Capítulo 7

Teorema Central del Límite

7.1. La Ley Débil de los Grandes Números.

Lema 7.1 (Desigualdad de Markov) *Sea $X \geq 0$ una variable aleatoria y a un número positivo, entonces*

$$P(X \geq a) \leq \frac{1}{a} \operatorname{E}(X). \tag{7.1}$$

Demostración Si $\operatorname{E}(X) = +\infty$, 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

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

Dividiendo por a se obtiene (7.1). ∎

Lema 7.2 (Desigualdad de Tchebichev) *Sea X una variable aleatoria con varianza finita y $\varepsilon > 0$. Entonces*

$$P(|X - \operatorname{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\operatorname{Var}(X). \tag{7.2}$$

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

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

que es (7.2). ∎

Teorema 7.1 (Ley Débil de los Grandes Números) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes,*

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

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

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0. \tag{7.3}$$

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

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

$$\mathrm{Var}(\frac{S_n}{n}) = \frac{1}{n^2}\,\mathrm{Var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Aplicando (7.2)

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

cuando $n \to \infty$. Esto prueba la ley débil de los grandes números. ■

7.1.1. 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

   $$P(|\hat{p}_n - p| \geq \varepsilon) \to 0 \quad \text{para todo } \varepsilon > 0 \text{ cuando } n \to \infty. \tag{7.4}$$

   La propiedad (7.4), de aproximación entre $\hat{p}_n$ (que es función de las observaciones empíricas que podemos efectuar), y el número p (que es un cierto parámetro del problema), nos indica que si el número n de observaciones es bastante grande, es entonces pequeña la probabilidad de que la distancia de $\hat{p}_n$ a p sea mayor que un número dado.

   La desigualdad de Tchebichev nos dá, además, una acotación de la velocidad con la que el primer miembro de (7.4) tiende a cero cuando $n \to \infty$. En el ejemplo que estamos considerando, como ya hemos visto en la sección 6.3,

   $$\mathrm{E}(\hat{p}_n) = p \qquad \mathrm{Var}(\hat{p}_n) = \frac{p(1-p)}{n}$$

   y entonces

   $$P(|\hat{p}_n - p| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\frac{p(1-p)}{n} \tag{7.5}$$

   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

$$P(|\hat{p}_n - p| \geq 0.01) \leq 0.05. \tag{7.6}$$

Usando (7.5) tenemos

$$P(|\hat{p}_n - p| \geq 0.01) \leq \frac{1}{(0.01)^2} \frac{p(1-p)}{n}. \tag{7.7}$$

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

$$\frac{1}{(0.01)^2} \frac{1}{4n} \leq 0.05, \tag{7.8}$$

resultará

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

Es sencillo ver que (7.8) se verifica si $n \geq 50.000$.

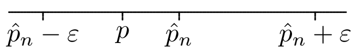


Figura 7.1

En principio, la acotación dada por la desigualdad de Tchebichev es grosera. En el ejemplo numérico que estamos analizando, se podrían dar acotaciones mucho más precisas del primer miembro que figura en la desigualdad (7.7), lo cual permitiría reducir considerablemente el valor de n a partir del cual se verifica (7.6). 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 (7.6) está acotado por 0.00001, es decir, es mucho menor que el segundo miembro.

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

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

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

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

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

$$\mathrm{E}(\hat{\lambda}_t) = \frac{1}{t} \mathrm{E}(N_t) = \frac{1}{t} \lambda t = \lambda$$

$$\mathrm{Var}(\hat{\lambda}_t) = \frac{1}{t^2}\,\mathrm{Var}(N_t) = \frac{1}{t^2}\lambda t = \frac{\lambda}{t} \to 0 \quad (t \to +\infty)$$

En consecuencia

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

para cada $\varepsilon > 0$,

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

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 $X_1, X_2, \ldots$ son independientes y tienen igual varianza. Se puede verificar, sin gran esfuerzo, que vale una demostración análoga si las variables no están correlacionadas y

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

donde $\sigma_i = \mathrm{Var}(X_i)$. 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 μ, es decir, la esperanza matemática, y el promedio experimental es

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

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 (7.3), 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.

7.2. Convergencia en Probabilidad y Convergencia Casi Segura.

Definición 7.1 Diremos que una sucesión de variables aleatorias $Y_1, Y_2, \ldots$ *converge en probabilidad* a la variable aleatoria Y, si se verifica que para todo $\varepsilon > 0$

$$P(|Y_n - Y| \geq \varepsilon) \to 0 \tag{7.9}$$

cuando $n \to \infty$. En este caso usamos la notación $Y_n \xrightarrow{P} Y \quad (n \to \infty)$.

Ejemplo Básico.
La ley débil de los grandes números nos da un ejemplo de convergencia en probabilidad. Allí $Y_n = S_n/n$ es el promedio de las n variables aleatorias $X_1, X_2, \ldots, X_n$ y la variable aleatoria Y es la constante μ, que es el valor esperado común de las X_n.

Definición 7.2 Diremos que una sucesión de variables aleatorias $Y_1, Y_2, \ldots$ *converge casi seguramente (o con probabilidad 1)* a la variable aleatoria Y, si se verifica que

$$P(Y_n \to Y \text{ cuando } n \to \infty) = 1 \tag{7.10}$$

es decir, si salvo para un conjunto de eventos ω cuya probabilidad es igual a cero, se verifica que exite el límite

$$\lim_{n\to\infty} Y_n(\omega) = Y(\omega).$$

Usaremos la notación $Y_n \to Y$ c.s. o también $Y_n \to Y$ c.p.1.

Ejemplo. Si $Y_n(\omega) \to Y(\omega)$ cuando $n \to \infty$ para todo $\omega \in \Omega$, es claro que se verifica (7.10), ya que en este caso el conjunto

$$\{\omega : Y_n(\omega) \text{ no converge a } Y(\omega)\} \tag{7.11}$$

de los eventos ω tales que $Y_n(\omega)$ o bien no tiene límite, o bien tiene un límite que no es $Y(\omega)$, no es otro que el vacío, y por lo tanto su probabilidad es igual a cero.

Obviamente, este ejemplo no es demasiado interesante. Importarán situaciones en las que hay convergencia casi segura, pero el conjunto (7.11) *no* es vacío, aunque tenga probabilidad nula.

La siguiente proposición y la observacin posterior aclaran la relación entre ambos modos de convergencia. La demostración puede ser consultada en el apéndice a este capítulo.

Proposición 7.1 *Sea $\{Y_n\}$ una sucesión de variables aleatorias que converge casi seguramente a la variable aleatoria Y. Entonces Y_n converge a Y en probabilidad.*

Observación 7.1 El recíproco de esta propiedad es falso. Ver ejercicio 5 de este capítulo.

7.3. La Ley Fuerte de los Grandes Números.

Teorema 7.2 (Ley Fuerte de los Grandes Números) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes tal que*

$$\mathrm{E}(X_n) = \mu, \qquad \mathrm{E}(X_n^4) \leq C \quad n = 1, 2, \dots$$

donde C es una constante. Entonces,

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

converge a μ casi seguramente.

La demostración de este resultado se puede encontrar en el apéndice.

7.3.1. Ejemplos y Comentarios.

1. En el caso de la distribución binomial, se verifican las hipótesis de la ley fuerte de los grandes números. En efecto, si tenemos n observaciones independientes del suceso A, $p = P(A)$, definimos

$$X_i = \begin{cases} 1 & \text{si } A \text{ ocurre en la } i\text{-ésima observación} \\ 0 & \text{si } A \text{ no ocurre en la } i\text{-ésima observación,} \end{cases}$$

   con lo cual $S_n = X_1 + X_2 + \cdots + X_n$ es el número de éxitos en la observación de A. Entonces $\mathrm{E}(X_i) = p$ para $i = 1, 2, \dots$ y $\mathrm{E}(X_i^4)$ es acotado (por la constante 1, por ejemplo, ya que $X_i^4 \leq 1$). Se concluye que $S_n/n \to p$ casi seguramente.

   Esta relación entre frecuencia relativa y probabilidad, precisa la respuesta originada en la ley débil de los grandes números y responde a la problemática planteada anteriormente, en torno a la vinculación entre observaciones experimentales y probabilidad.

2. La conclusión de la ley fuerte de los grandes números sigue valiendo bajo hipótesis más débiles que las del teorema que hemos demostrado. Por ejemplo, podemos reemplazar la hipótesis de que los momentos de orden 4 ($\mathrm{E}(X_i^4)$) estén acotados, por la hipótesis (menos exigente) de que

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathrm{Var}(X_n) < \infty$$

   y obtener el mismo resultado. Diversas generalizaciones son posibles, con aplicaciones diversas también. El lector interesado puede consultar el texto de W. Feller citado en la bibliografía.

7.4. 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 7.3 (Fórmula de Stirling) *Cuando $n \to \infty$, se tiene*

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

Demostración Se trata de probar que

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

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 $f(x) = \log x$ 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 $\overline{P_k\, P_{k+1}}$. 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

$$\begin{aligned} A_n = \int_1^n \log x\, dx &> \sum_{k=1}^{n-1} \frac{1}{2}(\log k + \log(k+1)) \\ &= \log 2 + \log 3 + \cdots + \log(n-1) + \frac{1}{2}\log n. \end{aligned} \tag{7.12}$$

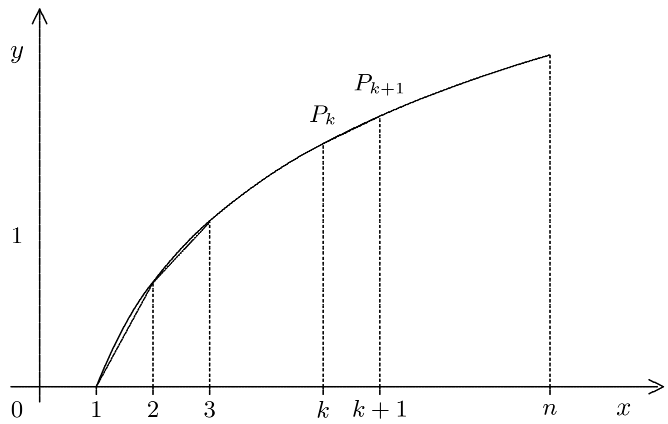


Figura 7.3

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.

Además, si llamamos b_k al área sombreada en la figura 7.4, donde el segmento $\overline{P'_k\, P'_{k+1}}$ es tangente al gráfico de $f(x) = \log x$ en el punto de abcisa $x = k + 1/2$, es claro que

$$a_n < \sum_{k=1}^{n-1} b_k, \qquad (n > 1), \tag{7.13}$$

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 (7.13)

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

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

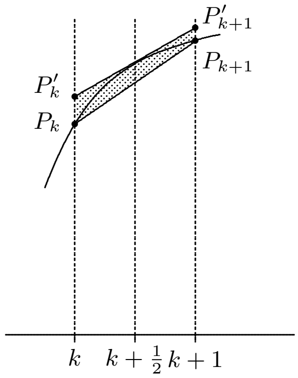


Figura 7.4

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

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

Tomando exponenciales en ambos miembros

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

de donde

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

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

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

de donde

$$I_n = \frac{n-1}{n} I_{n-2}. \tag{7.14}$$

Por lo tanto, considerando separadamente los casos en los cuales n es par e impar, y aplicando (7.14) reiteradamente, obtenemos

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

Además

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

o sea que

$$I_{2p} = \frac{2p-1}{2p}\frac{2p-3}{2p-2}\cdots\frac{3}{4}\frac{1}{2}\frac{\pi}{2}; \qquad I_{2p} = \frac{2p}{2p+1}\frac{2p-2}{2p-1}\cdots\frac{4}{5}\frac{2}{3}. \tag{7.15}$$

Observamos ahora que $\{I_n\}$ es una sucesión decreciente, ya que para $0 \le x \le \pi/2$ se tiene $0 \le \operatorname{sen} x \le 1$, $\operatorname{sen}^n x$ 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 (7.14) tenemos

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

y por lo tanto la sucesión intermedia

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

también tiende a 1 cuando $p \to \infty$. Usando ahora (7.15) esto indica que

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

y en consecuencia

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

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

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

Utilizamos ahora el resultado del paso 1. Sabemos que

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

Sustituyendo

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

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

7.5. 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

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{7.16}$$

Denotamos la función de distribución respectiva por

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt. \tag{7.17}$$

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 $\mathrm{E}(S_n) = np$, $\mathrm{Var}(S_n) = npq$.

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 (7.17).

Teorema 7.3 *Sean $a < b$ números reales, entonces*

$$P(a < S_n' \le b) \longrightarrow \Phi(b) - \Phi(a) = \int_a^b \phi(t)\, dt, \quad \text{cuando } n \to \infty. \tag{7.18}$$

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

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

donde la suma se extiende sobre los enteros k tales que

$$a < \frac{k - np}{\sqrt{npq}} \leq b. \tag{7.19}$$

Comencemos por dar una aproximación para cada sumando $p_{n,k}$, donde k verifica (7.18). Para facilitar la notación ponemos $\delta = k - np$. Si k verifica (7.18), entonces

$$\frac{\delta}{n} \to 0 \quad \text{cuando } n \to \infty \tag{7.20}$$

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 $1/\sqrt{n} \to 0$ tenemos

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

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

Teniendo en cuenta que bajo la condición (7.24) –que se verifica puesto que sólo nos interesan los valores de k que cumplen (7.18)– tanto k como $n - k$ tienden a $+\infty$ cuando $n \to \infty$, tenemos

$$\begin{aligned} p_{n,k} &= \frac{\sqrt{2\pi n} n^n e^{-n} p^k q^{n-k} e^{\gamma_n - \gamma_k - \gamma_{n-k}}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi (n-k)} (n-k)^{(n-k)} e^{-(n-k)}} \\ &= \frac{1}{\sqrt{2\pi}}\ \sqrt{\frac{n}{(np+\delta)(nq-\delta)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} e^{\gamma_n - \gamma_k - \gamma_{n-k}}. \end{aligned} \tag{7.21}$$

El primer factor lo podemos escribir como

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

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

Tomando logaritmos en el segundo 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 $|x| < 1$:

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

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

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

Apliquemos 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 $\frac{\delta}{n} \to 0$, 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

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

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 (7.21)

$$\begin{aligned}
p_{n,k} &= \frac{1}{\sqrt{2\pi npq}} \exp\{-\frac{\delta^2}{2npq} + B\frac{\delta^3}{n^2} + \theta_{n,k}\} && (7.22) \\
&= \frac{1}{\sqrt{npq}}\phi(x_{n,k}) \exp\{B\frac{\delta^3}{n^2} + \theta_{n,k}\} \\
&= \frac{1}{\sqrt{npq}}\phi(x_{n,k}) e^{\alpha_{n,k}} && (7.23)
\end{aligned}$$

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 $\frac{\delta^3}{n^2} \to 0$ cuando $n \to \infty$ 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 (7.18).

Volvamos ahora a nuestra suma inicial. Sustituyendo en ella la expresión hallada en (7.23) para $p_{n,k}$, obtenemos

$$\begin{aligned}
P(a < S'_n \leq b) &= \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}}\phi(x_{k,n}) e^{\alpha_{n,k}} \\
&= \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}}\phi(x_{k,n}) + \sum_{a < \frac{k-np}{\sqrt{npq}} \leq b} \frac{1}{\sqrt{npq}}\phi(x_{k,n})\left[e^{\alpha_{n,k}} - 1\right]. && (7.24)
\end{aligned}$$

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

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

el primer sumando tiende a

$$\int_a^b \phi(t)\, dt \tag{7.25}$$

cuando $n \to \infty$.

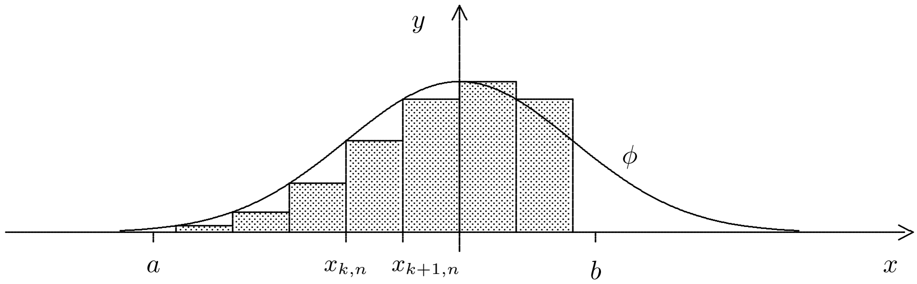


Figura 7.5

En cuanto al segundo sumando de (7.24), 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 $n \to \infty$ mientras que el segundo tiende a la integral (7.25) Por lo tanto el segundo sumando de (7.24) tiende a cero. Esto termina la demostración del teorema 7.3.

■

Teorema 7.4

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

Demostración Nos apoyamos en el teorema 7.3. Sea $\varepsilon < 0$, elegimos a_1 y a_2 tales que $a_1 < x < a_2$ (ver figura 7.6) y

$$\Phi(a_1) < \frac{\varepsilon}{2}, \qquad 1 - \Phi(a_2) < \frac{\varepsilon}{2}. \tag{7.26}$$

Entonces

$$1 - P(a_2 \geq S'_n > x) \geq 1 - P(S'_n > x) = P(S'_n \leq x) \geq P(a_1 < S'_n \leq x).$$

El primer miembro tiende a

$$1 - (\Phi(a_2) - \Phi(x)) = \Phi(x) + (1 - \Phi(a_2))$$

y el último a

$$\Phi(x) - \Phi(a_1)$$

cuando $n \to \infty$.

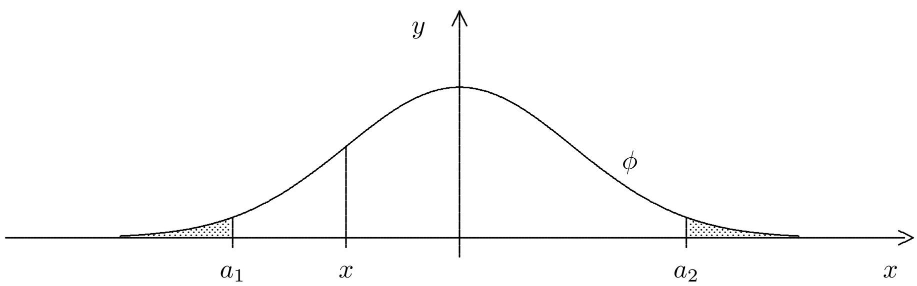


Figura 7.6

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

$$\Phi(x) + (1 - \Phi(a_2)) + \frac{\varepsilon}{2} \geq P(S'_n \leq x) \geq \Phi(x) - \Phi(a_1) - \frac{\varepsilon}{2}$$

y dada la forma en que han sido elegidos a_1, a_2, para $n \geq N$

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

o sea

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

■

Observación 7.2 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

$$\frac{\delta}{n} \to 0 \tag{7.27}$$

y

$$\frac{\delta^3}{n^2} \to 0. \tag{7.28}$$

Ahora bien, es claro que (7.31) implica (7.30) 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 (7.31) 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 7.3 y 7.4 para probar que si

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

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

$$P\left(\frac{k - np}{\sqrt{npq}} > a_n \right) \approx \int_{a_n}^{+\infty} \phi(t)dt \quad \text{cuando } n \to \infty \tag{7.29}$$

donde ϕ es, como antes, la densidad normal típica y el símbolo "$\approx$" 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 $n \to \infty$, 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

$$1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} dt. \tag{7.30}$$

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

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

En segundo lugar mostraremos que

$$1 - \Phi(x) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} \quad \text{cuando } x \to \infty \tag{7.31}$$

Queremos probar ahora que

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

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

$$\frac{-\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{1}{\sqrt{2\pi}} \left(-\frac{1}{x^2} e^{-x^2/2} + \frac{1}{x}(-xe^{-x^2/2}) \right)} = \frac{1}{\frac{1}{x^2} + 1} \overset{x\to\infty}{\longrightarrow} 1,$$

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 (7.31).

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, sustituímos 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 $1/2$ [1].

   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 (7.23), 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,

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

[1] 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.

Finalmente, en la fórmula (7.25), también aproximamos

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

mediante un desarrollo de Mac-Laurin.

Sustituyendo en (7.21), si agregamos, por ejemplo, la condición

$$|k-np| = |\delta| \leq C\ \sqrt{npq} \tag{7.32}$$

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

$$\frac{C}{\sqrt{npq}} < \frac{1}{3}, \tag{7.33}$$

obtenemos, en lugar de (7.26), una fórmula del tipo

$$p_{n,k} = \frac{1}{\sqrt{npq}}\phi(x_{k,n})e^{\varepsilon_n} \tag{7.34}$$

donde

$$|\varepsilon_n| \leq \frac{(1+C^3)(q-p)}{\sqrt{npq}} + \frac{(1+C^4)}{\sqrt{npq}}. \tag{7.35}$$

Observamos que la constante C que aparece en la condición (7.32), interviene en la acotación del error (7.34), 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 (7.33) puede ser cambiada, dado que

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

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 (7.35).

Agreguemos finalmente que la acotación (7.35) 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 (7.35) vale 0, y la convergencia a cero – cuando $n \to \infty$ – 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).

7.6. Ejemplos.

1. Verificar que, para el cálculo de límites en probabilidad, valen las mismas reglas que para el cálculo de límites ordinarios. Por ejemplo:

$$X_n \xrightarrow{P} X, \quad Y_n \xrightarrow{P} Y \Rightarrow X_n + Y_n \xrightarrow{P} X+Y, \quad X_nY_n \xrightarrow{P} XY, \text{ etc.}$$

$$\left.\begin{array}{l} X_n \xrightarrow{P} X \\ g:\mathbb{R}\to\mathbb{R} \text{ continua} \end{array}\right\} \Rightarrow g(X_n) \xrightarrow{P} g(X).$$

► Vamos a probar solamente el caso de la suma. Los restantes son análogos y no los expondremos en esta solución.

Sea entonces $X_n \xrightarrow{P} X$ e $Y_n \xrightarrow{P} Y$. Para $\varepsilon > 0$ tenemos la siguiente inclusión de eventos

$$\{|(X_n + Y_n) - (X + Y)| \geq \varepsilon\} \subset \{|X_n - X| \geq \frac{\varepsilon}{2}\} \cup \{|Y_n - Y| \geq \frac{\varepsilon}{2}\}. \tag{7.36}$$

Esta inclusión es cierta ya que si ocurre que

$$|(X_n + Y_n) - (X + Y)| \geq \varepsilon$$

alguno de los dos conjuntos que aparece en el segundo miembro tiene que ocurrir (y por lo tanto la unión), porque de no ocurrir ninguno de los dos se tendría

$$|X_n - X| < \frac{\varepsilon}{2} \quad \text{y} \quad |Y_n - Y| < \frac{\varepsilon}{2}$$

simultaneamente, de donde se obtiene que $|X_n - X + Y_n - Y| < \varepsilon$, contra lo supuesto.

De la relación (7.36):

$$P(|(X_n + Y_n) - (X + Y)| \geq \varepsilon) \leq P(|X_n - X| \geq \frac{\varepsilon}{2}) + P(|Y_n - Y| \geq \frac{\varepsilon}{2})$$

y como cada término del segundo miembro tiende a cero, lo mismo ocurre con el primero. ◄

2. Se considera una sucesión $X_1, X_2, \dots$ de variables aleatorias independientes con la misma distribución de probabilidad y momentos de cuarto orden finitos. Probar que

a. $\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \xrightarrow{P} (\mathrm{E}(X_1))^2$.

b. $\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 \xrightarrow{P} \mathrm{Var}(X_1)$, donde $\overline{X} = \frac{1}{n}\sum_1^n X_i$.

► a. Como

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mathrm{E}(X_1)$$

y $g(x) = x^2$ es una función continua, se obtiene el resultado.

b. Poniendo $\mathrm{E}(X_1) = \mu$, escribimos

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 &= \frac{1}{n}\sum_{i=1}^n (X_i - \mu + \mu - \overline{X})^2 \\
&= \frac{1}{n}\left(\sum_{i=1}^n (X_i - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^n (X_i - \mu) + n(\overline{X} - \mu)^2\right) \\
&= \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - (\overline{X} - \mu)^2.
\end{aligned}$$

El primer término en el último miembro es una suma de variables aleatorias independientes con igual distribución y segundo momento finito (porque lo es el de cuarto orden de X_i) y por lo tanto, le podemos aplicar la ley débil de los grandes números, es decir, que

$$\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 \xrightarrow{P} \mathrm{E}((X_1 - \mu)^2) = \mathrm{Var}(X_1).$$

En cuanto al segundo término, como $\overline{X} \xrightarrow{P} \mu$, se tiene que $(\overline{X} - \mu) \xrightarrow{(P)} 0$, obteniendose el resultado indicado en b. ◄

3. Se arroja 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 $X \sim b(6.000, 1/6)$.

$$
\begin{aligned}
P(990 \le X \le 1010) &= P\left(\frac{990 - 1.000}{\sqrt{6.000\frac{1}{6}\frac{5}{6}}} \le \frac{X - 1.000}{\sqrt{6.000\frac{1}{6}\frac{5}{6}}} \le \frac{1010 - 1.000}{\sqrt{6.000\frac{1}{6}\frac{5}{6}}}\right) \\
&\simeq \frac{1}{\sqrt{2\pi}} \int_{-0{,}346}^{0{,}346} e^{-x^2}\, dx \simeq 0,27.
\end{aligned}
$$

◄

4. 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 Tchebichev:

$$
n = 5.000, \qquad p = 0,5 \qquad np = 2.500
$$

$$
\begin{aligned}
P(S_n \ge 2.800) = P(S_n - np \ge 300) &= \frac{1}{2} P(|S_n - np| \ge 300) \\
&\le \frac{1}{2} \frac{1}{(300)^2} \operatorname{Var}(S_n) \\
&= \frac{1}{2} \frac{5.000}{4(300)^2} \simeq 0,0068.
\end{aligned}
$$

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

Si recurrimos a la aproximación mediante la distribución normal

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

con $a = 300/\sqrt{5.000/4} \simeq 8,48$.

La última integral se acota por

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

y reemplazando resulta para el último término una acotación por $0,12 \times 10^{-17}$, que es astronómicamente pequeño. ◄

5. Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes, cada una con distribución uniforme en $(0, 1)$ y

$$
m_n = \min\{X_1, X_2, \dots, X_n\}.
$$

a. Probar que $m_n \xrightarrow{P} 0$.

► a. Sea $0 < \varepsilon < 1$.

$$\begin{aligned} P(|m_n| \geq \varepsilon) &= P(m_n \geq \varepsilon) = P(X_1 \geq \varepsilon, X_2 \geq \varepsilon, \dots, \geq \varepsilon) \\ &= \prod_{i=1}^{n} P(X_i \geq \varepsilon) = (1-\varepsilon)^n \end{aligned}$$

que tiende a cero cuando $n \to \infty$, porque $0 < 1 - \varepsilon < 1$.

◄

Ejercicios

1. Calcular una aproximación de la probabilidad de que el número de 6 obtenidos al lanzar un dado perfecto 12.000 veces, este comprendido entre 1.900 y 2.150.

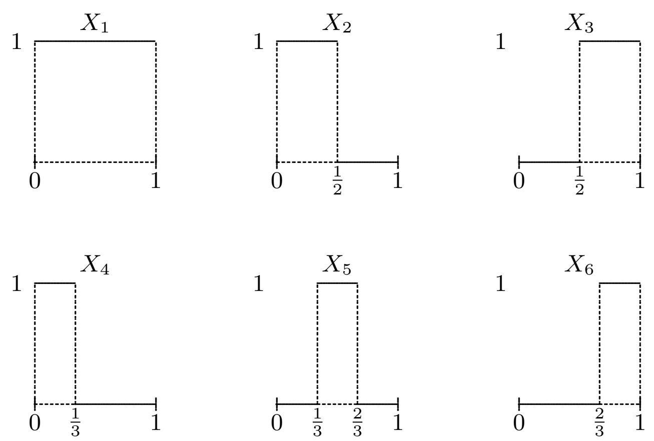
2. Encontrar el número k tal que la probabilidad de obtener entre 490 y k veces cara en 1.000 lanzamientos de una moneda, sea igual a $1/2$.

3. Se toma una muestra al azar con reposición, a efectos de estimar la fracción p de hembras en una población. Encontrar un tamaño de muestra que asegure que la estimación se hará con un error de menos de $0,005$, al menos con probabilidad 0,99.

4. Se desea estimar la probabilidad de falla p en un proceso de producción mediante la observación de n objetos producidos, elegidos independientemente. Se sabe que p está comprendida entre 0,1 y 0,3, en virtud de la información previa de que se dispone sobre el proceso. Se desea hallar el tamaño n de la muestra para que la probabilidad de que la frecuencia relativa de objetos defectuosos en la muestra difiera del verdadero valor p en más de 0,01 sea menor que 0,05.

5. Probar que el siguiente es un ejemplo de una sucesión de variables aleatorias $\{X_n\}$ que converge en probabilidad pero no casi seguramente.

   Tomamos $\Omega = [0, 1]$ con la probabilidad uniforme. Definimos la sucesión $\{X_n\}$ de acuerdo a lo que sugiere la figura 7.7.



   Probar que $P(|X_n| \geq \varepsilon) \to 0$ cuando $n \to \infty$ para todo $\varepsilon > 0$, pero que, en cambio, no sólo no es cierto que $X_n \to 0$ casi seguramente, sino que $X_n(\omega)$ no converge para ningún ω.

6. Aproximación de la distribución binomial por la distribución de Poisson.

Sea $\{X_n\}$ una sucesión de variables aleatorias con distribución binomial de parámetros (n, p_n). Supongamos que $n \to \infty$ y $p_n \to 0$ simultáneamente, y de tal manera que $np_n = \lambda > 0$, donde λ es una constante.

Probar que la distribución de probabilidad de X_n tiende a la distribución de Poisson con parámetro λ, en el sentido de que

$$P(X_n = k) \stackrel{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}e^{-\lambda} \quad \text{para } k = 0, 1, 2, \dots$$

7. Se observan n repeticiones independientes de un juego de ruleta. Hallar n para que, con probabilidad menor que 0,1, la frecuencia relativa del número de veces que ocurre la primera docena sea mayor que 1/3. Hacer el cálculo de dos maneras:

   a. Usando la desigualdad de Tchebichev.

   b. Usando la aproximación por la distribución normal.

   Nota. Los resultados posibles en un juego de ruleta son 0, 1, 2,..., 36 y la primera docena consta de los números 1, 2,..., 12.

8. Consideremos el espacio $\Omega = [0, 1)$ dotado de la probabilidad uniforme. Recordemos que cada número $x \in [0, 1)$ tiene una escritura decimal

$$x = 0, x_1x_2x_3 \dots \tag{7.37}$$

   donde los x_i son dígitos, es decir, enteros comprendidos entre 0 y 9 (para $i = 1, 2, \dots$).

   Recordemos también, que el significado de (7.40) no es otro que

$$x = \sum_{i=1}^{\infty} \frac{x_i}{10^i} \tag{7.38}$$

   Cuando la sucesión $\{x_i\}$ es tal que $x_i = 0$ para todo $i \geq i_0$, para alguna representación de x, se dice que x es un número decimal.

   a. Probar que la representación decimal (7.40) es única, salvo para los números decimales, que tienen dos representaciones. (En este último caso elegimos como representación aquella que tiene ceros de un índice en adelante).

   b. Consideremos las variables aleatorias $X_1, X_2, \dots$ definidas en Ω de la siguiente manera: si x tiene la representación (7.40),

$$X_i(x) = x_i \quad \text{para } i = 1, 2, \dots$$

   Probar que esta es una sucesión de variables aleatorias independientes, cada una con la distribución

$$P(X_i = k) = \frac{1}{10} \quad k = 0, 1, \dots, 9.$$

   c. Sea $S_n(x)$ el número de veces que aparece un cierto dígito (por ejemplo el 7) entre los primeros n decimales de la representación de x. Muestre que, salvo para los $x \in [0, 1)$ en un conjunto de probabilidad nula, se cumple que

$$\frac{S_n(x)}{n} \to \frac{1}{10} \quad \text{cuando } n \to \infty.$$

   Es decir que, salvo para estos x excepcionales (pruebe que son un conjunto no-numerable) la proporción del número de veces que aparece el dígito "7" tiende a 1/10. Lo mismo pasa, obviamente, con cualquier otro dígito.

9. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con igual distribución F. Llamamos "distribución empírica de la muestra $X_1, X_2, \ldots, X_n$" a la función $F_n : \mathbb{R} \to [0,1]$ definida por

$$F_n(x) = \frac{1}{n}(\sharp\{i : X_i \leq x\})$$

(observe que, para cada x, $F_n(x)$ es una variable aleatoria).

a. Representar gráficamente la función F_n y mostrar que ella es, como función de x, una función de distribución con saltos puros.

b. Probar que, para cada $x \in \mathbb{R}$, se cumple

$$F_n(x) \xrightarrow{(P)} F(x) \quad \text{cuando } n \to \infty.$$

c. Mostrar que también, para cada $x \in \mathbb{R}$

$$F_n(x) \longrightarrow F(x) \quad \text{casi seguramente cuando } n \to \infty.$$

10. En un cálculo numérico mediante una computadora, 1.000 números se reemplazan por el entero más próximo. Si los errores de redondeo cometidos son independientes y la distribución de probabilidad de cada uno es uniforme en el intervalo $\left(-\frac{1}{2}, \frac{1}{2}\right)$, calcular aproximadamente la probabilidad de que la suma de los 1.000 números redondeados difiera de la suma de los 1.000 números originales en menos de 10.

11. Al lanzar una moneda 10.000 veces se obtuvieron 5.500 caras. ¿Sospecharía Ud. que la moneda no es equilibrada?

12. Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes con igual distribución, y supongamos que

$$\sigma^2 = \mathrm{Var}(X_n) < \infty \quad n = 1, 2, \ldots$$

a. Calcular el límite en probabilidad de

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_{i+1} - X_i)^+.$$

b. En particular, indicar cuál es el resultado en a. cuando la distribución común de las X_i es la normal $(0, 1)$.

Sugerencia: No podemos aplicar directamente la ley débil de los grandes números. ya que las variables $\{(X_{i+1} - X_i)^+\}$ no forman una sucesión de variables aleatorias independientes. Sin embargo, si $i + 1 < j$, $(X_{i+1} - X_i)^+$ y $(X_{j+1} - X_j)^+$ si resultan independientes. Aproveche esta observación para acotar $\mathrm{Var}(Y_n)$ y probar que tiende a cero cuando $n \to \infty$.

Apéndice

Proposición 7.2 (Propiedad 1) *Sea $\{Y_n\}$ una sucesión de variables aleatorias que converge casi seguramente a la variable aleatoria Y. Entonces Y_n converge a Y en probabilidad.*

Observación 7.3 El recíproco de esta propiedad es falso. Ver ejercicio 5 de este capítulo.

Demostración Sea ε un número positivo cualquiera. Queremos probar que

$$P(|Y_n - Y| \geq \varepsilon) \to 0 \quad \text{cuando } n \to \infty. \tag{7.39}$$

Definimos los eventos

$$A_n = \{\omega : |Y_n(\omega) - Y(\omega)| \geq \varepsilon\}$$
$$\overline{A} = \overline{\lim_{n\to\infty}} A_n = \{\omega : \omega \in A_n \text{ para infinitos valores de } n\}$$

(este conjunto suele llamarse el *límite superior* de la sucesión de conjuntos $\{A_n\}$).

Es sencillo verificar que las siguientes proposiciones son ciertas.

a. $\overline{A} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$.

b. Si llamamos $B_m = \bigcup_{n=m}^{\infty} A_n$, entonces la sucesión de eventos $\{B_m\}$ es decreciente y $\overline{A} = \bigcap_{m=1}^{\infty} B_m$.

Observamos que $\overline{A} \subset \{\omega : Y_n(\omega) \text{ no tiende a } Y(\omega)\}$, porque si $\omega \in \overline{A}$, entonces, de acuerdo a la definición de $\overline{A}$

$$|Y_n(\omega) - Y(\omega)| \geq \varepsilon$$

para infinitos índices n, y esto implica que $Y_n(\omega)$ no tiende a $Y(\omega)$, según la definición del límite de las sucesiones de números reales.

Entonces, teniendo en cuenta que $Y_n \to Y$ casi seguramente,

$$P(\overline{A}) \leq P(\omega : Y_n(\omega) \text{ no tiende a } Y(\omega)) = 0$$

y en consecuencia $P(\overline{A}) = 0$.

Por otra parte, de las propiedades generales de los espacios de probabilidad, como la sucesión de eventos $\{B_n\}$ es decreciente, se tiene

$$\lim_{m\to\infty} P(B_m) = P(\cap_{m=1}^{\infty} B_m) = P(\overline{A}) = 0.$$

Finalmente, es claro de la definición de B_m que

$$A_m \subset B_m \Rightarrow P(A_m) \leq P(B_m) \Rightarrow \lim P(A_m) = 0.$$

que es justamente (7.39). ■

Teorema 7.5 (Propiedad 2. (Lema de Borel-Cantelli).) *Sea $\{A_n\}$ una sucesión de eventos en un espacio de probabilidad. Entonces*

1. $\sum_n P(A_n) < \infty \Rightarrow P(\overline{A}) = 0$.

2. $$\left.\begin{array}{l} \sum_n P(A_n) = +\infty \\ A_1, A_2, \ldots \text{ es una sucesión de eventos independientes} \end{array}\right\} \Rightarrow P(\overline{A}) = 1$$

Demostración Veamos la primera parte. Sabemos que

$$\overline{A} = \bigcap_{m=1}^{\infty} B_m \quad \text{con} \quad B_m = \bigcup_{n=m}^{\infty} A_n.$$

Entonces

$$P(B_m) \leq \sum_{n=m}^{\infty} P(A_n)$$

y como la serie $\sum_n P(A_n)$ es convergente, la cola de la serie que figura en el segundo miembro de esta desigualdad tiende a cero cuando $m \to \infty$. Por lo tanto,

$$P(B_m) \to 0 \quad \text{cuando } m \to \infty.$$

Pero, por otra parte, $\overline{A} \subset B_m$ para todo $m = 1, 2, \ldots$ implica que $P(\overline{A}) \leq P(B_m)$ para todo $m = 1, 2, \ldots$, y en consecuencia

$$P(\overline{A}) \leq \lim_{m\to\infty} P(B_m) \Rightarrow P(\overline{A}) = 0.$$

Esto prueba la primera parte.

Para la segunda parte basta ver que $P(\overline{A}^c) = 0$. En virtud de la reglas de De Morgan

$$\overline{A}^c = \bigcup_{m=1}^{\infty} B_m^c, \qquad B_m^c = \bigcap_{n=m}^{\infty} A_n^c.$$

Entonces $P(\overline{A}^c) \leq \sum_{m=1}^{\infty} P(B_m^c)$, y para probar (2) basta con probar

$$P(B_m^c) = 0 \quad \text{para todo } m = 1, 2, \ldots \tag{7.40}$$

Llamemos

$$C_{m,p} = \bigcap_{n=m}^{m+p} A_n^c.$$

Es claro que

$$B_m^c \subset C_{m,p} \quad \text{para todo } p = 1, 2, \ldots$$
$$P(B_m^c) \leq P(C_{m,p}) \quad \text{para todo } p = 1, 2, \ldots$$

Ahora bien, como los $\{A_n\}$ forman una sucesión de eventos independientes, también la sucesión $\{A_n^c\}$ es de eventos independientes, y por lo tanto

$$P(C_{m,p}) = P(\bigcap_{n=m}^{m+p} A_n^c) = \prod_{n=m}^{m+p} P(A_n^c) = \prod_{n=m}^{n+p} (1 - P(A_n)). \tag{7.41}$$

A estas alturas, aplicamos la desigualdad

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

válida para todo x real, y que puede ser demostrada por métodos elementales de cálculo.

Como además, todos los factores en el último miembro de (7.41) son no-negativos, resulta

$$P(C_{m,p}) \leq \prod_{n=m}^{m+p} e^{-P(A_n)} = \exp\{-\sum_{n=m}^{m+p} P(A_n)\}$$

Hagamos tender $p \to +\infty$ en esta desigualdad. Como la serie $\sum P(A_n)$ es divergente, el exponente de la derecha tiende a $-\infty$ y la exponencial a cero. Esto prueba que

$$P(B_m^c) \leq \lim_{p\to\infty} P(C_{m,p}) = 0$$

y de aquí que $P(B_m^c) = 0$. Esto es (7.40) y termina la demostración de la segunda parte.

Teorema 7.6 (Ley Fuerte de los Grandes Números) *Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes tal que*

$$\mathrm{E}(X_n) = \mu, \qquad \mathrm{E}(X_n^4) \leq C \quad n = 1, 2, \dots$$

donde C es una constante. Entonces,

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

converge a μ casi seguramente.

Demostración Queremos probar que el evento

$$\{\omega : Y_n(\omega) \text{ no tiende a } \mu\}$$

tiene probabilidad nula. Observamos que decir que Y_n no tiende a μ es lo mismo que decir que existe un número de la forma $1/k$, k entero positivo, tal que

$$|Y_n - \mu| \geq \frac{1}{k}$$

para infinitos valores de n. Es decir que se tiene la igualdad entre eventos

$$\{Y_n \text{ no tiende a } \mu\} = \bigcup_{k=1}^{\infty} \{|Y_n - \mu| \geq \frac{1}{k} \text{ para infinitos valores de } n\}$$

que implica

$$P(Y_n \text{ no tiende a } \mu) \leq \sum_{k=1}^{\infty} P(|Y_n - \mu| \geq \frac{1}{k} \text{ para infinitos valores de } n). \tag{7.42}$$

Para probar que el primer miembro vale cero, alcanza con probar que cada sumando del segundo miembro es igual a cero. Para esto, introducimos algo más de notación. k estará fijo en lo que sigue.

$$A_{n,k} = \{|Y_n - \mu| \geq \frac{1}{k}\}.$$

Observamos que el k-ésimo sumando del segundo miembro de la desigualdad (7.42) es la probabilidad de

$$\overline{\lim_{n\to\infty}} A_{n,k} = \overline{A}_k = \{\omega : \omega \in A_{n,k} \text{ para infinitos valores de } n\}.$$

Para demostrar que $P(\overline{A}_k) = 0$ usaremos el lema de Borel-Cantelli (propiedad 2 de la sección 7.3). Según éste, basta demostrar que la serie

$$\sum_n P(A_{n,k})$$

es convergente. Para ello acotaremos cada término por el término respectivo de una serie convergente, de la siguiente forma

$$P(A_{n,k}) = P(|Y_n - \mu| \geq \frac{1}{k}) = P((Y_n - \mu)^4 \geq \frac{1}{k^4}) \leq k^4\,\mathrm{E}((Y_n - \mu)^4).$$

La última desigualdad resulta directamente de aplicar la desigualdad de Markov.

Introducimos aún algo de notación:

$$\begin{aligned} Y_n - \mu &= \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{n} \\ &= \frac{(X_1 - \mu) + (X_2 - \mu) + \cdots + (X_n - \mu)}{n} \end{aligned}$$

y ahora poniendo $Z_i = X_i - \mu$ obtenemos

$$Y_n - \mu = \frac{Z_1 + Z_2 + \cdots + Z_n}{n}.$$

Entonces es claro que $\mathrm{E}(Z_n) = \mathrm{E}(Y_n) - \mu = 0$ y, por otra parte, desarrollando la cuarta potencia de Z_n,

$$\mathrm{E}(Z_n^4) = \mathrm{E}(X_n^4) - 4\mu\,\mathrm{E}(X_n^3) + 6\mu^2\,\mathrm{E}(X_n^2) - 4\mu^3\,\mathrm{E}(X_n) + \mu^4$$

y esta expresión está acotada por una constante fija, ya que, si los cuartos momentos de X_n están acotadas (por hipótesis), también lo están los momentos de orden menor, por la propiedad 7 de los valores esperados (ver sección 6.2).

Entonces, la hipótesis implica que, para todo $n = 1, 2, \dots,$

$$\mathrm{E}(Z_n^4) \leq C_1$$

donde C_1 es una constante. Podemos ahora acotar

$$\begin{aligned}
\mathrm{E}((Y_n - \mu)^4) &= \mathrm{E}\left(\left(\frac{Z_1 + Z_2 + \cdots + Z_n}{n}\right)^4\right) \\
&= \frac{1}{n^4}\,\mathrm{E}\left(\left(\sum_{i=1}^n Z_i\right)^4\right) \\
&= \frac{1}{n^4}\,\mathrm{E}\left(\sum_{i=1}^n Z_i \sum_{j=1}^n Z_j \sum_{k=1}^n Z_k \sum_{l=1}^n Z_l\right) \\
&= \frac{1}{n^4} \sum_{i,j,k,l} \mathrm{E}(Z_i Z_j Z_k Z_l).
\end{aligned}$$

Ahora tenemos que estudiar como son los términos $\mathrm{E}(Z_i Z_j Z_k Z_l)$ y para ello hay que considerar todos las cuadruplas (i, j, k, l) de enteros entre 1 y n.

Primero, si en la cuadrupla hay un elemento que no está repetido, entonces esta esperanza vale cero. En efecto, si el que no está repetido es el i (el mismo argumento vale si es algún otro índice), entonces

$$\mathrm{E}(Z_i Z_j Z_k Z_l) = \mathrm{E}(Z_i)\,\mathrm{E}(Z_j Z_k Z_l) = 0.$$

La primera igualdad se debe a que Z_i y $Z_j Z_k Z_l$ son independientes, y la segunda a que $\mathrm{E}(Z_i) = 0$.

Segundo, nos quedan entonces solamente los términos en los que no hay índices solos, y estos son de dos tipos: aquellos términos en los que los cuatro índices son iguales $i = j = k = l$ y aquellos en los que hay dos parejas de índices iguales, distintas entre si, por ejemplo, $i = j = 1; k = l = 2$. Más precisamente, resulta

$$\mathrm{E}((Y_n - \mu)^4) = \frac{1}{n^4}\left[\sum_{i=1}^n \mathrm{E}(Z_i^4) + \binom{4}{2} \sum_{1 \leq i < j \leq n} \mathrm{E}(Z_i^2 Z_j^2)\right]$$

El factor $\binom{4}{2}$ corresponde a todas las maneras de tener en la cuadrupla dos factores con índice i y dos con índice j. Todavía, para acotar estas sumas, usamos las desigualdades

$$\mathrm{E}(Z_i^4) \leq C_1$$

$$\mathrm{E}(Z_i^2 Z_j^2) \leq \sqrt{\mathrm{E}(Z_i^4)\ \mathrm{E}(Z_j^4)} \leq \sqrt{C_1 C_1} = C_1$$

donde la primera desigualdad de la última línea resulta de aplicar Cauchy-Schwarz. En definitiva

$$\mathrm{E}((Y_n - \mu)^4) \leq \frac{1}{n^4}\left(nC_1 + 6\frac{n(n-1)}{2}C_1\right) \leq \frac{C_2}{n^2}.$$

Aquí $n(n-1)/2$ es el número de términos en la suma doble y C_2 es una nueva constante.

Y ahora, reemplazando en (7.16) obtenemos

$$P(A_{n,k}) \leq \frac{C_2 k^4}{n^2},$$

y como la serie $\sum \frac{1}{n^2}$ es convergente, se deduce que $\sum_n P(A_{n,k})$ también lo es. Es lo que queríamos probar. ■

7.6.1. El Teorema Central del Límite: La Condición de Lindeberg.

El problema del estudio del comportamiento asintótico de una suma de variables aleatorias independientes cuando el número de sumandos tiende a infinito, es uno de los problemas clásicos de gran importancia en la teoría de probabilidades. Agreguemos, a título complementario, el enunciado de un teorema general que asegura, bajo ciertas condiciones, la convergencia a la distribución normal de la distribución de una suma de variables aleatorias independientes.

Teorema 7.7 (Lindeberg) *Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes con*

$$\mathrm{E}(X_n) = 0, \quad \sigma_n^2 = \mathrm{E}(X_n^2) < \infty \quad \textit{para } n = 1, 2, \dots$$

y pongamos $S_n = X_1 + X_2 + \cdots + X_n$. Es claro que

$$\mathrm{E}(S_n) = 0 \qquad s_n^2 \equiv \mathrm{Var}(S_n) = \sum_{i=1}^{n} \sigma_i^2.$$

Diremos que se cumple la condición de Lindeberg si para todo $\varepsilon > 0$

$$\frac{1}{s_n^2} \sum_{i=1}^{n} \mathrm{E}(X_i^2 \mathbf{1}_{A_{i,n}}) \overset{n\to\infty}{\longrightarrow} 0$$

donde $\mathbf{1}_{A_{i,n}}$ es la función indicatriz del evento

$$A_{i,n} = \{|X_i| > \varepsilon s_n\}.$$

Bajo esta condición se tiene

$$P\left(\frac{S_n}{s_n} \leq x\right) \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}.$$

Observamos que la condición $\mathrm{E}(X_n) = 0$ no tiene ninguna importancia restrictiva, ya que si $\mathrm{E}(X_n) \neq 0$, consideramos $X_n' = X_n - \mathrm{E}(X_n)$, y entonces $E(X_n') = 0$.

Para la demostración, el lector puede consultar los textos de M. Loève, W. Feller (Vol. 2), P. Billingsley o R. M. Dudley incluidos en la bibliografía.

Un caso de suma importancia en el que se cumple la condición de Lindeberg es aquel en el cual $X_1, X_2, \dots$ tienen todas la misma distribución. Tenemos

Teorema 7.8 *Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes con igual distribución F. Supongamos que*

$$\mathrm{E}(X_i) = \mu \qquad \mathrm{Var}(X_i) = \sigma^2 \quad \textit{para } n = 1, 2, \dots$$

Entonces

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}.$$

No es difícil probar que la sucesión de variables aleatorias independientes $X_n' = X_n - \mu$ cumple la condición de Lindeberg, y por lo tanto, la condición del teorema 7.7. Es claro además, que el teorema de De-Moivre - Laplace es un caso particular del teorema 7.8 (y como consecuencia, del 7.7). Basta considerar como distribución común F a la variables independientes $X_1, X_2, \dots$ la definida por

$$P(X_n = 1) = p, \qquad P(X_n = 0) = 1 - p.$$