frac{S_n}{n}\right) \right] &= \sum_{k=0}^n f\left(\frac{k}{n}\right) P(S_n = k) \\ &= \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \end{aligned}$$

Por la LDGN existe un entero  $n_0$  que no depende de  $p$ , tal que, si  $n \geq n_0$

$$P\left(\left|\frac{S_n}{n} - x\right| > \delta\right) < \varepsilon.$$

Usando la desigualdad triangular obtenemos

$$\begin{aligned} \left| \mathbb{E} \left[ f\left(\frac{S_n}{n}\right) \right] - f(x) \right| &= \left| \sum_{k=0}^n \left( f\left(\frac{k}{n}\right) - f(x) \right) P(S_n = k) \right| \\ &\leq \sum_{k=0}^n \left| f\left(\frac{k}{n}\right) - f(x) \right| P(S_n = k) \\ &\leq \sum_{\left|\frac{k}{n} - x\right| \leq \delta} \left| f\left(\frac{k}{n}\right) - f(x) \right| P(S_n = k) + \sum_{\left|\frac{k}{n} - x\right| > \delta} \left( \left| f\left(\frac{k}{n}\right) \right| + |f(x)| \right) P(S_n = k) \\ &\leq \sum_{\left|\frac{k}{n} - x\right| \leq \delta} \varepsilon P(S_n = k) + \sum_{\left|\frac{k}{n} - x\right| > \delta} 2 \sup_{0 \leq x \leq 1} |f(x)| P(S_n = k) \\ &= \varepsilon P\left(\left|\frac{S_n}{n} - x\right| \leq \delta\right) + 2 \sup_{0 \leq x \leq 1} |f(x)| P\left(\left|\frac{S_n}{n} - x\right| > \delta\right). \end{aligned}$$

En consecuencia, para todo  $n \geq n_0$ ,

$$\left| \mathbb{E} \left[ f\left(\frac{S_n}{n}\right) \right] - f(x) \right| \leq \varepsilon + 2\varepsilon \sup_{0 \leq x \leq 1} |f(x)|$$

y esto demuestre la proposición. ■

### 2.3. Desigualdad de Chernoff

La desigualdad de Chebychef nos da una cota general para la probabilidad  $P(|X - \mathbb{E}(X)| \geq \varepsilon)$  que en muchos casos, como hemos mencionado, puede ser mejorada y en esta sección veremos como hacer esto para variables con distribución binomial. Sin embargo, es importante resaltar que si queremos una desigualdad que valga para cualquier v.a. con segundo momento finito, no es posible mejorar la desigualdad de Chebychef.

Sea  $S_n = \sum_1^n X_i$  donde las variables  $X_i$  son independientes y tienen distribución de Bernoulli con probabilidad de éxito  $p$ .

**Lema 2.3** Para cualquier  $u \in \mathbb{R}$ ,

$$\mathbb{E}(e^{uS_n}) = (1 - p + pe^u)^n$$

*Demostración.* Como las variables  $X_i, 1 \leq i \leq n$  son independientes, las variables  $e^{uX_i}, 1 \leq i \leq n$  también lo son. Por otro lado, para cualquier  $i$ ,

$$\mathbb{E}(e^{uX_i}) = pe^u + (1 - p)e^0 = 1 - p + pe^u$$

Usando ahora la independencia

$$\begin{aligned} \mathbb{E}(e^{uS_n}) &= \mathbb{E}\left(\prod_{i=1}^n e^{uX_i}\right) = \prod_{i=1}^n \mathbb{E}(e^{uX_i}) \\ &= (1 - p + pe^u)^n \end{aligned}$$

■

Si  $X$  es una variable aleatoria, la función  $\psi(u) = \mathbb{E}(e^{uX})$  (para los valores de  $u$  tales que el valor esperado está bien definido) se conoce como la *función generadora de momentos* de  $X$ . En el lema anterior calculamos esta función para variables con distribución binomial.

**Teorema 2.3 (Desigualdad de Chernoff)** Sea  $S_n$  una variable con distribución binomial de parámetros  $n$  y  $p$  y sea  $\lambda = np = \mathbb{E}(S_n)$ . Para  $0 < \lambda + t < n$

$$P(S_n \geq \lambda + t) \leq \left(\frac{\lambda}{\lambda + t}\right)^{\lambda + t} \left(\frac{n - \lambda}{n - \lambda - t}\right)^{n - \lambda - t} \quad (2.20)$$

*Demostración.* Sea  $u > 0$ . Como la función  $f(x) = e^{ux}$  es una función creciente tenemos

$$P(S_n \geq \lambda + t) = P(e^{uS_n} \geq e^{u(\lambda + t)}) \leq e^{-u(\lambda + t)} \mathbb{E}(e^{uS_n})$$

donde la desigualdad se obtiene aplicando la desigualdad de Markov. Usando el resultado del lema 2.3 obtenemos

$$P(S_n \geq \lambda + t) \leq (1 - p + pe^u)^n e^{-u(\lambda + t)} \quad (2.21)$$

Pongamos  $g(u) = (1 - p + pe^u)^n e^{-u(\lambda + t)}$  y hallemos el valor de  $u$  que minimiza esta función. Derivando obtenemos

$$\begin{aligned} g'(u) &= -e^{-u(\lambda + t)}(\lambda + t)(1 - p + pe^u)^n + e^{-u(\lambda + t)}n(1 - p + pe^u)^{n-1}pe^u \\ &= e^{-u(\lambda + t)}(1 - p + pe^u)^{n-1} [pne^u - (\lambda + t)(1 - p + pe^u)] \\ &= e^{-u(\lambda + t)}(1 - p + pe^u)^{n-1} [ -(\lambda + t)(1 - p) + pe^u(n - \lambda - t) ] \end{aligned}$$

Para que esta expresión se anule es necesario que el tercer factor valga cero, pues los otros dos son positivos. Obtenemos

$$e^u = \frac{(\lambda + t)(1 - p)}{p(n - \lambda - t)}. \quad (2.22)$$

Para ver que este valor es un mínimo para  $g$  observamos que  $g(0) = 1$ ,  $g(u) \rightarrow \infty$  ( $u \rightarrow \infty$ ) porque  $\lambda + t < n$  y es fácil verificar que  $g'(0) < 0$ . Como la función es continua debe tener un mínimo en el valor de  $u$  definido por (2.22).

Regresamos a la expresión (2.21) y reemplazamos  $e^u$  por (2.22)

$$\begin{aligned} P(S_n \geq \lambda + t) &\leq \left(1 - p + \frac{(\lambda + t)(1 - p)}{(n - \lambda - t)}\right)^n \left(\frac{p(n - \lambda - t)}{(\lambda + t)(1 - p)}\right)^{(\lambda + t)} \\ &= \left(\frac{n(1 - p)}{n - \lambda - t}\right)^n \left(\frac{np(n - \lambda - t)}{n(\lambda + t)(1 - p)}\right)^{(\lambda + t)} \end{aligned}$$

recordando que  $\lambda = np$  y por lo tanto  $n(1-p) = n - \lambda$  obtenemos

$$= \left( \frac{n - \lambda}{n - \lambda - t} \right)^n \left( \frac{n - \lambda - t}{n - \lambda} \right)^{(\lambda+t)} \left( \frac{\lambda}{\lambda + t} \right)^{\lambda+t}$$

con lo que queda demostrada la desigualdad. ■

Para  $0 < x < 1$  definimos

$$\varphi(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}. \quad (2.23)$$

**Corolario 2.1** Para  $n \geq 1$  y  $\varepsilon \in (0, 1-p)$  se tiene

$$P\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq e^{-n\varphi(p+\varepsilon)} \quad (2.24)$$

*Demostración.* Recordando que  $\lambda = np$  y usando (2.20) obtenemos

$$P\left(\frac{S_n}{n} \geq p + \frac{t}{n}\right) \leq \left(\frac{p}{p + \frac{t}{n}}\right)^{np+t} \left(\frac{1-p}{1-p - \frac{t}{n}}\right)^{n-np-t}$$

siempre que  $0 < \lambda + t < n$ . Poniendo  $\varepsilon = t/n$  esta condición es  $0 < p + \varepsilon < 1$  y obtenemos

$$\begin{aligned} P\left(\frac{S_n}{n} \geq p + \varepsilon\right) &\leq \exp\left\{n(p+\varepsilon) \log \frac{p}{p+\varepsilon} + n(1-p-\varepsilon) \log \frac{1-p}{1-p-\varepsilon}\right\} \\ &= \exp\left\{-n\left[(p+\varepsilon) \log \frac{p+\varepsilon}{p} + (1-p-\varepsilon) \log \frac{1-p-\varepsilon}{1-p}\right]\right\} \\ &= \exp\{-n\varphi(p+\varepsilon)\} \end{aligned}$$

■

**Corolario 2.2** Si  $0 < \varepsilon < p$  entonces  $\varphi(p-\varepsilon)$  está bien definida, es positiva y

$$P\left(\frac{S_n}{n} \leq p - \varepsilon\right) \leq e^{-n\varphi(p-\varepsilon)} \quad (2.25)$$

*Demostración.* Este resultado sigue de intercambiar éxitos y fracasos,  $S_n$  y  $n - S_n$  y  $p$  y  $1-p$ . ■

**Corolario 2.3** Si  $0 < \varepsilon < \min(p, 1-p)$  entonces

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq e^{-n\varphi(p+\varepsilon)} + e^{-n\varphi(p-\varepsilon)} \quad (2.26)$$

*Demostración.* Usar los dos corolarios anteriores. ■

### 2.3.1. Comparación

La desigualdad de Chebychef en el caso binomial nos dice que

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Observamos que para  $n$  y  $p$  fijos, esta cota tiende a  $\infty$  cuando  $\varepsilon \rightarrow 0$ , de modo que la desigualdad deja de ser útil para valores muy pequeños de  $\varepsilon$ . Veamos algunos ejemplos para comparar los resultados que se obtienen con esta desigualdad y la cota de Chernoff.

**Ejemplos 2.1**

Vamos a considerar el caso de una distribución binomial con parámetro  $p = 0.5$ , que corresponde al caso simétrico.

1. Para  $n = 100$  y  $\varepsilon = 0.1$  la desigualdad de Chebychef nos da

$$P\left(\left|\frac{1}{100}S_{100} - \frac{1}{2}\right| > 0.1\right) \leq \frac{1}{4} = 0.25$$

mientras que con la desigualdad de Chernoff obtenemos  $h_+(0.1) = h_-(0.1) = 0.02014$  y

$$e^{-100h_+(0.1)} + e^{-100h_-(0.1)} = 0.267,$$

de modo que la desigualdad de Chernoff no mejora la de Chebychef en este caso. Si cambiamos el valor de  $\varepsilon$  a 0.05 obtenemos que la cota de Chebychef es 1, que no nos dice nada, mientras que la cota de Chernoff es 0.005

2. Para  $n = 1000$  y  $\varepsilon = 0.1$ , la desigualdad de Chebychef da una cota de 0.025 mientras que Chernoff da una cota menor que  $3.6 \times 10^{-9}$ . Si cambiamos a  $\varepsilon = 0.05$  con Chebychef obtenemos 0.1 mientras que para Chernoff la cota es menor que 0.0067.
3. Finalmente para  $n = 10,000$  y  $\varepsilon = 0.1$  obtenemos 0.0025 y  $7.14 \times 10^{-88}$  mientras que para  $\varepsilon = 0.05$  las cotas son 0.01 y  $1.8 \times 10^{-22}$ .

**Notación para el Comportamiento Asintótico de Funciones**

Introducimos la notación usual para comparación de sucesiones, propuesta por E. Landau. Sean  $f$  y  $g$ , dos funciones de números reales.

- Decimos que  $f = \mathcal{O}(g)$  o con mayor precisión que  $f(x) = \mathcal{O}(g(x))$  cuando  $x \rightarrow \infty$ , si existen una constante  $k > 0$  y  $x_0 \in \mathbb{R}$  tales que

$$|f(x)| \leq kg(x) \quad \text{para todo } x \geq x_0.$$

En este caso decimos que  $f$  es, a lo sumo, del mismo orden que  $g$  cuando  $x \rightarrow \infty$ .

- Decimos que  $f(x) = \mathcal{o}(g(x))$  cuando  $x \rightarrow \infty$  si para todo  $\varepsilon > 0$  existe  $x_0 \in \mathbb{R}$  tal que

$$|f(x)| \leq \varepsilon g(x) \quad \text{para todo } n \geq x_0$$

o equivalentemente, si

$$\frac{|f(x)|}{g(x)} \rightarrow 0 \quad \text{cuando } x \rightarrow \infty.$$

En particular, si  $f(x) \rightarrow 0$  cuando  $x \rightarrow \infty$  escribimos  $f(x) = \mathcal{o}(1)$  cuando  $x \rightarrow \infty$ .

- Escribimos  $f \sim g$  o  $f(x) \sim g(x)$  cuando  $x \rightarrow \infty$  si  $g(x) \neq 0$  para todo  $x$  grande y

$$\frac{f(x)}{g(x)} \rightarrow 1 \quad \text{cuando } x \rightarrow \infty.$$

Notaciones similares se usan cuando  $x \rightarrow 0$  o  $x \rightarrow -\infty$ , con las modificaciones obvias y también para sucesiones.

**Ejemplo 2.2**

La función  $f(x) = 2x^3 - x^2 + 5$  satisface las relaciones:  $f(x) = \mathcal{O}(x^3)$ ,  $f(x) = \mathcal{o}(x^4)$  y  $f(x) \sim 2x^3$  cuando  $x \rightarrow \infty$ .

Para el próximo resultado vamos a utilizar el desarrollo de la función  $\log(1+x)$  para  $|x| < 1$ , alrededor de  $x = 0$ , hasta el segundo grado:

$$\log(1+x) = x - \frac{x^2}{2} + \mathcal{O}(x^3) \quad (2.27)$$

cuando  $x \rightarrow 0$ , que se obtiene a partir de un desarrollo de Taylor.

**Proposición 2.2** *Cuando  $\varepsilon \rightarrow 0$  se tiene que*

$$\varphi(p + \varepsilon) = \frac{\varepsilon^2}{2p(1-p)} + \mathcal{O}(\varepsilon^3). \quad (2.28)$$

*Demostración.* Usando la ecuación (2.27) obtenemos

$$\begin{aligned} \varphi(p + \varepsilon) &= (p + \varepsilon) \log\left(1 + \frac{\varepsilon}{p}\right) + (1 - p - \varepsilon) \log\left(1 - \frac{\varepsilon}{1-p}\right) \\ &= (p + \varepsilon) \left( \frac{\varepsilon}{p} - \frac{1}{2} \left( \frac{\varepsilon}{p} \right)^2 + \mathcal{O}(\varepsilon^3) \right) + (1 - p - \varepsilon) \left( \frac{-\varepsilon}{1-p} - \frac{1}{2} \left( \frac{-\varepsilon}{1-p} \right)^2 + \mathcal{O}(\varepsilon^3) \right) \\ &= \frac{\varepsilon^2}{2p} + \frac{\varepsilon^2}{2(1-p)} + \mathcal{O}(\varepsilon^3) \\ &= \frac{\varepsilon^2}{2p(1-p)} + \mathcal{O}(\varepsilon^3). \end{aligned}$$

■

## 2.4. Desigualdad de Chernoff (2)

En esta sección veremos que la desigualdad de Chernoff (Corolario 2.1) es asintóticamente el mejor resultado posible. Para la demostración requerimos la fórmula de Stirling, que demostraremos en el siguiente capítulo.

**Lema 2.4 (Fórmula de Stirling)** *Cuando  $n \rightarrow \infty$ , se tiene*

$$n! \sim \sqrt{2\pi n} n^n e^{-n}.$$

**Proposición 2.3** *Para todo  $\varepsilon \in (0, 1-p)$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( P\left(\frac{S_n}{n} \geq p + \varepsilon\right) \right) = -\varphi(p + \varepsilon) \quad (2.29)$$

*Demostración.* El resultado del Corolario 2.1 dice que para  $n \geq 1$

$$\frac{1}{n} \log P(S_n \geq n(p + \varepsilon)) \leq -\varphi(p + \varepsilon)$$

Para hallar una cota inferior ponemos  $\alpha_n = 1 + [n(p + \varepsilon)]$  que representa al menor entero mayor que  $n(p + \varepsilon)$ . Entonces

$$P(S_n \geq n(p + \varepsilon)) \geq P(S_n = \alpha_n) = \frac{n!}{\alpha_n!(n - \alpha_n)!} p^{\alpha_n} (1-p)^{n-\alpha_n}$$

Usando la fórmula de Stirling para los factoriales que aparecen en la expresión anterior

$$P(S_n = \alpha_n) \sim \frac{1}{\sqrt{2\pi}} \left( \frac{n}{\alpha_n(n - \alpha_n)} \right)^{1/2} \left( \frac{np}{\alpha_n} \right)^{\alpha_n} \left( \frac{n(1-p)}{n - \alpha_n} \right)^{n-\alpha_n} \quad (2.30)$$

y por lo tanto los logaritmos de ambos lados de la relación también son asintóticamente equivalentes. A partir de las relaciones  $\alpha_n \sim n(p + \varepsilon)$  y  $n - \alpha_n \sim n(1 - p - \varepsilon)$  obtenemos

$$\left(\frac{n}{\alpha_n(n - \alpha_n)}\right)^{1/2} \sim \left(\frac{n}{n^2(p + \varepsilon)(1 - p - \varepsilon)}\right)^{1/2} = (n(p + \varepsilon)(1 - p - \varepsilon))^{-1/2}$$

y en consecuencia

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \left( \frac{n}{2\pi\alpha_n(n - \alpha_n)} \right)^{1/2} \right) = 0. \quad (2.31)$$

Por otro lado

$$\begin{aligned} \alpha_n \log \left( \frac{np}{\alpha_n} \right) &= n(p + \varepsilon) \log \left( \frac{np(p + \varepsilon)}{\alpha_n(p + \varepsilon)} \right) + (\alpha_n - n(p + \varepsilon)) \log \left( \frac{np}{\alpha_n} \right) \\ &= n(p + \varepsilon) \log \left( \frac{p}{p + \varepsilon} \right) + n(p + \varepsilon) \log \left( \frac{n(p + \varepsilon)}{\alpha_n} \right) + (\alpha_n - n(p + \varepsilon)) \log \left( \frac{np}{\alpha_n} \right) \end{aligned} \quad (2.32)$$

Usando de nuevo que  $\alpha_n \sim n(p + \varepsilon)$  obtenemos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left( n(p + \varepsilon) \log \left( \frac{n(p + \varepsilon)}{\alpha_n} \right) \right) = 0. \quad (2.33)$$

Por otro lado la diferencia  $\alpha_n - n(p + \varepsilon)$  está acotada y en consecuencia

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left( (\alpha_n - n(p + \varepsilon)) \log \left( \frac{np}{\alpha_n} \right) \right) = 0. \quad (2.34)$$

Usando (2.33) y (2.34) en (2.32) obtenemos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{np}{\alpha_n} \right)^{\alpha_n} = (p + \varepsilon) \log \left( \frac{p}{p + \varepsilon} \right).$$

De manera similar se demuestra que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{n(1 - p)}{n - \alpha_n} \right)^{n - \alpha_n} = (1 - p - \varepsilon) \log \left( \frac{1 - p}{1 - p - \varepsilon} \right)$$

y esto concluye la demostración. ■

## 2.5. La Fórmula de Stirling.

En la próxima sección estudiaremos la aproximación de la distribución binomial mediante la distribución normal. Para ello, utilizaremos la siguiente proposición sobre sucesiones de números reales.

**Lema 2.5 (Fórmula de Stirling)** *Cuando  $n \rightarrow \infty$ , se tiene*

$$n! \sim \sqrt{2\pi n} n^n e^{-n}.$$

*Demostración* Se trata de probar que

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{n} n^n e^{-n}} = \sqrt{2\pi}.$$

Dividimos la demostración en dos pasos. En el primero veremos que existe y es finito el límite indicado en el primer miembro; en el segundo veremos que dicho límite vale  $\sqrt{2\pi}$ .

**Paso 1.** Consideremos la función  $f(x) = \log x$  para  $x \geq 1$  y acotemos inferiormente el área  $A_n$  limitada por su gráfica, el eje  $0x$  y la vertical por  $x = n$ , sustituyendo cada trozo de la curva  $P_k P_{k+1}$  por el segmento  $\overline{P_k P_{k+1}}$  (ver Figura 2.6). Teniendo en cuenta que el área del trapecio  $k P_k P_{k+1} (k+1)$  es

$$\frac{1}{2}(\log k + \log(k+1))$$

obtenemos

$$\begin{aligned} A_n &= \int_1^n \log x \, dx > \sum_{k=1}^{n-1} \frac{1}{2}(\log k + \log(k+1)) \\ &= \log 2 + \log 3 + \cdots + \log(n-1) + \frac{1}{2} \log n. \end{aligned} \quad (2.35)$$

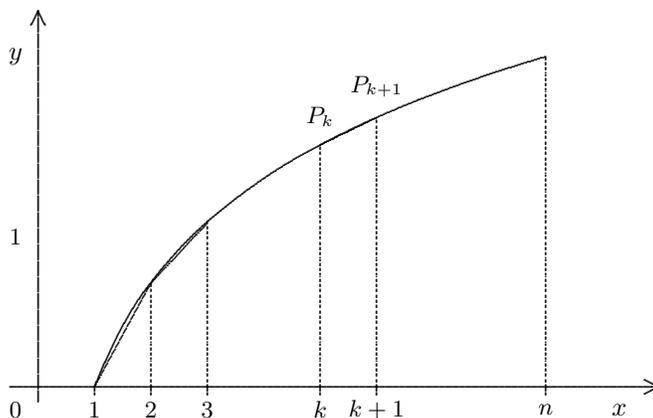


Figura 2.6 Aproximación a la función  $\log x$ .

Sea  $a_n = A_n - (\log 2 + \log 3 + \cdots + \log(n-1) + \frac{1}{2} \log n)$ , la diferencia entre el área  $A_n$  y la suma de las áreas de los trapecios. Es claro que la sucesión  $\{a_n\}$  es monótona creciente.

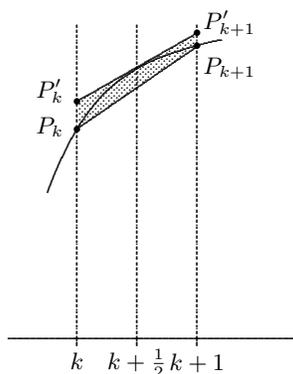


Figura 2.7 Acotación de la aproximación a la función  $\log x$ .

Además, si llamamos  $b_k$  al área sombreada en la figura 2.7, donde el segmento  $\overline{P'_k P'_{k+1}}$  es tangente al gráfico de  $f(x) = \log x$  en el punto de abscisa  $x = k + 1/2$ , es claro que

$$a_n < \sum_{k=1}^{n-1} b_k, \quad (n > 1), \quad (2.36)$$

donde  $b_k$  es la diferencia de las áreas de los trapecios  $k P'_k P'_{k+1} (k+1)$  y  $k P_k P_{k+1} (k+1)$ . Por lo tanto

$$\begin{aligned} b_k &= \log(k + \frac{1}{2}) - \frac{1}{2}[\log k + \log(k+1)] = \frac{1}{2} \log \frac{(k + \frac{1}{2})^2}{k(k+1)} \\ &= \frac{1}{2} \log(1 + \frac{1/4}{k(k+1)}) \leq \frac{1}{8k(k+1)} \end{aligned}$$

ya que  $\log(1+x) \leq x$ . Sustituyendo en (2.36)

$$a_n < \frac{1}{8} \sum_{k=0}^{n-1} \frac{1}{k(k+1)} < \frac{1}{8} \sum_{k=1}^{\infty} \frac{1}{k^2} = C < \infty$$

porque la serie  $\sum_k 1/k^2$  es convergente.

En consecuencia, la sucesión  $\{a_n\}$  es monótona creciente y acotada superiormente. Por lo tanto tiene un límite finito que llamaremos  $a$ . Por otra parte, integrando por partes

$$A_n = \int_1^n \log x \, dx = x \log x \Big|_1^n - \int_1^n dx = n \log n - n + 1$$

y en consecuencia

$$a_n = n \log n - n + 1 - (\log 2 + \dots + \log(n-1) + \frac{1}{2} \log n).$$

Tomando exponenciales en ambos miembros

$$e^{a_n-1} = \frac{n^n e^{-n} n^{1/2}}{n!},$$

de donde

$$\frac{n!}{n^n e^{-n} n^{1/2}} \rightarrow \frac{1}{e^{a-1}} = \alpha \text{ cuando } n \rightarrow \infty.$$

**Paso 2.** Probemos ahora que  $\alpha = \sqrt{2\pi}$ . Para ello consideramos

$$\begin{aligned} I_n &= \int_0^{\pi/2} \text{sen}^n x \, dx = \int_0^{\pi/2} \text{sen}^{n-2} x (1 - \cos^2 x) \, dx \\ &= I_{n-2} - \int_0^{\pi/2} \text{sen}^{n-2} x \cos x \cos x \, dx \\ &= I_{n-2} - \left[ \frac{1}{n-1} \text{sen}^{n-1} x \cos x \Big|_0^{\pi/2} + \int_0^{\pi/2} \frac{1}{n-1} \text{sen}^{n-1} x \text{sen} x \, dx \right] \\ &= I_{n-2} - \frac{1}{n-1} I_n \end{aligned}$$

de donde

$$I_n = \frac{n-1}{n} I_{n-2}. \tag{2.37}$$

Por lo tanto, considerando separadamente los casos en los cuales  $n$  es par e impar, y aplicando (2.37) reiteradamente, obtenemos

$$I_{2p} = \frac{2p-1}{2p} \frac{2p-3}{2p-2} \dots \frac{3}{4} \frac{1}{2} I_0; \quad I_{2p+1} = \frac{2p}{2p+1} \frac{2p-2}{2p-1} \dots \frac{4}{5} \frac{2}{3} I_1.$$

Además

$$I_0 = \int_0^{\pi/2} dx = \frac{\pi}{2}, \quad I_1 = \int_0^{\pi/2} \operatorname{sen} x \, dx = 1,$$

o sea que

$$I_{2p} = \frac{2p-1}{2p} \frac{2p-3}{2p-2} \cdots \frac{3}{4} \frac{1}{2} \frac{\pi}{2}; \quad I_{2p} = \frac{2p}{2p+1} \frac{2p-2}{2p-1} \cdots \frac{4}{5} \frac{2}{3}. \quad (2.38)$$

Observamos ahora que  $\{I_n\}$  es una sucesión decreciente, ya que para  $0 \leq x \leq \pi/2$  se tiene  $0 \leq \operatorname{sen} x \leq 1$ ,  $\operatorname{sen}^n x$  decrece con  $n$ , y por lo tanto, también la integral  $I_n$ . De aquí resulta que:

$$\frac{I_{2p+2}}{I_{2p}} < \frac{I_{2p+1}}{I_{2p}} < 1.$$

Usando (2.37) tenemos

$$\frac{I_{2p+2}}{I_{2p}} = \frac{2p+1}{2p+2} \rightarrow 1 \quad \text{cuando } p \rightarrow \infty$$

y por lo tanto la sucesión intermedia

$$\frac{I_{2p+1}}{I_{2p}}$$

también tiende a 1 cuando  $p \rightarrow \infty$ . Usando ahora (2.38) esto indica que

$$\frac{I_{2p+1}}{I_{2p}} = \frac{[2p(2p-2) \cdots 4 \times 2]^2}{(2p+1)[(2p-1)(2p-3) \cdots 5 \times 3]^2} \frac{2}{\pi} \rightarrow 1 \quad \text{cuando } p \rightarrow \infty,$$

y en consecuencia

$$\frac{2p(2p-2) \cdots 4 \times 2}{(2p-1)(2p-3) \cdots 5 \times 3} \sqrt{\frac{2}{\pi(2p+1)}} \rightarrow 1 \quad \text{cuando } p \rightarrow \infty.$$

Multiplicando numerador y denominador por

$$2p(2p-2) \cdots 4 \times 2 = 2^p p(p-1)(p-2) \cdots 1 = 2^p p!$$

obtenemos

$$\frac{(2^p p!)^2}{(2p)!} \sqrt{\frac{2}{\pi(2p+1)}} \rightarrow 1 \quad \text{cuando } p \rightarrow \infty.$$

Utilizamos ahora el resultado del paso 1. Sabemos que

$$n! \sim \alpha n^n e^{-n} \sqrt{n}.$$

Sustituyendo

$$\frac{(2^p p!)^2}{(2p)!} \sqrt{\frac{2}{\pi(2p+1)}} \sim \frac{(2^p \alpha p^p e^{-p} \sqrt{p})^2}{\alpha (2p)^{2p} e^{-2p} \sqrt{2p}} \sqrt{\frac{2}{\pi(2p+1)}} \sim \frac{\alpha}{\sqrt{2\pi}}.$$

Como el límite es 1, debe ser  $\alpha = \sqrt{2\pi}$ . Esto termina la demostración de la fórmula de Stirling. ■

## 2.6. El Teorema de De-Moivre - Laplace.

Podemos ahora probar el teorema de De-Moivre - Laplace, que permite aproximar la distribución binomial por la distribución normal. Recordemos que la distribución normal típica – con parámetros  $(0,1)$  – es aquella cuya densidad es la función

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (2.39)$$

Denotamos la función de distribución respectiva por

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt. \quad (2.40)$$

Como anteriormente, llamaremos  $S_n$  a una variable aleatoria con distribución binomial, que representa el número de veces que ocurre el evento  $A$  en  $n$  observaciones independientes. Sea  $p = P(A)$ ,  $0 < p < 1$  y  $q = 1 - p$ . La distribución de  $S_n$  está dada por

$$p_{n,k} = P(S_n = k) = \binom{n}{k} p^k q^{n-k} \quad k = 0, 1, \dots, n,$$

y tiene  $E(S_n) = np$ ,  $\text{Var}(S_n) = npq$ .

En los próximos dos teoremas probaremos que cuando el número de observaciones crece indefinidamente, la distribución de la variable aleatoria

$$S'_n = \frac{S_n - np}{\sqrt{npq}}$$

tiende a la distribución normal dada por (2.40).

**Teorema 2.4** Sean  $a < b$  números reales, entonces

$$P(a < S'_n \leq b) \longrightarrow \Phi(b) - \Phi(a) = \int_a^b \phi(t) dt, \quad \text{cuando } n \rightarrow \infty. \quad (2.41)$$

*Demostración* Tenemos que estudiar el límite cuando  $n \rightarrow \infty$  de

$$\begin{aligned} P(a < S'_n \leq b) &= P\left(a < \frac{S_n - np}{\sqrt{npq}} \leq b\right) \\ &= \sum_{a < \frac{k - np}{\sqrt{npq}} \leq b} P(S_n = k) \\ &= \sum_{a < \frac{k - np}{\sqrt{npq}} \leq b} p_{n,k} \end{aligned}$$

donde la suma se extiende sobre los enteros  $k$  tales que

$$a < \frac{k - np}{\sqrt{npq}} \leq b. \quad (2.42)$$

Comencemos por dar una aproximación para cada sumando  $p_{n,k}$ , donde  $k$  verifica (2.42). Para facilitar la notación ponemos  $\delta = k - np$ . Si  $k$  verifica (2.42), entonces

$$\frac{\delta}{n} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty \quad (2.43)$$

ya que entonces,  $\frac{|\delta|}{\sqrt{n}} = \frac{|k - np|}{\sqrt{n}}$  es una sucesión acotada, porque

$$a \sqrt{pq} < \frac{k - np}{\sqrt{n}} \leq b \sqrt{pq}$$

y como  $1/\sqrt{n} \rightarrow 0$  tenemos

$$\frac{\delta}{n} = \frac{\delta}{\sqrt{n}} \frac{1}{\sqrt{n}} \rightarrow 0.$$

Por otro lado

$$p_{n,k} = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}.$$

Utilizamos ahora la fórmula de Stirling con la siguiente notación: pondremos

$$n! = \sqrt{2\pi n} n^n e^{-n} e^{\gamma_n}$$

donde  $\gamma_n \rightarrow 0$  (y por lo tanto  $e^{\gamma_n} \rightarrow 1$ ) cuando  $n \rightarrow \infty$ .

Teniendo en cuenta que bajo la condición (2.43) –que se verifica puesto que sólo nos interesan los valores de  $k$  que cumplen (2.42)– tanto  $k$  como  $n-k$  tienden a  $+\infty$  cuando  $n \rightarrow \infty$ , tenemos

$$\begin{aligned} p_{n,k} &= \frac{\sqrt{2\pi n} n^n e^{-n} p^k q^{n-k} e^{\gamma_n - \gamma_k - \gamma_{n-k}}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi(n-k)} (n-k)^{(n-k)} e^{-(n-k)}} \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{(np+\delta)(nq-\delta)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} e^{\gamma_n - \gamma_k - \gamma_{n-k}}. \end{aligned} \quad (2.44)$$

El primer factor lo podemos escribir como

$$\sqrt{\frac{n}{(np+\delta)(nq-\delta)}} = \frac{1}{\sqrt{n(p+\frac{\delta}{n})(q-\frac{\delta}{n})}} = \frac{1}{\sqrt{npq}} e^{\gamma'_n}$$

con  $\gamma'_n \rightarrow 0$  cuando  $n \rightarrow \infty$ .

Tomando logaritmos en el segundo y tercer factor tenemos

$$k \log\left(\frac{np}{k}\right) + (n-k) \log\left(\frac{nq}{n-k}\right) = -(np+\delta) \log\left(1 + \frac{\delta}{np}\right) - (nq-\delta) \log\left(1 - \frac{\delta}{nq}\right).$$

Usamos el desarrollo de MacLaurin de la función  $\log(1+x)$  cuando  $|x| < 1$ :

$$\log(1+x) = x - \frac{x^2}{2} + \frac{1}{3} \frac{1}{(1+\theta x)^3} x^3, \quad (0 < \theta < 1).$$

Si  $|x| < 1/2$  y  $A = \frac{1}{3} \frac{1}{(1+\theta x)^3}$  se tiene  $|A| < 3$ , es decir, que en esta situación vale

$$\log(1+x) = x - \frac{x^2}{2} + Ax^3 \quad \text{con } |A| < 3.$$

Aplicamos este desarrollo a  $\log(1 + \frac{\delta}{np})$  y  $\log(1 - \frac{\delta}{nq})$ , lo cual es posible si  $n$  es suficientemente grande, ya que  $\frac{\delta}{n} \rightarrow 0$ , y en consecuencia, tanto  $\frac{\delta}{np}$  como  $\frac{\delta}{nq}$  son menores que  $1/2$  a partir de un cierto valor de  $n$  en adelante. Resulta

$$\begin{aligned} &k \log\left(\frac{np}{k}\right) + (n-k) \log\left(\frac{nq}{n-k}\right) \\ &= -(np+\delta) \left(\frac{\delta}{np} - \frac{\delta^2}{2n^2 p^2} + A \frac{\delta^3}{n^3 p^3}\right) - (nq-\delta) \left(-\frac{\delta}{nq} - \frac{\delta^2}{2n^2 q^2} + A_1 \frac{\delta^3}{n^3 q^3}\right) \\ &= -\frac{\delta^2}{2n} \left(\frac{1}{p} + \frac{1}{q}\right) + B \frac{\delta^3}{n^2}, \end{aligned}$$

donde  $|A| < 3$ ,  $|A_1| < 3$  y  $|B|$  está acotado por una cierta constante fija, digamos  $|B| \leq M$ . Tomando el antilogaritmo del segundo factor y sustituyendo en (2.44)

$$p_{n,k} = \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{\delta^2}{2npq} + B\frac{\delta^3}{n^2} + \theta_{n,k}\right\} \quad (2.45)$$

$$\begin{aligned} &= \frac{1}{\sqrt{npq}} \phi(x_{n,k}) \exp\left\{B\frac{\delta^3}{n^2} + \theta_{n,k}\right\} \\ &= \frac{1}{\sqrt{npq}} \phi(x_{n,k}) e^{\alpha_{n,k}} \end{aligned} \quad (2.46)$$

con

$$\theta_{n,k} = \gamma_n - \gamma_k - \gamma_{n-k} - \gamma'_n, \quad \alpha_{n,k} = B\frac{\delta^3}{n^2} + \theta_{n,k} \quad \text{y} \quad x_{n,k} = \frac{\delta}{\sqrt{npq}} = \frac{k - np}{\sqrt{npq}},$$

donde  $\frac{\delta^3}{n^2} \rightarrow 0$  cuando  $n \rightarrow \infty$  ya que

$$\frac{\delta^3}{n^2} = \left(\frac{\delta}{\sqrt{n}}\right)^3 \frac{1}{\sqrt{n}}$$

y el primer factor permanece acotado cuando  $k$  verifica (2.42).

Volvamos ahora a nuestra suma inicial. Sustituyendo en ella la expresión hallada en (2.46) para  $p_{n,k}$ , obtenemos

$$\begin{aligned} P(a < S'_n \leq b) &= \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}} \phi(x_{k,n}) e^{\alpha_{n,k}} \\ &= \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}} \phi(x_{k,n}) + \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}} \phi(x_{k,n}) [e^{\alpha_{n,k}} - 1]. \end{aligned} \quad (2.47)$$

Dado que  $\frac{1}{\sqrt{npq}}$  es justamente el incremento  $x_{k+1,n} - x_{k,n}$ , y que

$$\frac{1}{\sqrt{npq}} \rightarrow 0$$

el primer sumando tiende a

$$\int_a^b \phi(t) dt \quad (2.48)$$

cuando  $n \rightarrow \infty$ .

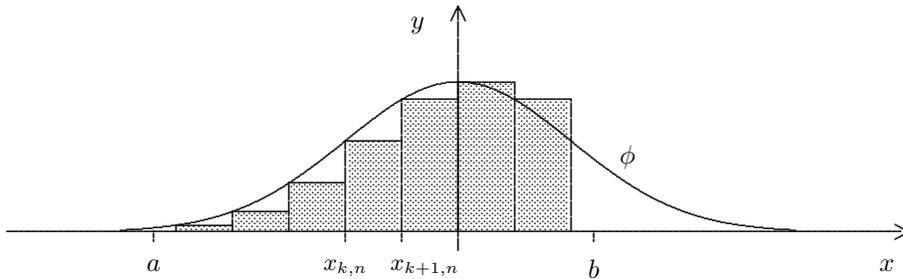


Figura 2.8 Aproximación a la densidad Gaussiana.

En cuanto al segundo sumando de (2.47), dado que

$$|e^x - 1| \leq e^{|x|} - 1,$$

el mismo se acota por

$$\sup_k |\alpha_{n,k}| \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}} \phi(x_{k,n}),$$

y ahora el primer factor tiende a cero cuando  $n \rightarrow \infty$  mientras que el segundo tiende a la integral (2.48) Por lo tanto el segundo sumando de (2.47) tiende a cero. Esto termina la demostración del teorema 2.4. ■

### Teorema 2.5

$$P(S'_n \leq x) \rightarrow \Phi(x) = \int_{-\infty}^x \phi(t) dt \quad \text{cuando } n \rightarrow \infty$$

*Demostración* Nos apoyamos en el teorema 7.3. Sea  $\varepsilon > 0$ , elegimos  $a$  y  $b$  tales que  $a < x < b$  (ver figura 7.6) y

$$\Phi(a) < \frac{\varepsilon}{2}, \quad 1 - \Phi(b) < \frac{\varepsilon}{2}. \quad (2.49)$$

Entonces

$$P(a < S'_n \leq x) \leq P(S'_n \leq x) = 1 - P(S'_n > x) \leq 1 - P(b \geq S'_n > x).$$

El primer miembro tiende a

$$\Phi(x) - \Phi(a)$$

y el último a

$$1 - (\Phi(b) - \Phi(x)) = \Phi(x) + (1 - \Phi(b))$$

cuando  $n \rightarrow \infty$ .

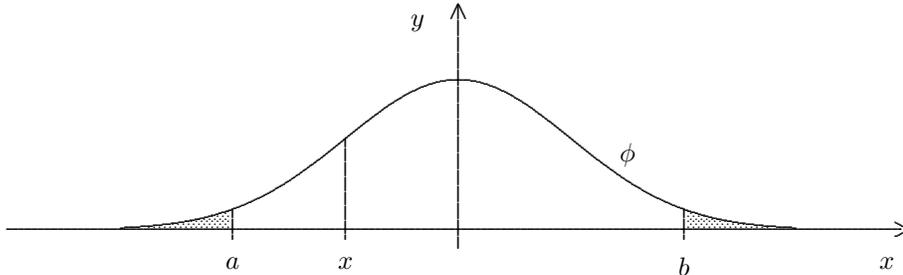


Figura 2.9 Selección de los números  $a$  y  $b$ .

Por lo tanto, existe  $N$  tal que para  $n \geq N$  se tiene

$$\Phi(x) - \Phi(a) - \frac{\varepsilon}{2} \leq P(S'_n \leq x) \leq \Phi(x) + (1 - \Phi(b)) + \frac{\varepsilon}{2}$$

y dada la forma en que han sido elegidos  $a, b$ , para  $n \geq N$

$$\Phi(x) - \varepsilon \leq P(S'_n \leq x) \leq \Phi(x) + \varepsilon$$

o sea

$$|P(S'_n \leq x) - \Phi(x)| \leq \varepsilon.$$

y con esto alcanza para demostrar el resultado ■

**Observación 2.1** 1. Para generalizar en diversos sentidos el teorema de De-Moivre - Laplace, se puede utilizar una técnica de demostración esencialmente análoga. Observando lo que hemos hecho, se ve que el procedimiento depende solamente de que

$$\frac{\delta}{n} \rightarrow 0 \quad (2.50)$$

y

$$\frac{\delta^3}{n^2} \rightarrow 0. \quad (2.51)$$

Ahora bien, es claro que (2.51) implica (2.50) ya que

$$\frac{\delta}{n} = \frac{1}{n^{1/3}} \left( \frac{\delta^3}{n^2} \right)^{1/3}.$$

Por lo tanto, siempre que se verifique (2.51) para todos los enteros  $k$  tales

$$a \leq \frac{k - np}{\sqrt{npq}} \leq b,$$

se podrá obtener la misma conclusión. Esta observación permite generalizar el resultado anterior al caso en el cual los límites  $a$  y  $b$  varían con  $n$ . A título de ejemplo, se puede utilizar la misma demostración de los teoremas 2.4 y 2.5 para probar que si

$$\frac{a_n^3}{\sqrt{n}} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty$$

es decir, que  $a_n$  puede tender a  $+\infty$ , pero sólo más lentamente que  $n^{1/6}$ , entonces

$$P\left(\frac{S_n - np}{\sqrt{npq}} > a_n\right) \sim \int_{a_n}^{+\infty} \phi(t) dt \quad \text{cuando } n \rightarrow \infty \quad (2.52)$$

donde  $\phi$  es, como antes, la densidad normal típica y el símbolo “ $\sim$ ” dice que ambos términos son infinitésimos equivalentes.

Estos resultados permiten estudiar “desviaciones grandes” del número de aciertos  $S_n$  (con distribución binomial), del valor esperado  $np$ . Para un uso eficaz de esta relación, conviene saber cómo tiende a cero, cuando  $n \rightarrow \infty$ , la expresión

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

Veamos en primer lugar que si  $x > 0$  entonces

$$1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi} x} e^{-x^2/2} dt. \quad (2.53)$$

En efecto,

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

ya que en la última integral  $t/x \geq 1$ . Entonces

$$\begin{aligned} 1 - \Phi(x) &\leq \frac{1}{\sqrt{2\pi} x} \int_x^{\infty} t e^{-t^2/2} dt \\ &= \left( -\frac{1}{\sqrt{2\pi} x} e^{-t^2/2} \right) \Big|_x^{+\infty} = \frac{1}{\sqrt{2\pi} x} e^{-x^2/2}. \end{aligned}$$

En segundo lugar mostraremos que

$$1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \quad \text{cuando } x \rightarrow \infty \quad (2.54)$$

Queremos probar ahora que

$$\lim_{x \rightarrow \infty} \frac{1 - \Phi(x)}{\frac{1}{\sqrt{2\pi}x} e^{-x^2/2}} = 1,$$

Aplicamos la regla de L'Hôpital. El cociente de las derivadas es

$$\frac{-\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{1}{\sqrt{2\pi}} \left(-\frac{1}{x^2} e^{-x^2/2} + \frac{1}{x} (-x e^{-x^2/2})\right)} = \frac{1}{\frac{1}{x^2} + 1} \xrightarrow{x \rightarrow \infty} 1,$$

de modo que tiene límite el cociente de las derivadas y por lo tanto también el de las funciones, y vale lo mismo. Esto prueba (2.54).

2. En las aplicaciones, generalmente interesa saber no sólo que cuando el número de observaciones crece indefinidamente, la distribución binomial tiende a la distribución normal, sino además, cuál es la velocidad de convergencia. Dicho de otra manera, interesa conocer una acotación del error cometido cuando, para un  $n$  dado, sustituimos la distribución binomial por la distribución normal. Dicho error depende, naturalmente, del valor de  $n$ , pero además depende del valor de  $p$ ; la convergencia es tanto mas rápida cuanto más cercano es  $p$  a  $1/2$ <sup>1</sup>.

Siguiendo el procedimiento de la demostración del teorema de De-Moivre - Laplace, para dar una acotación del error en consideración, el paso fundamental es el de afinar la fórmula (2.46), que aproxima la función de probabilidad de la distribución binomial, mediante la densidad de la distribución normal. Para ello debe darse, por un lado, una estimación del error en la aplicación de la fórmula de Stirling – que puede obtenerse por el mismo método que hemos seguido para probarla – y por otro, tomar más términos en el desarrollo de  $\log(1+x)$ . Para fijar las ideas, podemos tomar como mejora de la fórmula de Stirling

$$n! = \sqrt{2\pi n} n^n e^{-n} e^{\alpha_n}$$

donde  $0 < \alpha_n < \frac{1}{12n}$ .

En cuanto al desarrollo del logaritmo podemos tomar, por ejemplo,

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{1}{5} \frac{x^5}{(1+\theta x)^5} \quad \text{para } 0 < \theta < 1, |x| < 1.$$

Finalmente, en la fórmula (7.25), también aproximamos

$$\frac{1}{\sqrt{\left(1 + \frac{\delta}{np}\right)\left(1 - \frac{\delta}{nq}\right)}} \left(1 + \frac{\delta}{np}\right)^{-1/2} \left(1 - \frac{\delta}{nq}\right)^{-1/2}$$

mediante un desarrollo de Mac-Laurin.

Sustituyendo en (2.44), si agregamos, por ejemplo, la condición

$$|k - np| = |\delta| \leq C \sqrt{npq} \quad (2.55)$$

donde  $C$  es una constante, y tomamos  $n$  lo bastante grande como para que

$$\frac{C}{\sqrt{npq}} < \frac{1}{3}, \quad (2.56)$$

<sup>1</sup>Cuando  $p$  está cerca de 0 ó de 1 y  $n$  no es muy grande, puede ser más precisa la aproximación de la distribución binomial por la de Poisson. Sobre este y otros aspectos de aproximación, así como para la exposición de diversos ejemplos, recomendamos al lector consultar el Vol. 1 del libro de W. Feller, incluido en la bibliografía.

obtenemos, en lugar de (2.46), una fórmula del tipo

$$p_{n,k} = \frac{1}{\sqrt{npq}} \phi(x_{k,n}) e^{\varepsilon_n} \quad (2.57)$$

donde

$$|\varepsilon_n| \leq \frac{(1+C^3)(q-p)}{\sqrt{npq}} + \frac{(1+C^4)}{\sqrt{npq}}. \quad (2.58)$$

Observamos que la constante  $C$  que aparece en la condición (2.55), interviene en la acotación del error (2.57), en el sentido de que, cuanto mayor es  $C$ , para un mismo  $n$ , menos precisa es la aproximación. Lo que indica  $C$ , es cuán distantes del valor medio  $np$  pueden ser los valores  $k$  de la variable aleatoria  $S_n$  que estamos considerando. Cuanto más lejano del promedio  $np$  son dichos valores de  $k$ , mayor es el  $n$  necesario para obtener la misma aproximación.

La condición (2.56) puede ser cambiada, dado que

$$\frac{C}{\sqrt{npq}} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

Cuanto más pequeño es el segundo miembro – que en este caso es  $1/3$  – es decir, cuanto mayor es  $n$  para un valor dado de  $C$ , más precisa es la cota del error que se obtiene en lugar de (2.58).

Agreguemos finalmente que la acotación (2.58) sugiere la dependencia del error cometido al sustituir la distribución binomial por la normal, en función de los valores de  $p$ . Si  $p = q = 1/2$ , el primer término del segundo miembro de (2.58) vale 0, y la convergencia a cero – cuando  $n \rightarrow \infty$  – es esencialmente más rápida. Por otra parte, para un  $n$  dado, la cota depende de  $npq$ , o sea que es tanto menos precisa cuanto más distante es  $p$  de  $1/2$  (o sea, cuanto más próximo es  $p$  de 0 ó de 1).

## 2.7. Ejemplos.

1. Se lanza un dado 6,000 veces. Aproximar mediante la distribución normal la probabilidad de obtener el número 6 entre 990 y 1,010 veces.
- Sea  $X$  el número de veces que obtenemos 6 en 6,000 lanzamientos. Sabemos que  $X \sim b(6,000, 1/6)$ .

$$\begin{aligned} P(990 \leq X \leq 1010) &= P\left(\frac{990 - 1,000}{\sqrt{6,000 \frac{1}{6} \frac{5}{6}}} \leq \frac{X - 1,000}{\sqrt{6,000 \frac{1}{6} \frac{5}{6}}} \leq \frac{1010 - 1,000}{\sqrt{6,000 \frac{1}{6} \frac{5}{6}}}\right) \\ &\simeq \frac{1}{\sqrt{2\pi}} \int_{-0.346}^{0.346} e^{-x^2} dx \simeq 0.27. \end{aligned}$$

◀

2. En 5,000 lanzamientos de una moneda se obtuvieron 2,800 caras. ¿Es razonable suponer que la moneda no está cargada?
- La pregunta puede reformularse también de este modo: Si la moneda no está cargada, ¿cuán excepcional es el evento de que el número de caras ocurridas en 5,000 lanzamientos exceda de 2,500 al menos en 300? Si la probabilidad de este evento es muy pequeña, más que atribuir el resultado a un rarísimo acontecimiento, uno tenderá a atribuirlo a que, en realidad, la moneda está cargada, y en consecuencia la probabilidad ha sido calculada sobre una base errónea. Veamos

Primero acotamos dicha probabilidad usando la desigualdad de Chebychef:

$$n = 5,000, \quad p = 0.5 \quad np = 2,500$$

$$\begin{aligned}
P(S_n \geq 2,800) &= P(S_n - np \geq 300) = \frac{1}{2}P(|S_n - np| \geq 300) \\
&\leq \frac{1}{2} \frac{1}{(300)^2} \text{Var}(S_n) \\
&= \frac{1}{2} \frac{5,000}{4(300)^2} \simeq 0.0068.
\end{aligned}$$

Es decir que la probabilidad del suceso excepcional, está acotada superiormente por 0,0068.

Si recurrimos a la aproximación mediante la distribución normal

$$P(S_n - np \geq 300) = P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq \frac{300}{\sqrt{np(1-p)}}\right) \simeq \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-x^2/2} dx$$

con  $a = 300/\sqrt{5,000/4} \simeq 8.48$ .

La última integral se acota por

$$\frac{1}{\sqrt{2\pi} a} e^{-a^2/2}$$

y reemplazando resulta para el último término una acotación por  $0.12 \times 10^{-17}$ , que es astronómicamente pequeño. ◀

- Entre los años 2000 y 2009 inclusive nacieron en México 13,249,775 niñas y 13,230,179 niños. Queremos ver si este resultado es compatible con la hipótesis de que el sexo de los recién nacidos se distribuye al azar con probabilidad 0.5.

- ▶ Para ver consideramos que el sexo de cada recién nacido es una variable de Bernoulli con probabilidad de éxito 0.5, y diremos que ocurre un éxito si nace una niña. El número total de ensayos es 26,479,954 y el valor esperado del número de niñas es  $np = 13,239,977$ . La diferencia entre el número de niñas y su valor esperado es 9,798 y queremos hallar la probabilidad de observar una diferencia mayor o igual a este valor:

$$\begin{aligned}
P(S_n \geq 13,239,977) &= P\left(\frac{S_n - np}{\sqrt{npq}} \geq \frac{13,239,977 - np}{\sqrt{npq}}\right) \\
&\simeq P(N \geq 3,808) \\
&\simeq 7 \times 10^{-5}
\end{aligned}$$

Este número es suficientemente pequeño como para pensar que la distribución no tiene probabilidad uniforme. ◀

## 2.8. Simulación de Ensayos de Bernoulli

En esta sección revisamos algunos aspectos de la simulación de sucesiones de variables de Bernoulli y sus variables asociadas, usando R. Comenzamos por recordar la simulación de variables aleatorias en R.

### 2.8.1. Generación de Variables Aleatorias en R

El lenguaje R tiene incorporadas una serie de rutinas para generar variables aleatorias. La sintaxis precisa de la instrucción correspondiente depende de la distribución, pero todas tienen el formato común `rdist`, donde *dist* designa la distribución; por ejemplo, para generar valores a partir de la distribución normal usamos `rnorm`. Según la distribución, puede ser necesario especificar uno o varios parámetros. La tabla que presentamos a continuación incluye las distribuciones más comunes, los parámetros requeridos y sus valores por defecto. `n` representa siempre el tamaño de la muestra.

Distribución	Función en R
Binomial	<code>rbinom(n, size, prob)</code>
Poisson	<code>rpois(n, lambda)</code>
Geométrica	<code>rgeom(n, prob)</code>
Hipergeométrica	<code>rhyper(nn, m, n, k)</code>
Binomial Negativa	<code>rnbinom(n, size, prob)</code>
Multinomial	<code>rmultinom(n, size, prob)</code>
Uniforme	<code>runif(n, min=0, max=1)</code>
Exponencial	<code>rexp(n, rate=1)</code>
Gaussiana	<code>rnorm(n, mean=0, sd=1)</code>
Gamma	<code>rgamma(n, shape, scale=1)</code>
Weibull	<code>rweibull(n, shape, scale=1)</code>
Cauchy	<code>rcauchy(n, location=0, scale=1)</code>
Beta	<code>rbeta(n, shape1, shape2)</code>
t	<code>rt(n, df)</code>
Fisher	<code>rf(n, df1, df2)</code>
$\chi^2$	<code>rchisq(n, df)</code>
Logística	<code>rlogis(n, location=0, scale=1)</code>
Lognormal	<code>rlnorm(n, meanlog=0, sdlog=1)</code>

Tabla 2.1: Funciones para generar valores de distintas distribuciones en R.

Además, R tiene la función `sample` que permite obtener muestras con o sin reposición de conjuntos finitos de valores. La sintaxis es

```
sample(x, size, replace = FALSE, prob = NULL)
```

donde

- `x` es el conjunto a partir del cual queremos obtener la muestra, escrito como un vector,
- `size` es el tamaño de la muestra,
- `replace` permite indicar si se permiten repeticiones (`replace = TRUE`) o no y finalmente
- `prob` es un vector de probabilidades si se desea hacer un muestreo pesado y no uniforme.

Adicionalmente, asociada a cada distribución hay otras tres funciones que corresponden a la densidad, la función de distribución y la función de cuantiles. Estas funciones tienen una sintaxis similar a la que hemos estudiado en esta sección para la generación de variables aleatorias, pero cambiando la primera letra por `d`, `p` o `q`. Así, para la distribución normal, `dnorm(x)` corresponde los valores de la densidad en los puntos del vector `x`, `pnorm(q)` corresponde los valores de la función de distribución en los puntos del vector `q` y `qnorm(p)` nos da los cuantiles correspondientes al vector de probabilidades `p`.

### 2.8.2. Ensayos de Bernoulli

Para simular variables de Bernoulli en R usamos los comandos para generar variables binomiales con parámetro  $n = 1$ . Con las siguientes instrucciones generamos 100 variables de Bernoulli con parámetro  $p = 0.1$  y las graficamos. La primera instrucción fija la semilla del generador de números aleatorios para que los resultados se puedan reproducir.

```
set.seed(192837)
index <- 1:100
muestra <- rbinom(100,1,0.1)
plot(index, muestra, type='h',lwd=2)
```

Las siguientes instrucciones sirven para graficar en cuatro paneles, muestras de tamaño 100 para variables de Bernoulli con probabilidades éxito  $p = 0.1; 0.3; 0.5$  y  $0.7$ . El resultado se muestra en la figura 2.1

```
par(mfrow=c(4,1))
plot(index,rbinom(100,1,0.1), type='h',lwd=2, main='Ensayos de Bernoulli p=0.1',
     ylab='',bty='n',yaxt='n')
plot(index,rbinom(100,1,0.3), type='h',lwd=2, main='Ensayos de Bernoulli p=0.3',
     ylab='',bty='n',yaxt='n')
plot(index,rbinom(100,1,0.5), type='h',lwd=2, main='Ensayos de Bernoulli p=0.5',
     ylab='',bty='n',yaxt='n')
plot(index,rbinom(100,1,0.7), type='h',lwd=2, main='Ensayos de Bernoulli p=0.7',
     ylab='',bty='n',yaxt='n')
par(mfrow=c(1,1))
```

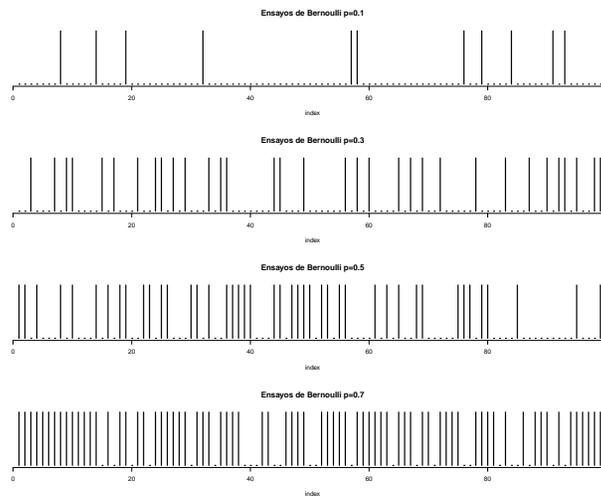


Figura 2.1: Muestras de 100 ensayos de Bernoulli con probabilidades de éxito 0.1, 0.3, 0.5 y 0.7. Las barras verticales indican los éxitos.

Una de las distribuciones asociadas a sucesiones de ensayos de Bernoulli es la geométrica, que corresponde a la variable que cuenta el número de ensayos hasta el primer éxito. La siguiente instrucción indica en cuál ensayo ocurre el primer éxito en el vector `muestra`:

```
index[muestra==1] [1]
[1] 7
```

Si, en cambio, quisiéramos el lugar del tercer éxito la expresión sería

```
index[muestra==1] [3]
[1] 28
```

Para facilitar la repetición de este procedimiento vamos a definir una función que nos permita hallar el primer éxito:

```
geo1 <- function(n=1000,p=0.1) (1:n)[rbinom(n,1,p)==1] [1]
```

La función tiene dos parámetros, el tamaño de la muestra, `n` que tiene valor por defecto igual a 1000, y la probabilidad de éxito `p`, que tiene valor por defecto 0.1. Ambos pueden fijarse al llamar la función, pero al tener valores asignados por defecto, la función se ejecuta aún cuando no se indiquen valores para estos parámetros:

```
geo1()
```

[1] 7

cuyo resultado es el lugar que ocupa el primer éxito. El número de ensayos se incluye como parámetro pues si queremos usar valores muy pequeños de la probabilidad de éxito, puede ser necesario aumentar en proporción su valor. Si queremos repetir esta función para obtener una muestra podemos usar la función `replicate`:

```
replicate(10,geo1())
[1] 1 3 5 9 19 7 8 19 1 28
```

que nos da una muestra de tamaño 10 de la distribución geométrica. A continuación usamos la función `tabulate` para crear una tabla de valores de una muestra de tamaño 10,000 de la distribución geométrica con  $p = 0.1$ . Truncamos los valores de la tabla en 50 y los graficamos. El resultado se muestra en la figura 2.2 (izq).

```
valores <- tabulate(replicate(10000,geo1()),50)
plot(1:50,valores,pch=16), main='Distribución Geométrica con p=0.1'
abline(h=0, col='red')
abline(v=0, col='red')
```

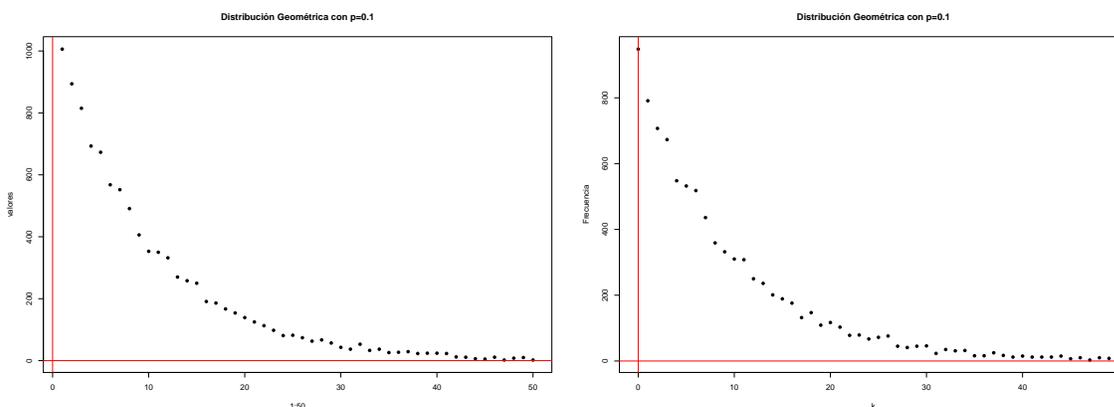


Figura 2.2: Frecuencia empírica de los valores para una distribución geométrica con parámetro  $p = 0.1$  basado en una muestra simulada de tamaño 10,000: (izq) usando la función `geo1` que construimos, (der) usando la función `rgeom`.

En R existe la función `rgeom` que genera muestras de la distribución geométrica, pero no es exactamente la misma que simulamos anteriormente. La función `rgeom` cuenta el número de fracasos previos al primer éxito, y no incluye el ensayo en el cual éste ocurre. En consecuencia, los valores difieren en 1 y la distribución que simula R incluye como valor posible 0, que corresponde a una sucesión en la cual el primer resultado es un éxito. Construimos una gráfica similar a la que acabamos de realizar con esta función, que se muestra en la figura 2.2 (der).

```
plot(0:49,tabulate(rgeom(10000,0.1),50),pch=16,xlab='k',ylab='Frecuencia',
      main='Distribución Geométrica con p=0.1')
abline(h=0, col='red')
abline(v=0, col='red')
```

En consecuencia, para obtener valores que correspondan a la distribución que definimos anteriormente, debemos añadir una unidad a los valores generados con la función `rgeom` o, equivalentemente, desplazar los valores hacia la derecha una unidad al realizar las gráficas.

Como mencionamos en la sección 2.8.1, la función `dgeom` nos da los valores de la densidad (o función de probabilidad) de la distribución geométrica. Usamos esta función a continuación para hacer gráficas de

las funciones de probabilidad correspondientes a distribuciones geométricas con parámetros 0.1, 0.3, 0.5 y 0.7. El resultado se presenta en la figura 2.3.

```
xval <- 0:40; clr <- rainbow(50)
plot(xval+1, dgeom(xval,0.8), type='l',main='Densidad de la Distribución Geométrica',
      lwd=2, ylab='Densidad', xlab='x',col=clr[4])
lines(xval+1, dgeom(xval,0.5),col=clr[14],lwd=2)
lines(xval+1, dgeom(xval,0.3),col=clr(50)[34],lwd=2)
lines(xval+1, dgeom(xval,0.1),col=clr[44],lwd=2)
legend('topright', c('0.7','0.5','0.3','0.1'),lwd=rep(2,4),
      col=c(clr[4],clr[14],clr[34],clr[44]))
abline(h=0, col='red')
abline(v=0, col='red')
```

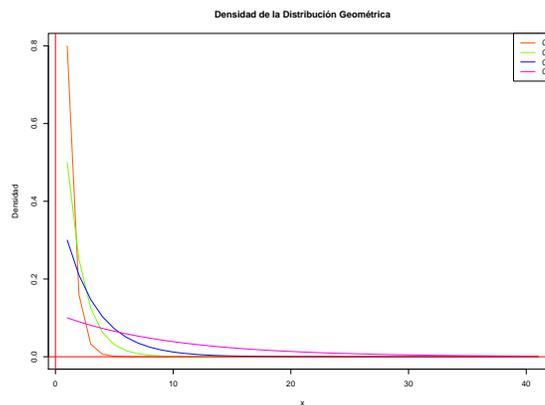


Figura 2.3: Densidades para la distribución geométrica con parámetros 0.1, 0.3, 0.5 y 0.7.

Sin embargo, la representación de la función de probabilidad que hemos hecho para estas distribuciones no es la más adecuada, pues las variables son discretas. A continuación haremos una representación más apropiada, usando barras verticales para los valores de la función de probabilidad. La primera instrucción guarda en `op` los valores por defecto de los parámetros gráficos, de modo que si los modificamos para lograr una gráfica mejor, los podemos restaurar al final con la última instrucción. En el proceso de elaboración de las gráficas cambiamos en dos ocasiones los márgenes, cuyos valores por defecto tienden a ser muy grandes cuando se usan varios paneles en una misma representación. El resultado se presenta en la figura 2.4.

```
op <- par(no.readonly = TRUE)
par(mfrow=c(2,2))
par(mar=c(4,4,4,2)+0.1)
plot(xval+1, dgeom(xval,0.7), type='h',main='Distribución Geométrica, p=0.7',lwd=2,
      ylab='Densidad', xlab='x')
plot(xval+1, dgeom(xval,0.5),col=clr[4],lwd=2,type='h',
      main='Distribución Geométrica, p=0.5',ylab='Densidad', xlab='x')
par(mar=c(5,4,3,2)+0.1)
plot(xval+1, dgeom(xval,0.3),col=clr[28],lwd=2,type='h',
      main='Distribución Geométrica, p=0.3',ylab='Densidad', xlab='x')
plot(xval+1, dgeom(xval,0.1),col=clr[34],lwd=2,type='h',
      main='Distribución Geométrica, p=0.1',ylab='Densidad', xlab='x')
```

```
par(mfrow=c(1,1))
par(op)
```

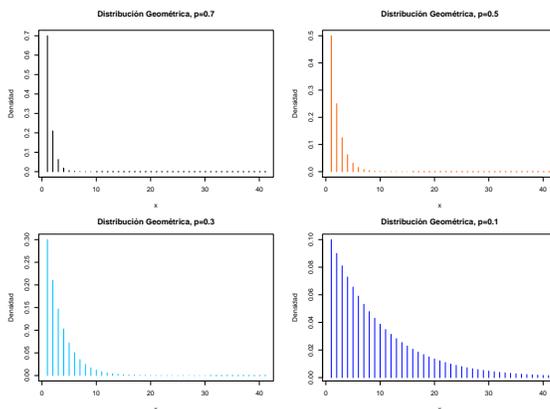


Figura 2.4: Densidades para la distribución geométrica con parámetros 0.1, 0.3, 0.5 y 0.7.

Otra distribución asociada a una sucesión de ensayos de Bernoulli es la binomial negativa, que cuenta el número de ensayos necesarios hasta obtener el  $k$ -ésimo éxito. Una manera de simular esta variable es sumar  $k$  variables con distribución geométrica, como en el ejemplo que sigue, en el cual  $k = 3$ .

```
set.seed(1234)
sum(replicate(3,geo1()))
[1] 35
```

Otra alternativa es modificar la función `geo1` para que halle el  $k$ -ésimo éxito en lugar del primero:

```
bineg <- funcion(k=2,nn=1000,pp=0.1) (1:nn)[rbinom(nn,1,pp)==1][k]
bineg()
[1] 10
bineg(10)
[1] 95
```

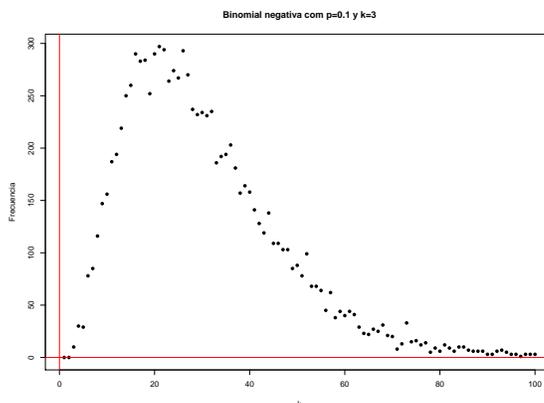


Figura 2.5: Frecuencia empírica de los valores para una distribución binomial negativa con parámetro  $p = 0.1$  para  $k = 3$  basado en una muestra simulada de tamaño 10,000.

A continuación simulamos una muestra de tamaño 10,000 y graficamos la frecuencia de los resultados hasta el valor  $k = 100$ . El resultado se muestra en la figura 2.5.

```
valoresbn <- tabulate(replicate(10000,bineg(3)),100)
plot(1:100,valoresbn,pch=16,main='Binomial negativa com p=0.1 y k=3',
     xlab='k',ylab='Frecuencia')
abline(h=0, col='red')
abline(v=0, col='red')
```

En R también existe una función `rnbinom` para generar muestras de la distribución binomial negativa, así como funciones para la densidad `dnbinom`, función de distribución `pnbinom` y función de cuantiles `qnbinom`, pero al igual que en el caso de la distribución geométrica, los valores de esta variable en R no incluyen los éxitos y solo cuentan el número de fracasos hasta obtener el  $k$ -ésimo éxito. Por lo tanto, para que los valores en R coincidan con los que hemos definido en este curso, debemos sumar  $k$  al valor de esta variable.

A continuación presentamos algunas gráficas de la densidad o función de probabilidad de variables con distribución binomial negativa con diversos valores de los parámetros.

```
op <- par(no.readonly = TRUE)
par(mfrow=c(2,2))
par(mar=c(4,4,4,2)+0.1)
plot(xval+3, dnbinom(xval,3,0.7), type='h',main='Distribución Binomial Negativa,
     k=3,p=0.7',lwd=2, ylab='Densidad', xlab='x',xlim=c(0,50))
plot(xval+3, dnbinom(xval,3,0.5),col=clrs[4],lwd=2,type='h',ylab='Densidad',
     main='Distribución Binomial Negativa, k=3, p=0.5', xlab='x',xlim=c(0,50))
par(mar=c(5,4,3,2)+0.1)
plot(xval+3, dnbinom(xval,3,0.3),col=clrs[28],lwd=2,type='h',ylab='Densidad',
     main='Distribución Binomial Negativa, k=3, p=0.3', xlab='x',xlim=c(0,50))
plot(xval+3, dnbinom(xval,3,0.1),col=clrs[34],lwd=2,type='h',ylab='Densidad',
     main='Distribución Binomial Negativa, k=3, p=0.1', xlab='x',xlim=c(0,50))

par(mar=c(4,4,4,2)+0.1)
plot(xval+3, dnbinom(xval,3,0.3), type='h',main='Distribución Binomial Negativa,
     k=3,p=0.3',lwd=2,ylab='Densidad', xlab='x',xlim=c(0,50))
plot(xval+5, dnbinom(xval,5,0.3),col=clrs[4],lwd=2,type='h',ylab='Densidad'
     ,main='Distribución Binomial Negativa, k=5, p=0.3', xlab='x',xlim=c(0,50))
par(mar=c(5,4,3,2)+0.1)
plot(xval+7, dnbinom(xval,7,0.3),col=clrs[28],lwd=2,type='h',ylab='Densidad',
     main='Distribución Binomial Negativa, k=7, p=0.3', xlab='x',xlim=c(0,50))
plot(xval+9, dnbinom(xval,9,0.3),col=clrs[34],lwd=2,type='h',ylab='Densidad',
     main='Distribución Binomial Negativa, k=9, p=0.3', xlab='x',xlim=c(0,50))
par(mfrow=c(1,1))
par(op)
```

Los resultados se muestran en las dos siguientes gráficas.

Finalmente consideramos la distribución binomial, que está vinculada a los ensayos de Bernoulli de manera natural, ya que una variable con distribución de parámetros  $n$  y  $p$  cuenta el número de éxitos en una sucesión de  $n$  ensayos de Bernoulli independientes con probabilidad de éxito igual a  $p$ . Presentamos varias animaciones realizadas en R. Para comenzar fijamos  $n = 10$  y variamos  $p$  entre 0 y 1 con pasos de 0.01

```
binom <- function(p){
  plot(0:10,dbinom(0:10,10,p),type='h',lwd=5,ylim=c(0,0.5),
     xlab='Valores',ylab='Probabilidad')
  Sys.sleep(0.1)}
ignorar <- sapply((0:100)/100,binom)
```

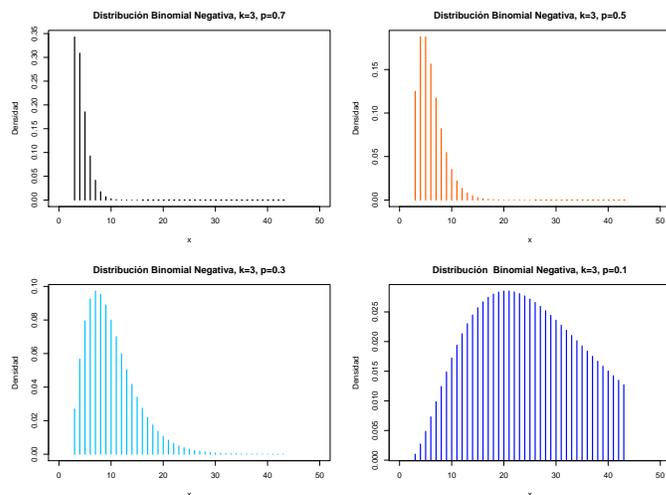


Figura 2.6: Función de probabilidad binomial negativa con  $k = 3$  y  $p = 0.7, 0.5, 0.3$  y  $0.1$ .

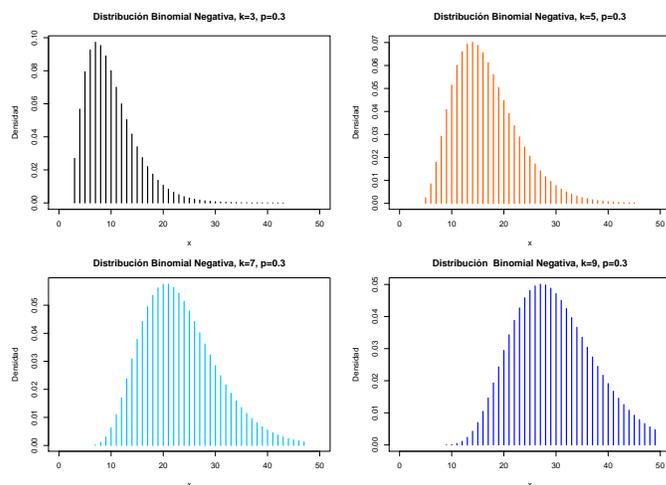


Figura 2.7: Función de probabilidad binomial negativa con  $p = 0.3$  y  $pk = 3, 5, 7$  y  $9$ .

A continuación fijamos  $p = 0.5$  y variamos  $n$  de 1 a 100:

```
binom2<- function(n){
  plot(0:n,dbinom(0:n,n,0.5),type='h',lwd=5,ylim=c(0,0.5),
  xlab='Valores',ylab='Probabilidad')
  Sys.sleep(0.1)}
ignorar <- sapply((0:100),binom2)
```

En las siguientes gráficas cambiamos la escala en el eje  $x$  y graficamos para  $n$  hasta 200.

```
binom3<- function(n){
  plot(0:n,dbinom(0:n,n,0.5),type='h',lwd=5,ylim=c(0,0.5),
  xlab='Valores',ylab='Probabilidad',
  xlim=c((n/2)-2*sqrt(n),(n/2)+2*sqrt(n)))
  Sys.sleep(0.08) }
ignorar <- sapply((0:200),binom3)
```

Finalmente cambiamos el valor de  $p$  a 0.25 y simulamos con  $n$  entre 0 y 200. La primera gráfica es sin modificar la escala.

```
binom4<- function(n){
  plot(0:n,dbinom(0:n,n,0.25),type='h',lwd=5,ylim=c(0,0.5),
  xlab='Valores',ylab='Probabilidad')
  Sys.sleep(0.08)}
ignorar <- sapply((0:200),binom4)
binom5<- function(n){
  plot(0:n,dbinom(0:n,n,0.25),type='h',lwd=5,ylim=c(0,0.5),
  xlab='Valores',ylab='Probabilidad',
  xlim=c((n/4)-2*sqrt(n),(n/4)+2*sqrt(n)))
  Sys.sleep(0.08)}
ignorar <- sapply((0:200),binom4)
```