5

Convergencia de Variables Aleatorias

5.1. Funciones Generadoras de Momentos

Dada una variable aleatoria X, o su función de distribución F, vamos a definir otra función generadora, como

$$M_X(t) = \mathrm{E}(e^{tX}).$$

siempre que este valor esperado exista.

Notemos que cuando X toma valores en los enteros no-negativos, $M_X(t) = \phi_X(e^t)$, donde ϕ_X es la f.g.p. de la variable X. Si X está acotada, M_X está bien definida para todo t real; en cambio, si X no está acotada, es posible que el dominio de M no sea el conjunto de todos los reales. En todo caso, p siempre está definida en cero, y $M(0) = 1$.

Es posible demostrar que si la f.g.m. de la v.a. X existe en un entorno de 0, entonces para todo $k > 0$,

$$\mathrm{E}(|X|^k) < \infty.$$

Más aún, la serie

$$M_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}\big(1 + \sum_{k=1}^{\infty} \frac{t^k X^k}{k!}\big) = 1 + \sum_{n=1}^{\infty} \frac{t^k}{k!}\,\mathrm{E}(X^k) \tag{5.1}$$

es convergente y se puede derivar término a término. Obtenemos

$$M'_X(0) = \mathrm{E}(X); \quad M''_X(0) = \mathrm{E}(X^2)$$

y en general

$$M_X^{(k)}(0) = \mathrm{E}(X^k).$$

Es por esta última propiedad que esta función se conoce como *función generadora de momentos* (f.g.m.).

Ejemplos 5.1

1. Si $X \sim \mathcal{U}(a, b)$, su densidad es $f(x) = 1/(b-a)$ para $a < x < b$ y

$$M(t) = \int_a^b \frac{e^{tx}}{b-a}dx = \frac{e^{tb} - e^{ta}}{t(b-a)} \tag{5.2}$$

En el caso particular de la distribución uniforme en $(0, 1)$ tenemos $M(t) = (e^t - 1)/t$.

Si queremos derivar (5.2) para hallar los momentos de esta distribución, tenemos que aplicar la regla de L'Hôpital repetidas veces. Una alternativa es usar el desarrollo de la función exponencial en serie de potencias y juntar los términos de igual orden, apoyándonos en la existencia de la f.g.m.:

$$\begin{aligned} M(t) &= \frac{1}{t(b-a)}\left(e^{tb}-e^{ta}\right) = \frac{1}{t(b-a)}\left(1+\sum_{n=1}^{\infty}\frac{(tb)^n}{n!}-1-\sum_{n=1}^{\infty}\frac{(ta)^n}{n!}\right) \\ &= \frac{1}{b-a}\left(\sum_{n=1}^{\infty}\frac{(b^n-a^n)}{n!}t^{n-1}\right) \end{aligned}$$

Este es el desarrollo de MacLaurin para $M(t)$ en 0 y por lo tanto, necesariamente,

$$M^{(k)}(0) = \frac{k!}{(k+1)!}\frac{b^{k+1}-a^{k+1}}{b-a} = \frac{b^{k+1}-a^{k+1}}{(k+1)(b-a)} \tag{5.3}$$

Vemos que

$$M'(0) = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

mientras que

$$M''(0) = \frac{b^3-a^3}{3(b-a)} = \frac{a+ab+b^2}{3}$$

y un cálculo sencillo muestra que

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{(a+b)^2}{12}.$$

2. Si $X \sim Bin(n,p)$ veamos que $M(t) = (pe^t+1-p)^n$): Un cálculo directo muestra que

$$M(t) = \sum_{j=0}^{n} e^{jt}\binom{n}{j}p^j(1-p)^{n-j} = (pe^t+1-p)^n,$$

.

Otra manera de obtener este resultado es recordar que $\phi_X(s) = (ps+q)^n$ y utilizar

$$M_X(t) = \phi_X(e^t) = (pe^t+q)^n$$

Derivando es fácil obtener que

$$\mathrm{E}(X) = M'(0) = np, \qquad \mathrm{E}(X^2) = M''(0) = npq + (np)^2$$

y en consecuencia $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = npq$.

3. Si $X \sim \mathcal{E}xp(\lambda)$, es decir, si $P(X \le x) = 1 - e^{-\lambda x}$, para $x \ge 0$, entonces $M(t) = \lambda/(\lambda - t)$ para $t \le \lambda$.

El resultado se obtiene a partir del cálculo

$$M(t) = \int_0^{\infty} \lambda e^{-\lambda x}\, e^{tx}dx = \lambda\, \frac{e^{(t-\lambda)x}}{t-\lambda}\bigg|_0^{\infty} = \frac{\lambda}{\lambda-t}.$$

Observamos que en este caso, $M(t)$ no está definida si $t \ge \lambda$. Para $t < \lambda$

$$\frac{\lambda}{\lambda-t} = \frac{1}{1-\frac{t}{\lambda}} = \sum_{k=0}^{\infty}\left(\frac{t}{\lambda}\right)^k$$

y usando la ecuación (5.1) obtenemos los momentos de X:

$$\mathrm{E}(X^k) = \lambda^{-k}k!, \qquad k \ge 1.$$

4. Si $X \sim \mathcal{N}(0,1)$, es decir, si $P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$, entonces $M(t) = e^{t^2/2}$.

   Calculemos

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \, e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} \, e^{t^2/2} \mathrm{d}x = e^{t^2/2}$$

   ya que $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} \mathrm{d}x = 1$ puesto que el integrando es la densidad de una variable aleatoria con distribución $\mathcal{N}(t,1)$

   Recordando el desarrollo en serie de potencias de la función exponencial obtenemos

$$M(t) = e^{t^2/2} = \sum_{n=1}^{\infty} \frac{t^{2n}}{2^{2n} n!} = \sum_{n=1}^{\infty} \frac{(2n)!}{2^{2n} n!} \frac{t^{2n}}{(2n)!}$$

   Comparando con (5.1) obtenemos que los momentos de $X \sim \mathcal{N}(0,1)$ están dados por

$$\mathrm{E}(X^n) = \begin{cases} 0 & \text{si } n \text{ es impar,} \\ \frac{n!}{(n/2)!} 2^{-n/2} & \text{si } n \text{ es par.} \end{cases}$$

5. Sea X una variable aleatoria con distribución de Pareto de índice $\alpha > 0$, con densidad dada por

$$f(x) = \frac{\alpha}{x^{1+\alpha}} \quad \text{para } x > 1$$

   y 0 en otro caso. Es fácil ver que esta variable no tiene f.g.m. ya que

$$\lim_{x \to \infty} \frac{\alpha e^{tx}}{x^{1+\alpha}} = \infty$$

   y por lo tanto la integral que define la f.g.m.

$$\int_1^{\infty} f(x) e^{tx} \, dx = \int_1^{\infty} \frac{\alpha}{x^{1+\alpha}} e^{tx} \, dx$$

   no converge.

▲

El último ejemplo que hemos visto muestra que no siempre existe la f.g.m. Una condición necesaria para su existencia es que la variable X tenga momentos finitos de todos los órdenes, pero esta condición no es suficiente

Sea X una v.a. con f.g.m. M_X y sea $Y = aX + b$ una transformación lineal de X, entonces

$$M_Y(t) = \mathrm{E}(e^{tY}) = \mathrm{E}(e^{t(aX+b)}) = \mathrm{E}(e^{taX} e^{tb}) = e^{tb} \, \mathrm{E}(e^{taX}) = e^{tb} M_X(at) \tag{5.4}$$

Ejemplo 5.2
Si $X \sim \mathcal{N}(0,1)$ sabemos que $Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$. Por (5.4)

$$M_Y(t) = e^{t\mu} M_X(\sigma t) = \exp(\mu t + (\sigma t)^2/2).$$

▲

Observación 5.1 Por la forma en la cual hemos definido la función generadora de momentos, cuando las f.g.m. de dos variables aleatorias X_1, X_2 coinciden para todos los valores de t en un entorno de $t = 0$, entonces las distribuciones de probabilidad de X_1 y X_2 deben ser idénticas. Este resultado lo enunciamos en el próximo teorema, sin demostración

Teorema 5.1 *Si X tiene función generadora de momentos $M(t)$ que está definida en un entorno $(-a, a)$ de 0, entonces $M(t)$ caracteriza a la distribución de X, es decir, si otra variable Y tiene la misma función generadora de momentos, las distribuciones de X e Y coinciden.*

Si X, Y son v.a.i. con f.g.m. respectivas M_x y M_Y que existen en un dominio común $|t| < d$ entonces la f.g.m. de la suma $X + Y$ está dada por

$$M_{X+Y}(t) = \mathrm{E}[e^{t(X+Y)}] = \mathrm{E}[e^{tX}e^{tY}] = \mathrm{E}[e^{tX}]\ \mathrm{E}[e^{tY}] = M_X(t)M_Y(t) \tag{5.5}$$

para $|t| < d$. Este resultado se extiende a la suma de n variables aleatorias independientes: Si $S_n = X_1 + \cdots + X_n$,

$$M_{S_n}(t) = \prod_{i=1}^{n} \mathrm{E}(e^{tX_i}) = \prod_{i=1}^{n} M_{X_i}(t).$$

Ejemplo 5.3
Consideremos n v.a.i. con distribución de Poisson de parámetros λ_i, respectivamente. Tenemos

$$\begin{aligned} M_{X_i}(t) = \mathrm{E}(e^{tX_i}) &= \sum_{k=0}^{\infty} e^{tk}\frac{\lambda_i^k}{k!}e^{-\lambda_i} \\ &= e^{-\lambda_i}\sum_{k=0}^{\infty}\frac{(\lambda_i e^t)^k}{k!} \\ &= e^{\lambda_i(e^t-1)} \end{aligned}$$

En consecuencia

$$M_{S_n}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{(\sum_{i=1}^{n}\lambda_i(e^t-1))}$$

y por el teorema 5.1 vemos que S_n tiene distribución de Poisson de parámetro $\sum_1^n \lambda_i$.

▲

5.1.1. Teorema de Continuidad

La función generadora de momentos resulta particularmente útil cuando consideramos sucesiones de variables aleatorias, como lo muestra el siguiente teorema que enunciamos sin demostración.

Teorema 5.2 (de Continuidad) *Sea $F_n(x)$, $n \geq 1$ una sucesión de f.d. con funciones generadores de momento respectivas $M_n(t)$, $n \geq 1$, que están definidas para $|t| < b$. Supongamos que cuando $n \to \infty$, $M_n(t) \to M(t)$ para $|t| \leq a < b$, donde $M(t)$ es la función generadora de momentos de la distribución $F(x)$. Entonces $F_n(x) \to F(x)$ cuando $n \to \infty$ para todo punto x en el cual F es continua.*

Ejemplo 5.4
Sea Y_n una v.a. con distribución uniforme en los enteros $\{1, 2, \dots, n\}$. Hallemos la f.g.m. de Y_n/n:

$$\begin{aligned}
M_n(t) &= \mathrm{E}[e^{tY_n/n}] = \frac{1}{n}\big(e^{t/n} + e^{2t/n} + \cdots + e^{nt/n}\big) \\
&= \frac{1}{n}\Big(\frac{e^{t/n} - e^{(n+1)t/n}}{1 - e^{t/n}}\Big) \\
&= \frac{1}{n}\Big(\frac{1 - e^t}{e^{-t/n} - 1}\Big) \\
&= \frac{1 - e^t}{t - \frac{t^2}{2n} + \frac{t^3}{6n^2}} \\
&\to \frac{1 - e^t}{t} \quad \text{cuando } n \to \infty,
\end{aligned}$$

que es la f.g.m. de una distribución uniforme en $(0, 1)$. Por lo tanto, para $0 \leq x \leq 1$

$$P(\frac{Y_n}{n} \leq x) \to P(U \leq x) = x$$

donde U tiene distribución $\mathcal{U}(0, 1)$. ▲

Ejemplo 5.5
Sea Y_n una v.a. con distribución geométrica con parámetro $p = \lambda/n$. La f.g.p. de esta variables está dada por

$$\phi_{Y_n}(s) = \mathrm{E}(s^{Y_n}) = \sum_{k=1}^{\infty} q^{k-1} p s^k = \frac{ps}{1 - qs}$$

La f.g.p. de Y_n/n es

$$\phi_n(s) = \mathrm{E}(s^{Y_n/n}) = \mathrm{E}((s^{1/n})^{Y_n}) = \phi_{Y_n}(s^{1/n}) = \frac{ps^{1/n}}{1 - qs^{1/n}}$$

Recordemos que si la f.g.p. de X es $\phi_X(s)$ entonces la f.g.m. está dada por $M(t) = \phi_X(e^t)$. Por lo tanto, la f.g.m. de Y_n/n está dada por

$$\begin{aligned}
M_n(t) = \phi_n(e^t) &= \frac{pe^{t/n}}{1 - qe^{t/n}} \\
&= \frac{\lambda}{n} e^{t/n}\Big(1 - \big(1 - \frac{\lambda}{n}\big)e^{t/n}\Big)^{-1}
\end{aligned}$$

Recordemos que $e^x = 1 + x + \mathcal{O}(x^2)$ cuando $x \to 0$. Desarrollando las exponenciales en la expresión anterior

$$\begin{aligned}
M_n(t) &= \frac{\lambda}{n}\big(1 + \frac{t}{n} + \mathcal{O}\big((\frac{t}{n})^2\big)\Big(1 - \big(1 - \frac{\lambda}{n}\big)\big(1 + \frac{t}{n} + \mathcal{O}\big((\frac{t}{n})^2\big)\Big)^{-1} \\
&= \frac{\lambda}{n}\Big(1 + \mathcal{O}\big(\frac{1}{n}\big)\Big)\Big(\frac{\lambda - t}{n} + \mathcal{O}\big(\frac{1}{n^2}\big)\Big)^{-1} \\
&\to \frac{\lambda}{\lambda - t}
\end{aligned}$$

que es la f.g.m. de una variable exponencial con parámetro λ. Por lo tanto

$$P\Big(\frac{X_n}{n} \leq x\Big) \to 1 - e^{-\lambda x}$$

para $x > 0$. ▲

Veamos una aplicación del teorema anterior para demostrar el Teorema de de Moivre y Laplace.

Teorema 5.3 (de Moivre-Laplace) *Sea $S_n \sim Bin(n,p)$ para $n \geq 1$ y $q = 1-p$. Definimos*

$$T_n = \frac{S_n - np}{(npq)^{1/2}}$$

Entonces para todo $x \in \mathbb{R}$,

$$P(T_n \leq x) \to \Phi(x) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{-x^2/2} dx.$$

Demostración. Recordemos que S_n es la suma de n v.a.i. con distribución de Bernoulli de parámetro p: $S_n = \sum_1^n X_i$. Usamos esto para calcular la función generadora de momentos de T_n.

$$\begin{aligned}
\mathrm{E}(e^{tT_n}) &= \mathrm{E}\Big[\exp\Big(\frac{t(S_n - np)}{(npq)^{1/2}}\Big)\Big] = \mathrm{E}\Big[\exp\Big(\frac{t(\sum_1^n (X_i - p))}{(npq)^{1/2}}\Big)\Big] \\
&= \mathrm{E}\Big[\prod_{i=1}^{n} \exp\Big(\frac{t(X_i - p)}{(npq)^{1/2}}\Big)\Big] = \prod_{i=1}^{n} \mathrm{E}\Big[\exp\Big(\frac{t(X_i - p)}{(npq)^{1/2}}\Big)\Big] \\
&= \Big(E\Big[\exp\Big(\frac{t(X_1 - p)}{(npq)^{1/2}}\Big)\Big]\Big)^n \\
&= \Big(p \exp\Big(\frac{t(1-p)}{(npq)^{1/2}}\Big) + q \exp\Big(\frac{-pt}{(npq)^{1/2}}\Big)\Big)^n. \qquad (5.6)
\end{aligned}$$

Ahora hacemos un desarrollo de Taylor para las dos exponenciales que aparecen en esta última expresión para obtener

$$p \exp\Big(\frac{t(1-p)}{(npq)^{1/2}}\Big) = p\Big(1 + \frac{qt}{(npq)^{1/2}} + \frac{q^2t^2}{2npq} + \frac{C_1 q^3 t^3}{3!(npq)^{3/2}}\Big) \qquad (5.7)$$

$$q \exp\Big(\frac{-pt}{(npq)^{1/2}}\Big) = q\Big(1 - \frac{pt}{(npq)^{1/2}} + \frac{p^2t^2}{2npq} + \frac{C_2 p^3 t^3}{3!(npq)^{3/2}}\Big). \qquad (5.8)$$

La suma de estas dos expresiones nos da $1 + \frac{t^2}{2n} + O(n^{-3/2})$ y sustituyendo en (5.6) obtenemos

$$\mathrm{E}(e^{tT_n}) = \big(1 + \frac{t^2}{2n} + O(n^{-3/2})\big)^n \to e^{t^2/2}$$

que es la f.g.m. de la distribución normal típica. ∎

5.1.2. Sumas de Variables Aleatorias y F.G.M.

Recordemos que si X es una v.a. con valores en los enteros no negativos $0, 1, 2, \dots$ definimos la función generadora de probabilidad de X por

$$\phi_X(s) = \mathrm{E}(s^X) = \sum_{k=0}^{\infty} s^k P(X = k).$$

Si tenemos ahora una sucesión de v.a.i. X_i con la misma distribución que X y otra v.a. N, independiente de ellas y con valores en $0, 1, 2, \dots$, la f.g.p. de la suma aleatoria $T = \sum_{i=1}^N X_i$, donde la suma vale 0 si $N = 0$, tiene f.g.m. dada por

$$\phi_T(s) = \mathrm{E}(s^T) = \phi_N(\phi_X(s)),$$

es decir, es la composición de las f.g.p. de N y X.

Supongamos ahora que X tiene distribución con f.g.m. $M_X(t) = \mathrm{E}(e^{tX})$ y consideramos, al igual que antes, la suma aleatoria $T = \sum_{i=1}^N X_i$ con la misma convención de que $T = 0$ si $N = 0$. ¿Cuál es ahora la f.g.m. de T? Tenemos

$$\begin{aligned} M_T(t) &= \mathrm{E}(e^{tT}) = \mathrm{E}(e^{t\sum_1^N X_i}) \\ &= \sum_{n=1}^{\infty} \mathrm{E}\left(e^{t\sum_1^n X_i}\right) P(N = n) \\ &= \sum_{n=1}^{\infty} M_X^n(t) P(N = n) \\ &= \phi_N(M_X(t)) \end{aligned}$$

que es un resultado similar al anterior, cambiando la f.g.p. de X por la f.g.m.

Ejemplo 5.6
Supongamos que X_i tiene distribución exponencial de parámetro λ y N es geométrica con parámetro p. Hemos visto que

$$M_X(t) = \frac{\lambda}{\lambda - t} \qquad \text{y} \qquad \phi_N(s) = \frac{ps}{1 - qs}.$$

Usando el resultado que acabamos de demostrar obtenemos

$$M_T(t) = \frac{pM_X(t)}{1 - qM_X(t)} = \frac{p\lambda}{\lambda - t - q\lambda} = \frac{p\lambda}{p\lambda - t}.$$

Esta última expresión corresponde a una distribución exponencial de parámetro $p\lambda$.Observamos que la suma de un número fijo de variables exponenciales tiene distribución Gamma.

▲

5.1.3. Procesos de Ramificación

Consideremos una partícula (neutrones, bacterias, virus informático, etc.) que puede generar nuevas partículas del mismo tipo. El grupo inicial de individuos pertenece a la generación 0 y suponemos que cada individuo produce una cantidad aleatoria ξ de descendientes con distribución de probabilidad

$$P(\xi = k) = p_k, \qquad k = 0, 1, 2, \dots \tag{5.9}$$

donde $p_k \geq 0$, $\sum_k p_k = 1$. Suponemos que todos los individuos actúan de manera independiente, que todos viven el mismo período de tiempo y todos siguen la misma ley P dada por (5.9) para generar su descendencia (ver figura 5.1). El proceso $(X_n)_{n\geq 1}$ donde X_n representa el tamaño de la n-ésima generación, es una cadena de Markov y se conoce como un *proceso de ramificación*.

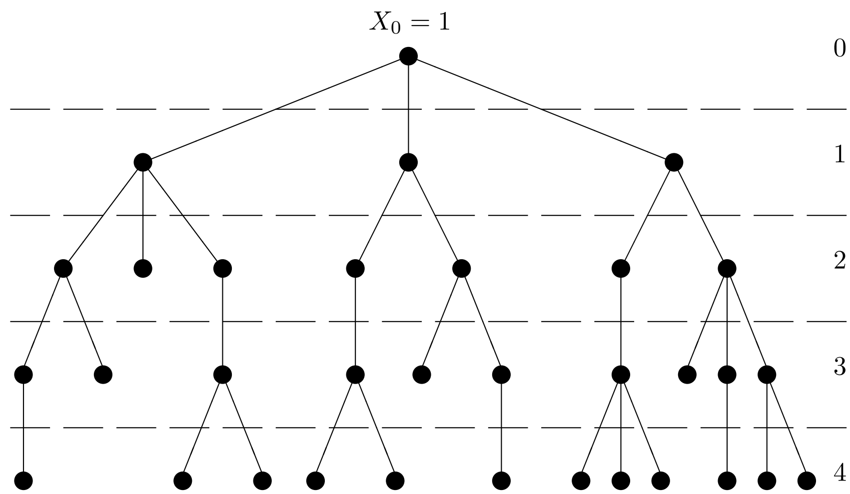

Figura 5.1

El espacio de estados de esta cadena es $\{0, 1, 2, \dots\}$ donde 0 es un estado absorbente. Por otro lado, si $X_n = k$, los k miembros de esta generación producen

$$\xi_1^n + \xi_2^n + \cdots + \xi_k^n = X_{n+1} \tag{5.10}$$

descendientes, que forman la siguiente generación de modo que

$$P_{kj} = P(\xi_1^n + \xi_2^n + \cdots + \xi_k^n = j | X_n = k). \tag{5.11}$$

Si una partícula produce $\xi = 0$ descendientes, lo interpretamos como que la partícula muere o desaparece. Puede ocurrir que luego de varias generaciones todos los descendientes de la partícula inicial hayan muerto o desaparecido. Decimos entonces que todos los descendientes de la partícula inicial se extinguieron. Un problema interesante es calcular la probabilidad de extinción U_∞ de un proceso de ramificación que comienza con una sola partícula. Una vez que resolvamos este problema, podemos hallar la probabilidad de que una cadena que comienza con k partículas se extinga, pues como las partículas actúan independientemente, esta probabilidad es $(U_\infty)^k$.

Probabilidades de Extinción.

La población se extingue cuando el tamaño de la población es 0. El instante (aleatorio) de extinción N es el primer índice n para el cual $X_n = 0$ y luego, obviamente, $X_k = 0$ para todo $k \geq N$. 0 es un estado absorbente y podemos calcular la probabilidad de extinción haciendo un análisis de la primera transición. Llamemos

$$U_n = P_1(N \leq n) = P_1(X_n = 0) \tag{5.12}$$

la probabilidad de extinción antes de n o en n. El único miembro inicial de la población produce $\xi_1^{(0)} = k$ descendientes. Por su parte, cada uno de estos descendientes generará una población de descendientes y cada una de estas líneas de descendencias debe desaparecer en $n-1$ generaciones o antes.

Las k poblaciones generadas por el individuo inicial son independientes entre sí y tienen las mismas propiedades estadísticas que la población inicial. Por lo tanto, la probabilidad de que una cualquiera de ellas desaparezca en $n-1$ generaciones es U_{n-1} por definición, y la probabilidad de que las k subpoblaciones mueran en $n-1$ generaciones es $(U_{n-1})^k$, por independencia.

Por la ley de la probabilidad total

$$U_n = \sum_{k=0}^{\infty} p_k (U_{n-1})^k, \qquad n = 1, 2, \dots \tag{5.13}$$

con $U_0 = 0$ y $U_1 = p_0$.

Recordemos que si ξ es una v.a. con valores enteros positivos y distribución de probabilidad $P(\xi = k) = p_k, \quad k = 0, 1, \dots,$ la función generadora de probabilidad (f.g.p.) $\phi(s)$ asociada a la v.a. ξ (o equivalentemente a su distribución (p_k)) se define por

$$\phi(s) = \mathrm{E}[s^\xi] = \sum_{k=0}^{\infty} s^k p_k, \qquad 0 \le s \le 1. \tag{5.14}$$

Podemos ahora escribir la relación (5.13) en términos de la función generadora:

$$U_n = \sum_{k=0}^{\infty} p_k (U_{n-1})^k = \phi(U_{n-1}) \tag{5.15}$$

es decir, si conocemos la función generadora de probabilidades $\phi(s)$, podemos calcular iterativamente las probabilidades de extinción U_n comenzando con $U_0 = 0$: $U_1 = \phi(U_0) = \phi(0)$, $U_2 = \phi(U_1)$, etc.

Ejemplo 5.7

En esta población un individuo no tiene descendientes con probabilidad $1/4$ o tiene dos descendientes con probabilidad $3/4$. La relación recursiva (5.13) es en este caso

$$U_n = \frac{1 + 3(U_{n-1})^2}{4}.$$

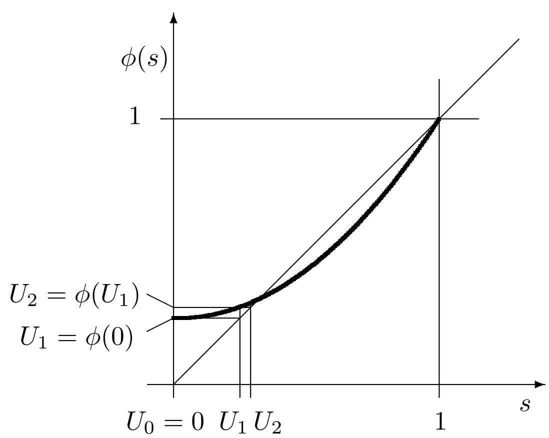

Figura 5.2

La función generadora es

$$\phi(s) = \mathrm{E}[s^\xi] = 1 \cdot \frac{1}{4} + s^2 \cdot \frac{3}{4} = \frac{1 + 3s^2}{4}$$

y vemos que $U_n = \phi(U_{n-1})$. Representamos esta función en la Figura 5.2. Podemos ver que las probabilidades de extinción convergen de manera creciente a la menor solución de la ecuación $u = \phi(u)$.

Esto también ocurre en el caso general: Si U_∞ es la menor solución de la ecuación $u = \phi(u)$, entonces U_∞ es la probabilidad de que la población se extinga en algún momento finito. La alternativa es que la población exista indefinidamente, lo que ocurre con probabilidad $1 - U_\infty$.

En el ejemplo que estamos considerando, la ecuación $u = \phi(u)$ es

$$u = \frac{1}{4} + \frac{3}{4} u^2,$$

con soluciones 1 y 1/3, y la menor solución es 1/3. ▲

Puede ocurrir que $U_\infty = 1$, en cuyo caso es seguro que la población desaparece en algún momento.

Ejemplo 5.8
Si las probabilidades son ahora $p_0 = 3/4$ y $p_2 = 1/4$, la función generadora es

$$\phi(s) = \frac{3}{4} + \frac{1}{4}s^2.$$

La ecuación $u = \phi(u)$ es ahora

$$u = \frac{3}{4} + \frac{1}{4}u^2$$

con soluciones 1 y 3. Como la menor solución es 1, $U_\infty = 1$.

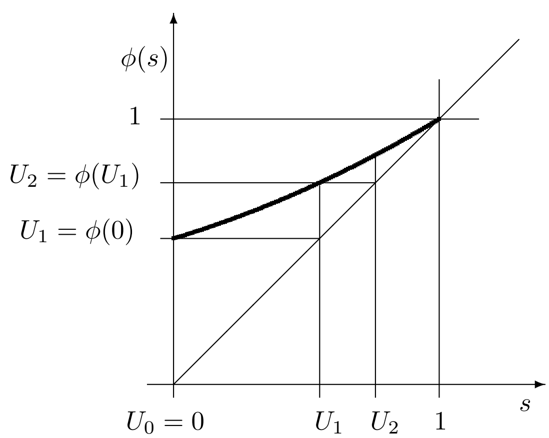

Figura 5.3

▲

Para determinar en cuál caso nos encontramos hay que ver si la curva de la función generadora $\phi(s)$ cruza la recta $y = x$ por debajo de 1, y esto se puede determinar por la pendiente de ϕ en 1:

$$\phi'(1) = \left.\frac{d\phi(s)}{ds}\right|_{s=1} = \mathrm{E}(\xi) = \mu.$$

En el razonamiento que sigue usaremos el hecho de que si $p_k > 0$ para algún $k > 1$, la f.g.p. correspondiente es estrictamente convexa en $[0, 1]$. Esto es consecuencia de que ϕ es una serie de potencias con coeficientes positivos.

- *Si* $0 < \mu \leq 1$ *entonces* $\phi(t) > t$, *para todo* $t \in [0, 1)$. Para probarlo, definimos una función $g(t) = \phi(t) - t$, esta función satisface que $g(0) = \phi(0)$, $g(1) = 0$ y es estrictamente decreciente puesto que su derivada $g'(t) = \phi'(t) - 1$ es estrictamente negativa, y esto se debe al hecho que ϕ' es estrictamente creciente y $\phi'(1) = \mu \leq 1$. Entonces, $g(t) > 0$, para $0 \leq t < 1$. En particular, la ecuación $\phi(t) = t$, no tiene raíces en $(0, 1)$.
- *Si* $\mu > 1$, *entonces la ecuación* $\phi(t) = t$ *tiene una única solución en* $[0, 1)$. Esto implica que $\lim_{t\uparrow 1} \phi'(t) = \phi'(1) = \mu > 1$. Por continuidad existe un $t_0 < 1$, tal que $\phi'(t) > 1$ para todo $t_0 < t \leq 1$, por el teorema del valor intermedio vemos que

$$\frac{\phi(1) - \phi(t_0)}{1 - t_0} = \frac{1 - \phi(t_0)}{1 - t_0} = \phi'(t') > 1, \qquad \text{para algún } t' \in (t_0, 1),$$

  de donde sigue que $g(t_0) = \phi(t_0) - t_0 < 0$, y puesto que g es continua y $g(0) = P(\xi = 0) > 0$, podemos afirmar que existe un $0 < \eta < 1$ con $g(\eta) = 0$. Por la convexidad estricta de ϕ es claro que g no puede tener ninguna otra raíz en $(\eta, 1)$, ni en $(0, \eta)$.

Sea η la raíz más pequeña de la ecuación $\phi(t) = t$, en $[0, 1]$. Los hechos anteriores implican que esta solución existe y además: *si* $\mu \leq 1$, *entonces* $\eta = 1$; *si* $\mu > 1$, *entonces* $\eta < 1$.

Tenemos entonces

$$\begin{array}{ccc} \phi'(1) < 1 & \text{no hay cruce} & U_\infty = 1, \\ \phi'(1) > 1 & \text{hay cruce} & U_\infty < 1. \end{array}$$

Pero hemos visto que $\phi'(1) = \mathrm{E}[\xi]$ y por lo tanto,

$$\begin{array}{ccc} \mathrm{E}[\xi] < 1 & \Rightarrow & U_\infty = 1, \\ \mathrm{E}[\xi] > 1 & \Rightarrow & U_\infty < 1. \end{array}$$

El caso límite corresponde a $\mathrm{E}[\xi] = 1$, donde $E[X_n|X_0 = 1] = 1$ para todo n, de modo que el tamaño promedio de la población es constante pero la población desaparece con seguridad, a menos que la varianza sea 0, es decir, que con probabilidad 1 todo individuo tenga exactamente un descendiente, en cuyo caso la población no se extingue nunca.

Ejemplo 5.9

Supongamos que el tamaño de las familias se distribuye según una ley geométrica con parámetro q,

$$P(\xi = k) = qp^k, \qquad k \geq 0, \qquad \text{para algún } p \in (0, 1).$$

Es fácil de calcular la función generadora ϕ,

$$\phi(s) = \sum_{n=0}^{\infty} qs^n p^n = \frac{q}{1 - ps}, \qquad |s| < p^{-1}.$$

La media vale $\mu = \frac{p}{q}$. Se puede verificar usando un argumento de inducción que la n-ésima composición de ϕ consigo misma puede ser escrita como

$$\phi_n(s) = \begin{cases} \dfrac{n - (n-1)s}{n + 1 - ns} & \text{si } \ p = q = 1/2, \\[2ex] \dfrac{q\left[p^n - q^n - ps(p^{n-1} - q^{n-1})\right]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} & \text{si } p \neq q. \end{cases}$$

¿Cuál es la probabilidad de extinción en este caso? Usaremos la forma explícita de ϕ_n para responder a esta pregunta. Recordemos que

$$P(X_n = 0) = \phi_n(0) = \begin{cases} \dfrac{n}{n+1} & \text{si } \ p = q = 1/2, \\[2ex] \dfrac{q\,(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{si } p \neq q, \end{cases}$$

por lo que al hacer $n \to \infty$, vemos que

$$\lim_{n\to\infty} P(X_n = 0) = \begin{cases} 1 & \text{si } \ p \leq q \\ \frac{q}{p}, & \text{si } \ p > q. \end{cases}$$

Observemos que si para algún $n \geq 1$, $X_n = 0$ entonces $X_{n+k} = 0$, para todo $k \geq 0$. Es decir que la población se extingue en un tiempo anterior o igual a n. Tenemos que

$$\begin{aligned} P(\text{extinción en un tiempo finito}) &= \lim_{n\to\infty} P(\text{extinción antes del instante } n) \\ &= \lim_{n\to\infty} P(X_n = 0) \\ &= \begin{cases} 1 & \text{si } \ p \leq q \\ \frac{q}{p}, & \text{si } \ p > q. \end{cases} \end{aligned}$$

Conclusión: La extinción ocurre con probabilidad 1, solamente en el caso en que $p/q = \mu = \mathrm{E}(X_1) \leq 1$; esta condición es bastante natural, puesto que $\mathrm{E}(X_n) = \mathrm{E}(X_1)^n \leq 1$, y es de esperarse que $X_n = 0$ tarde o temprano.

Veamos que el resultado de los ejercicios anteriores es consecuencia de un resultado más general.

Teorema 5.4 *Si $X_0 = 1$ tenemos que*

$$U_\infty = \lim_{n\to\infty} P_1(X_n = 0) = P(\textit{extinción en un tiempo finito}) = \eta,$$

donde η es la menor solución a la ecuación, $\phi(t) = t$. Además, $\eta = 1$, si $\mu < 1$, (el caso subcrítico) y $\eta < 1$, si $\mu > 1$ (caso super-crítico), mientras que en el caso en que $\mu = 1$, (el caso crítico) $\eta = 1$ si el tamaño de las familias tiene varianza estrictamente positiva.

Demostración. Sea $U_n = P_1(X_n = 0)$. Sabemos que

$$U_n = \phi(U_{n-1}).$$

Es claro que $\{X_n = 0\} \subset \{X_{n+1} = 0\}$, para todo $n \geq 1$, entonces U_n es una sucesión creciente y acotada; en consecuencia el limite de U_n existe y por la continuidad de ϕ debe de satisfacer

$$\eta = \lim_{n\to\infty} U_n \leq 1, \qquad \eta = \phi(\eta).$$

Veamos ahora que si ν es otra raíz positiva de la ecuación, entonces $\eta \leq \nu$. Dado que ϕ es una función estrictamente creciente, tenemos que

$$U_1 = \phi(0) \leq \phi(\nu) = \nu,$$

y se sigue que

$$U_2 = \phi(U_1) \leq \phi(\nu) = \nu,$$

y por inducción se ve que $U_n \leq \nu$, para todo $n \geq 1$, y por lo tanto que $\eta \leq \nu$. En consecuencia, η es la menor solución de la ecuación $t = \phi(t)$.

Ya vimos que si $\mu > 1$ entonces la ecuación $\phi(t) = t$, tiene una única solución η en $[0, 1)$, y de hecho la otra solución a la ecuación es $t = 1$. La menor solución es $\eta < 1$. Por otro lado, en el caso en que $\mu < 1$, vimos que $\phi(t) > t$ para todo $t \in [0, 1)$, y es claro que $\phi(1) = 1$, por lo tanto la solución positiva más pequeña a la ecuación $\phi(t) = t$ es $\eta = 1$. En el caso especial en que $\mu = 1$, el caso crítico, necesitamos distinguir entre el caso en que $\sigma^2 = 0$, donde $\phi(s) = s$ y por lo tanto $\eta = 0$, y el caso $\sigma^2 > 0$, donde $\phi(s) > s$, $s \in [0, 1)$ y por lo tanto $\eta = 1$. ■

5.2. Convergencia de Variables Aleatorias

Hay varios modos de convergencia en la Teoría de Probabilidades. Vamos a considerar algunos de ellos a continuación. Sea $X_n, n \geq 1$ una sucesión de variables aleatorias definidas en un espacio de probabilidad común $(\Omega, \mathcal{F}, P)$ y sea X otra variable definida sobre este mismo espacio.

Definición 5.1 La sucesión X_n converge *puntualmente* a X si para todo $\omega \in \Omega$ se cumple que

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Notación: $X_n \to X$.

Definición 5.2 La sucesión X_n converge *casi seguramente* o *con probabilidad 1* a X si existe un conjunto nulo $N \in \mathcal{F}$ tal que para todo $\omega \notin N$ se cumple que

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Notación: $X_n \to X$ $c.s.$ o $c.p.1$, o también $X_n \xrightarrow{c.s.} X$.

Definición 5.3 La sucesión X_n converge *en probabilidad* a X si dado cualquier $\varepsilon > 0$ se tiene que

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

Notación: $X_n \xrightarrow{P} X$.

Definición 5.4 La sucesión X_n converge *en L^p*, $1 \leq p < \infty$, a X si $E[|X_n|^p] < \infty$ y

$$\lim_{n\to\infty} \mathrm{E}[|X_n - X|^p] = 0.$$

Notación: $X_n \xrightarrow{L^p} X$ o también $X_n \to X$ en L^p.

Definición 5.5 La sucesión X_n converge *en distribución* a X

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \qquad \text{para todo } x \in \mathcal{C}(F_X),$$

donde $\mathcal{C}(F_X)$ es el conjunto de puntos de continuidad de F_X.
Notación: $X_n \xrightarrow{D} X$. También usaremos la notación $X_n \xrightarrow{D} F_X$

Observación 5.2

1. En la convergencia c.s., para cada $\omega \in \Omega$ consideramos si la sucesión de números reales $X_n(\omega)$ converge al número real $X(\omega)$. Si esto ocurre para todo ω fuera de un conjunto N de medida 0, decimos que hay convergencia c.s.

2. La convergencia en L^2 se conoce usualmente como convergencia en media cuadrática.

3. En la definición de convergencia en distribución, las variables sólo aparecen a través de sus funciones de distribución. Por lo tanto las variables no tienen que estar definidas en un mismo espacio de probabilidad.

4. Es posible demostrar que una función de distribución tiene a lo sumo una cantidad numerable de discontinuidades. Como consecuencia $\mathcal{C}(F_X)$ es la recta real, excepto, posiblemente, por un conjunto numerable de puntos.

5. Es posible demostrar que en cualquiera de estos modos de convergencia el límite es (esencialmente) único.

▲

Ejemplo 5.10
Sea $X_n \sim \Gamma(n, n)$. Veamos que $X_n \xrightarrow{P} 1$ cuando $n \to \infty$.

Observamos que $\mathrm{E}[X_n] = 1$ mientras que $\mathrm{Var}[X] = 1/n$. Usando la desigualdad de Chebyshev obtenemos que para todo $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{n\varepsilon^2} \to 0 \quad \text{cuando } n \to \infty.$$

▲

Ejemplo 5.11
Sean $X_1, X_2, \dots$ v.a.i. con densidad común

$$f(x) = \begin{cases} \alpha x^{-\alpha-1}, & \text{para } x > 1, \alpha > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

y sea $Y_n = n^{-1/\alpha} \max_{1 \leq k \leq n} X_k,\ n \geq 1$. Demuestre que Y_n converge en distribución y determine la distribución límite.

Para resolver este problema vamos a calcular la f.d. común:

$$F(x) = \int_1^x \alpha x^{-\alpha-1} dy = 1 - x^{-\alpha}$$

siempre que $x > 1$ y vale 0 si no. Por lo tanto, para cualquier $x > 1$,

$$\begin{aligned} F_{Y_n}(x) &= P(\max_{1 \leq k \leq n} X_k \leq x n^{1/\alpha}) = \big(F(xn^{1/\alpha})\big)^n \\ &= \big(1 - \frac{1}{nx^\alpha}\big)^n \to e^{-x^{-\alpha}} \quad \text{cuando } n \to \infty. \end{aligned}$$

▲

Ejemplo 5.12 (La Ley Débil de los Grandes Números)
Esta es una versión débil de la LGN. Sean $X_1, X_2, \dots$ v.a.i.i.d. con media μ y varianza finita σ^2 y pongamos $S_n = X_1 + \cdots + X_n$, $n \geq 1$. La Ley (Débil) de los Grandes Números dice que para todo $\varepsilon > 0$,

$$P(|\frac{S_n}{n} - \mu| > \varepsilon) \to 0 \quad \text{cuando } n \to \infty,$$

es decir

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad \text{cuando } n \to \infty.$$

La prueba de esta proposición sigue de la desigualdad de Chebyshev:

$$P(|\frac{S_n}{n} - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{cuando } n \to \infty.$$

▲

Ejemplo 5.13 (Aproximación de Poisson)
Sea $X_n \sim \mathcal{B}in(n, \frac{\lambda}{n})$, entonces

$$X_n \xrightarrow{D} \mathcal{P}ois(\lambda)$$

Vemos esto

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \big(\frac{\lambda}{n}\big)^k \big(1 - \frac{\lambda}{n}\big)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \big(\frac{\lambda}{n}\big)^k \big(1 - \frac{\lambda}{n}\big)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \big(1 - \frac{\lambda}{n}\big)^{n-k} \\ &= \big(1 - \frac{1}{n}\big)\big(1 - \frac{2}{n}\big) \cdots \big(1 - \frac{k-1}{n}\big) \frac{\lambda^k}{k!} \big(1 - \frac{\lambda}{n}\big)^{n-k} \\ &= \frac{\big(1 - \frac{1}{n}\big)\big(1 - \frac{2}{n}\big) \cdots \big(1 - \frac{k-1}{n}\big)}{\big(1 - \frac{\lambda}{n}\big)^k} \frac{\lambda^k}{k!} \big(1 - \frac{\lambda}{n}\big)^n \end{aligned}$$

Si ahora hacemos $n \to \infty$ la primera fracción tiende a 1 porque k y λ están fijos, mientras que

$$\lim_{n \to \infty} \big(1 - \frac{\lambda}{n}\big)^n = e^{-\lambda}$$

Por lo tanto

$$\lim_{n \to \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

▲

5.2.1. Relación entre los Distintos Tipos de Convergencia

En esta sección nos planteamos investigar la relación entre los distintos tipos de convergencia que hemos definido, y en particular exploramos la posibilidad de poder ordenar estos conceptos.

I. Convergencia c.s. implica Convergencia en Probabilidad.

Fijemos $\omega \in \Omega$. Recordemos que la convergencia $X_n(\omega) \to X(\omega)$ quiere decir que para todo $\varepsilon > 0$ existe $k = k(\omega)$ tal que si $n \geq k$ se tiene que

$$|X_n(\omega) - X(\omega)| < \varepsilon.$$

Es fácil ver que, sin pérdida de generalidad, podemos tomar ε de la forma $1/r$ para $r \in \mathbb{N}$. Por lo tanto, $X_n(\omega) \to X(\omega)$ si y solo si para todo $r \in \mathbb{N}$ existe $k = k(\omega)$ tal que

$$n \geq k(\omega) \quad \Rightarrow \quad |X_n(\omega) - X(\omega)| < \frac{1}{r}. \tag{5.16}$$

Como consecuencia tenemos que el conjunto de $\omega \in \Omega$ para los cuales la sucesión $X_n(\omega)$ converge a $X(\omega)$ se puede expresar como

$$C = \bigcap_{r=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{r}\} \tag{5.17}$$

En efecto, si $\omega \in C$ entonces tiene que estar en todos los conjuntos que figuran en la primera intersección (*para todo* r), en alguno de los conjuntos de la unión (*existe* k) y en todos los conjuntos de la segunda intersección (*para todo* $n \geq k$) y la condición que aparece en la definición del conjunto entre llaves es precisamente la desigualdad de la derecha en (5.16). Teniendo en cuenta que todas las funciones son medibles, esto muestra que el conjunto C es medible. La convergencia con probabilidad 1 dice que $P(C) = 1$.

Observamos que si $P(\cap_{r\geq 1} D_r) = 1$, necesariamente $P(D_r) = 1$ para todo $r \geq 1$, pues si para algún r_0 se tiene $P(D_{r_0}) < 1$ entonces, por monotonía de la probabilidad P,

$$P(\cap_{r\geq 1} D_r) \leq P(D_{r_0}) < 1$$

lo cual es una contradicción. En consecuencia, para todo r

$$P(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{r}\}) = 1$$

Teniendo en cuenta que para r fijo, la sucesión de conjuntos $\cap_{n=k}^{\infty}\{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{r}\}$ es creciente en k y usando de nuevo la monotonía de la probabilidad P tenemos

$$\begin{aligned} 1 = P(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{r}\}) &= P(\lim_{k\to\infty} \bigcap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{r}\}) \\ &= \lim_{k\to\infty} P(\bigcap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{r}\}) \end{aligned}$$

Este razonamiento demuestra que $X_n \xrightarrow{c.s.} X$ cuando $n \to \infty$ sí y sólo sí para todo $\varepsilon > 0$ y δ, $0 < \delta < 1$, existe n_0 tal que, para todo $n > n_0$

$$P(\bigcap_{m>n} \{|X_m - X| < \varepsilon\}) > 1 - \delta \tag{5.18}$$

o equivalentemente

$$P(\bigcup_{m>n}\{|X_m - X| > \varepsilon\}) < \delta.$$

Como, para $m > n$,

$$\{|X_m - X| > \varepsilon\} \subset \bigcup_{m>n}\{|X_m - X| > \varepsilon\},$$

la sucesión también converge en probabilidad. El siguiente ejemplo muestra que el recíproco es falso.

Ejemplo 5.14
Sean $X_1, X_2, \dots$ v.a.i. tales que

$$P(X_n = 1) = 1 - \frac{1}{n} \quad \text{y} \quad P(X_n = n) = \frac{1}{n}, \quad n \geq 1.$$

Claramente,

$$P(|X_n - 1| > \varepsilon) = P(X_n = n) = \frac{1}{n} \to 0, \quad \text{cuando } n \to \infty,$$

para todo $\varepsilon > 0$, es decir,

$$X_n \xrightarrow{P} 1 \quad \text{cuando } n \to \infty.$$

Veamos ahora que X_n no converge c.s. a 1 cuando $n \to \infty$. Para todo $\varepsilon > 0$, $\delta \in (0,1)$ y $N > n$ tenemos

$$\begin{aligned}
P(\bigcap_{m>n}\{|X_m - X| < \varepsilon\}) &= P(\lim_N \bigcap_{m=n+1}^{N}\{|X_m - X| < \varepsilon\}) \\
&= \lim_N P(\bigcap_{m=n+1}^{N}\{|X_m - X| < \varepsilon\}) \\
&= \lim_N \prod_{m=n+1}^{N} P(|X_m - 1| < \varepsilon) \\
&= \lim_N \prod_{m=n+1}^{N} P(X_m = 1) = \lim_N \prod_{m=n+1}^{N}\big(1 - \frac{1}{m}\big) \\
&= \lim_N \prod_{m=n+1}^{N} \frac{m-1}{m} = \lim_N \frac{n}{N} = 0,
\end{aligned}$$

para cualquier n. Esto muestra que no existe n_0 para el cual (5.18) valga, y por lo tanto X_n no converge c.s. a 1 cuando $n \to \infty$.

▲

II. Convergencia en L^p implica Convergencia en Probabilidad

Usando la desigualdad de Markov, para $\varepsilon > 0$ fijo

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p}\,\mathrm{E}[|X_n - X|^p] \to 0$$

cuando $n \to \infty$, lo que muestra la conclusión.

En este caso el recíproco tampoco es cierto. Para empezar, $\mathrm{E}[|X_n - X|]$ no tiene por qué existir, pero aun si existe puede ocurrir que haya convergencia en probabilidad sin que haya convergencia en L^p.

Ejemplo 5.15
Sea $\alpha > 0$ y sea $X_1, X_2, \ldots$ v.a. tales que

$$P(X_n = 1) = 1 - \frac{1}{n^\alpha} \quad \text{y} \quad P(X_n = n) = \frac{1}{n^\alpha}, \quad n \geq 1.$$

Como

$$P(|X_n - 1| > \varepsilon) = P(X_n = n) = \frac{1}{n^\alpha} \to 0, \quad \text{cuando } n \to \infty,$$

para todo $\varepsilon > 0$, tenemos que

$$X_n \xrightarrow{P} 1 \quad \text{cuando } n \to \infty.$$

Por otro lado

$$\mathrm{E}[|X_n - 1|^p] = 0^p \cdot \big(1 - \frac{1}{n^\alpha}\big) + |n-1|^p \frac{1}{n^\alpha} = \frac{(n-1)^p}{n^\alpha},$$

de donde obtenemos que

$$\mathrm{E}[|X_n - 1|^p] \to \begin{cases} 0, & \text{para } p < \alpha, \\ 1, & \text{para } p = \alpha, \\ +\infty, & \text{para } p > \alpha, \end{cases} \tag{5.19}$$

Esto muestra que $X_n \stackrel{L^p}{\to} 1$ cuando $n \to \infty$ si $p < \alpha$ pero X_n no converge en L^p si $p \geq \alpha$. Por lo tanto, convergencia en L^p es más fuerte que convergencia en probabilidad. ▲

Observación 5.3 Si $\alpha = 1$ y las variables son independientes, cuando $n \to \infty$

$$\begin{aligned} &X_n \xrightarrow{P} 1, \\ &X_n \stackrel{c.s.}{\nrightarrow}, \\ &\mathrm{E}[X_n] \to 2, \\ &X_n \xrightarrow{L^p} 1 \quad \text{para } 0 < p < 1, \\ &X_n \stackrel{L^p}{\nrightarrow} \quad \text{para } p \geq 1. \end{aligned}$$

▲

Observación 5.4 Si $\alpha = 2$ y las variables son independientes, cuando $n \to \infty$

$$\begin{aligned} &X_n \xrightarrow{P} 1, \\ &X_n \xrightarrow{c.s.} 1, \\ &\mathrm{E}[X_n] \to 1, \quad \text{y} \quad \mathrm{Var}[X_n] \to 1 \\ &X_n \xrightarrow{L^p} 1 \quad \text{para } 0 < p < 2, \\ &X_n \stackrel{L^p}{\nrightarrow} \quad \text{para } p \geq 2. \end{aligned}$$

▲

III. Convergencia en L^p y Convergencia c.s. son Independientes

Ninguna de las dos implica la otra, y esto lo podemos ver de las observaciones anteriores. En el primer caso, para $0 < p < 1$ hay convergencia en L^p mientras que no hay convergencia c.s. En el segundo hay convergencia c.s. pero no hay convergencia en L^p para $p \geq 2$.

IV. Convergencia en Probabilidad implica Convergencia en Distribución

Sea $\varepsilon > 0$, entonces

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(\{X_n \leq x\} \cap \{|X_n - X| \leq \varepsilon\}) + P(\{X_n \leq x\} \cap \{|X_n - X| > \varepsilon\}) \\ &\leq P(\{X \leq x + \varepsilon\} \cap \{|X_n - X| \leq \varepsilon\}) + P(|X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon) \end{aligned}$$

es decir,

$$F_{X_n}(x) \leq F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon). \tag{5.20}$$

De manera similar se demuestra que

$$F_X(x - \varepsilon) \leq F_{X_n}(x) + P(|X_n - X| > \varepsilon). \tag{5.21}$$

Como $X_n \xrightarrow{P} X$ cuando $n \to \infty$ obtenemos, haciendo $n \to \infty$ en (5.20) y (5.21),

$$F_X(x - \varepsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$$

Esta relación es válida para todo x y todo $\varepsilon > 0$. Para demostrar la convergencia en distribución suponemos que $x \in \mathcal{C}(F_X)$ y hacemos $\varepsilon \to 0$. Obtenemos

$$F_X(x) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x),$$

por lo tanto

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

y como esto vale para cualquier $x \in \mathcal{C}(F_X)$ obtenemos la convergencia en distribución. ▲

Observación 5.5 Si F_X tiene un salto en x, sólo podemos concluir que

$$F_X(x-) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x),$$

y $F_X(x) - F_X(x-)$ es el tamaño del salto. Esto explica por qué sólo se toman en cuenta los puntos de continuidad en la definición de convergencia en distribución.

Como mencionamos anteriormente, la convergencia en distribución no requiere que las variables estén definidas en un mismo espacio de probabilidad. El siguiente ejemplo muestra que aun cuando las variables estén definidas en un espacio común, existen sucesiones que sólo convergen en distribución.

Ejemplo 5.16
Sea X una variable con distribución simétrica, continua y no-degenerada y definimos $X_1, X_2, \dots$ por $X_{2n} = X$ y $X_{2n-1} = -X$, $n = 1, 2, \dots$. Como $X_n \stackrel{D}{=} X$ para todo n, tenemos, en particular, $X_n \xrightarrow{D} X$ cuando $n \to \infty$. Por otro lado, como X tiene distribución no-degenerada existe $a > 0$ tal que $P(|X| > a) > 0$ (¿por qué?). En consecuencia, para todo $\varepsilon > 0$, $0 < \varepsilon < 2a$,

$$P(|X_n - X| > \varepsilon) = \begin{cases} 0, & \text{para } n \text{ par}, \\ P(|X| > \frac{\varepsilon}{2}) > 0, & \text{para } n \text{ impar}. \end{cases}$$

Esto muestra que X_n no puede converger en probabilidad a X cuando $n \to \infty$, y en consecuencia tampoco c.s. o en L^p. ▲

Podemos resumir todo lo anterior en el siguiente teorema.

Teorema 5.5 *Sean X y $X_1, X_2, \dots$ variables aleatorias, entonces, cuando $n \to \infty$,*

$$\begin{array}{ccccc} X_n \xrightarrow{c.s.} X & \Longrightarrow & X_n \xrightarrow{P} X & \Longrightarrow & X_n \xrightarrow{D} X \\ & & \Uparrow & & \\ & & X_n \xrightarrow{L^p} X & & \end{array}$$

Ninguna de las implicaciones se puede invertir.

5.3. La Ley de Grandes Números

En el capítulo 2 y en el ejemplo 5.12 de este capítulo demostramos dos versiones de la ley débil de los grandes números. Consideraremos en esta sección, sin demostración, otras versiones (débiles y fuertes) de este teorema de convergencia. Como hemos visto, la LDGN es un teorema sobre la convergencia en probabilidad del promedio (empírico) de una sucesión de v.a.i. a la media de la distribución común. Este resultado lo demostramos bajo la hipótesis de la existencia de los dos primeros momentos de la distribución.

En realidad es posible demostrar resultados más fuertes, tanto en el modo de convergencia como en las hipótesis que se requieren para probarlos. El costo es una mayor complejidad en las demostraciones que requieren herramientas más avanzadas. Por esta razón presentamos un resumen de los principales resultados, sin incluir sus demostraciones.

5.3.1. Ley Débil de Grandes Números

Consideremos una sucesión de variables aleatorias no correlacionadas con $\mathrm{E}(X_i) = \mu_i$, $\mathrm{Var}(X_i) = \sigma_i^2$. Sea $\bar{X}_n = \sum_{i=1}^n X_i/n$, entonces

$$\mathrm{E}(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^n \mu_i = \bar{\mu}_n, \qquad \mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2.$$

Teorema 5.6 (Chebychef) *Si se satisface que*

$$\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0 \quad (n \to 0) \tag{5.22}$$

entonces

$$\bar{X}_n - \bar{\mu}_n = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{P} 0 \tag{5.23}$$

La demostración es una simple aplicación de la desigualdad de Chebychef. Observamos que si $\mathrm{Var}(X_i) = \sigma^2$ para todo i, la condición (5.22) se satisface.

5.3.2. Ley Fuerte de Grandes Números

El primer resultado corresponde al caso de variables independientes con igual distribución.

Teorema 5.7 *Sea $X_n, n \geq 1$ una sucesión de v.a.i.i.d. y sea $S_n = \sum_{k=1}^n X_k$. Entonces para $A \in \mathbb{R}$*

$$\frac{1}{n}S_n \xrightarrow{c.s.} A \quad \Longleftrightarrow \quad \mathrm{E}(|X_1|) < \infty$$

y en este caso $A = \mathrm{E}(X_1)$. Por otro lado, si $\mathrm{E}(X_1) = \pm\infty$ entonces

$$\frac{1}{n}S_n \xrightarrow{c.s.} \pm\infty.$$

Para variables aleatorias que no necesariamente tienen igual distribución tenemos el siguiente resultado, que se debe a Kolmogorov.

Teorema 5.8 (Kolmogorov) *Sea $X_n, n \geq 1$ una sucesión de v.a.i. y sea $S_n = \sum_{k=1}^n X_k$. Supongamos que las variables X_n son centradas y tienen varianzas σ_k^2 que satisfacen*

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty. \tag{5.24}$$

Entonces

$$\frac{1}{n} S_n \xrightarrow{c.s.} 0.$$

Corolario 5.1 *Si en el teorema anterior se satisfacen todas las condiciones excepto que las variables X_k en lugar de estar centradas tienen media μ_k, entonces*

$$\frac{1}{n}\big(S_n - \sum_{k=1}^n \mu_k\big) = \frac{1}{n}\big(\sum_{k=1}^n (X_k - \mu_k)\big) \xrightarrow{c.s.} 0.$$

Corolario 5.2 (Borel) *Consideremos una sucesión de ensayos de Bernoulli $X_n, n \geq 1$ con probabilidad de éxito p. Entonces*

$$\frac{1}{n} S_n \xrightarrow{c.s.} p.$$

Ejemplo 5.17
Consideremos la sucesión de variables aleatorias (X_k) con función de probabilidad

$$P(X_k = k) = P(X_k = -k) = \frac{1}{2\sqrt{k}}, \quad P(X_k = 0) = 1 - \frac{1}{\sqrt{k}}.$$

Veamos si estas variables satisfacen las condiciones de los teoremas de Chebychef y de Kolmogorov. Es fácil ver que las variables son centradas y

$$\sigma_k^2 = \mathrm{Var}(X_k) = \mathrm{E}(X_k^2) = k^2 \frac{1}{\sqrt{k}} = k^{3/2}$$

y vemos que la condición (5.24) del teorema de Kolmogorov no se satisface. Para verificar si se satisface la condición (5.22) del teorema de Chebychef usamos el hecho de que la función $x^{3/2}$ es creciente. Tenemos

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = \frac{1}{n^2} \sum_{i=1}^n i^{3/2} \sim \frac{1}{n^2} \int_0^n x^{3/2}\, dx = \frac{2}{5} n^{1/2} \to \infty$$

y vemos que esta condición tampoco se satisface. ▲

Ejemplo 5.18
Si modificamos la función de probabilidad del ejemplo anterior de modo que

$$P(X_k = k) = P(X_k = -k) = \frac{1}{2k^\alpha}, \quad P(X_k = 0) = 1 - \frac{1}{k^\alpha}.$$

para algún $\alpha > 1$ vemos que

$$\sigma_k^2 = \mathrm{Var}(X_k) = \mathrm{E}(X_k^2) = k^2 \frac{1}{k^\alpha} = k^{2-\alpha}$$

y la condición (5.24) se satisface siempre que $\alpha > 1$.

5.3.3. Aplicaciones

Supongamos que nos interesa estimar la probabilidad de un cierto evento, por ejemplo, para $a < b$ nos interesa estimar

$$p = P(X \in (a, b])$$

Una posibilidad es considerar una sucesión independiente $(X_k)_{k=1}^n$ de realizaciones de esta variable y estimar p por la proporción de veces que se satisface $X_i \in (a, b]$. Definimos $Y_i = \mathbf{1}_{(a,b]}(X_i)$, de modo que Y_i vale 1 si $X_i \in (a, b]$ y 0 en otro caso.

Con esta definición las variables Y_i forman una sucesión de variables de Bernoulli independientes con probabilidad de éxito la probabilidad p que deseamos estimar. Recordando que $\mathrm{E}(Y_i) = p$, la LFGN nos dice que

$$\hat{p}_n = \frac{1}{n} S_n \xrightarrow{c.s.} p.$$

Este resultado nos dice que si estimamos p por $\hat{p}_n$, con probabilidad 1 cuando $n \to \infty$ el estimador converge al valor real (desconocido) del parámetro. Esta propiedad se conoce como *consistencia*.

Como caso particular, si tenemos una variable discreta X con valores $x_1, \dots, x_k$, podemos estimar $p_j = P(X = x_j)$ usando un procedimiento similar. En este caso el estimador es

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_j}(X_i).$$

La LFGN garantiza que estos estimadores son consistentes.

¿Qué sucede si la variable X tiene distribución continua? Supongamos que f es la densidad de X y queremos estimar el valor de f en $x = a$. Podemos considerar el intervalo $(a - h, a + h)$ para algún valor pequeño de $h > 0$ y usar las aproximaciones

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(a-h,a+h)}(X_i) \approx P(X \in (a-h, a+h)) = \int_{a-h}^{a+h} f(x)\,dx \approx 2hf(a).$$

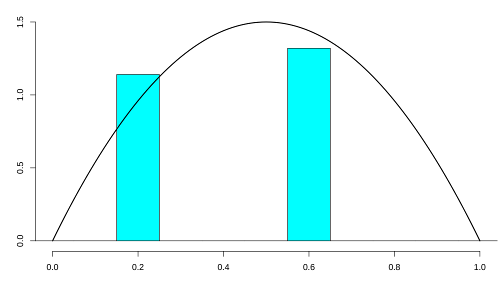

Figura 5.4

Esto sugiere estimar la densidad en a por

$$\hat{f}(a) = \frac{1}{2h} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(a-h,a+h)}(X_i) = \frac{1}{2h} \bar{Y}_n \tag{5.25}$$

donde ponemos $Y_i = \mathbf{1}_{(a-h,a+h)}(X_i)$ y $\bar{Y}_n = \sum_{i=1}^{n} Y_i / n$.

En la figura 5.4 presentamos la aproximación que obtuvimos para los valores de una densidad beta de parámetros $(2,2)$ en los puntos $x = 0.2$ y $x = 0.6$ a partir de 500 simulaciones de una variable con esta distribución. La línea continua representa la densidad de la distribución $\beta(2,2)$ y los rectángulos la aproximación a partir de intervalos de ancho $h = 0.1$ centrados en los valores de x. Como podemos observar, el valor estimado para $x = 0.2$ está por encima del verdadero valor mientras que para $x = 0.6$ la estimación cae por debajo.

A partir de la ecuación (5.25) vemos que el área del rectángulo es

$$2h \times \hat{f}(a) = 2h \times \frac{\bar{Y}_n}{2h} = \bar{Y}_n.$$

Si repetimos este procedimiento en una red de puntos equidistantes que cubran el dominio de la densidad y representamos las aproximaciones obtenidas por rectángulos, obtenemos un *histograma*. Este nombre fue propuesto por Karl Pearson en 1881, aunque la idea de este tipo de representación es muy anterior. En la figura 5.5 presentamos el histograma correspondiente a esta muestra. Vemos que, globalmente, la aproximación es razonable.

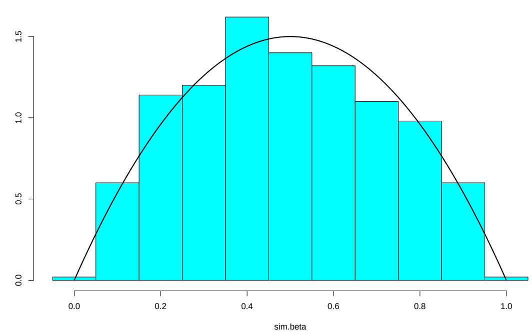

Figura 5.5

5.4. El Teorema Central del Límite

Otro resultado fundamental de la Teoría de Probabilidad clásica es el Teorema Central del Límite (TCL). En esta sección describiremos, de nuevo sin incluir las demostraciones, las principales versiones de este importante teorema. En el capítulo 2 vimos una versión elemental que muestra que la distribución normal aparece como límite de sucesión de variables aleatorias binomiales. El enunciado preciso es el siguiente.

Teorema 5.9 (TCL de DeMoivre-Laplace) *Sean $a < b$ números reales y $X_n, n \geq 1$ una sucesión de variables aleatorias con distribución de Bernoulli de parámetro $p \in (0,1)$ y sea $S_n = \sum_1^n X_i$ entonces*

$$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{w} \mathcal{N}(0,1) \tag{5.26}$$

cuando $n \to \infty$.

La siguiente versión muestra que el teorema se puede extender a sumas de v.a.i. con igual distribución, siempre que tengan segundo momento finito.

Teorema 5.10 (TCL) *Sea $X_n, n \geq 1$ una sucesión de v.a.i.i.d. con media $\mu = \mathrm{E}(X_i)$ y varianza $\sigma^2 = \operatorname{Var} X_i$. Sea $S_n = \sum_1^n X_i$, entonces*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{w} \mathcal{N}(0,1)$$

cuando $n \to \infty$.

La demostración de este teorema requiere herramientas que no están al nuestro alcance, pero para dar la idea, haremos la prueba para un acaso particular, suponiendo que las variables X_n tienen función generadora de momentos $M(t) = \mathrm{E}(e^{tX})$, que existe para $|t| < \delta$ para algún $\delta > 0$.

Demostración Definimos $Y_i = (X_i - \mu)/\sigma$, de modo que $\mathrm{E}(Y_i) = 0$ y $\operatorname{Var}(Y_i) = 1$. Usando (5.1) obtenemos

$$M_Y(t) = 1 + \frac{t^2}{2} + \frac{\mathrm{E}[(X-\mu)^3]}{6\sigma^3} t^3 + \cdots$$

Ahora calculamos las f.g.m. de las sumas, normalizadas

$$\begin{aligned}
\mathrm{E}\Big[\exp\Big\{t\Big(\frac{S_n - n\mu}{n^{1/2}\sigma}\Big)\Big\}\Big] &= \mathrm{E}\Big[\exp\Big\{\frac{t}{n^{1/2}}\sum_{i=1}^n Y_i\Big\}\Big] \\
&= \Big(M_Y\big(\frac{t}{n^{1/2}}\big)\Big)^n \\
&= \Big(1 + \frac{t^2}{2n} + \frac{\mathrm{E}[(X-\mu)^3]}{6\sigma^3 n^{3/2}} t^3 + \cdots\Big)^n \\
&\to e^{t^2/2}.
\end{aligned}$$

La función límite es la f.g.m. de la distribución normal típica y por el teorema de continuidad obtenemos el resultado. ■

¿Qué sucede si las variables X_i no tienen la misma distribución? En este caso es posible tener un resultado similar, pero se requiere una condición, debida a Lindeberg, que garantiza que ninguno de los sumandos sea determinante en el valor de la suma. Esta condición se expresa en términos de las varianzas. Supongamos que la variable X_n tiene varianze σ_n^2 y sea

$$s_n^2 = \operatorname{Var}(S_n) = \sum_1^n \sigma_i^2,$$

La condición de Lindeberg pide que para todo $\varepsilon > 0$,

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathrm{E}\left[X_i^2 \mathbf{1}_{(|X_i| \geq \varepsilon s_n)}\right] \to 0 \qquad (n \to \infty).$$

Teorema 5.11 (Lindeberg) *Sea $X_n, n \geq 1$ una sucesión de v.a.i. pero no necesariamente con la misma distribución, con $\mathrm{E}(X_i) = 0$ y $\operatorname{Var}(X_i) = \sigma_i^2 > 0$. Sea $S_n = \sum_1^n X_i$. Si la condición de Lindeberg se satisface entonces*

$$\frac{S_n}{s_n} \xrightarrow{w} \mathcal{N}(0,1)$$

cuando $n \to \infty$.

El significado de la condición de Lindeberg es poco claro, pero sirve para garantizar que ninguna variable en la suma tenga una contribución determinante al valor total. En particular, implica que

$$\frac{1}{s_n}\max_{k \leq n} \sigma_k^2 \to 0 \qquad (n \to \infty).$$

Una manera alternativa de presentar la condición de Lindeberg es la siguiente. Definimos las variables

$$Y_i = \begin{cases} X_i & \text{si } |X_i| \geq \varepsilon s_n, \\ 0 & \text{si } |X_i| < \varepsilon s_n. \end{cases} \tag{5.27}$$

Estas variables 'truncadas' coinciden con las variables originales si sus valores son grandes, pero toman el valor 0 si no es así. La condición de Lindeberg es

$$\frac{1}{s_n^2} \sum_{i=1}^{n} \mathrm{E}\left[Y_i^2\right] \to 0 \qquad (n \to \infty). \tag{5.28}$$

Ejemplos 5.19

1. Consideremos la sucesión de v.a.i. con distribución

$$P(X_n = n^\alpha) = P(X_n = -n^\alpha) = \frac{1}{2n^{2\alpha}}, \quad P(X_n = 0) = 1 - \frac{1}{n^{2\alpha}}$$

Es sencillo verificar que $\mathrm{E}(X_n) = 0$ y $\sigma_n = \mathrm{Var}(X_n) = 1$, de modo que $s_n = \sqrt{n}$. De acuerdo a la ecuación (5.27) para ε fijo, $Y_i = 0$ si $|X_i| = i^\alpha < \varepsilon\sqrt{n}$, es decir, si $i < \varepsilon^{1/\alpha} n^{1/2\alpha}$. Llamemos $n_0 = n_0(\alpha, \varepsilon) = [\varepsilon^{1/\alpha} n^{1/2\alpha}]$, con esta notación la suma en la condición (5.28) es

$$\frac{1}{s_n^2} \sum_{i=1}^{n} \mathrm{E}\left[Y_i^2\right] = \frac{1}{n} \sum_{i=n_0}^{n} \mathrm{E}\left[X_i^2\right] = \frac{n - n_0}{n} \tag{5.29}$$

siempre que $n_0 < n$. Consideremos tres casos:

- $\alpha < 1/2$. Tenemos
$$\frac{n_0}{n} \sim \varepsilon^{1/\alpha} \frac{n^{1/2\alpha}}{n} \to \infty$$
En consecuencia, existe N tal que si $n > N$, $n < n_0$ y la condición $|X_i| \geq \varepsilon s_n$ no se satisface para ningún índice $i \leq n$, de modo que, por el truncamiento $Y_i = 0$ para todo i. En este caso se satisface (5.28) y el TCL es válido.
- $\alpha > 1/2$. Tenemos
$$\frac{n_0}{n} \sim \varepsilon^{1/\alpha} \frac{n^{1/2\alpha}}{n} \to 0$$
En consecuencia, dado $\delta, 0 < \delta < 1$ existe N tal que si $n > N$, $n_0 < \delta n$ y la expresión en la ecuación (5.29) es
$$\frac{1}{s_n^2} \sum_{i=1}^{n} \mathrm{E}\left[Y_i^2\right] = \frac{1}{n} \sum_{i=n_0}^{n} \mathrm{E}\left[X_i^2\right] = \frac{n - n_0}{n} \geq 1 - \delta.$$
Esto muestra que la condición (5.28) no se satisface y el TCL no es válido.
- $\alpha = 1/2$. En este caso $n_0/n \sim \varepsilon^2$, que no depende de n y por lo tanto un argumento similar al del inciso anterior muestra que la condición (5.28) tampoco se satisface y el TCL no es válido.

2. Consideremos la sucesión de v.a.i. con distribución

$$P(X_k = k^\alpha) = P(X_k = -k^\alpha) = \frac{1}{2}, \quad \alpha > 0,$$

y queremos determinar si se satisfacen la LDGN y el TCL . En el primer caso hay que verificar si se satisface la condición (5.22). Por simetría tenemos que $\mathrm{E}(X_k) = 0$ y $\sigma_k^2 = \mathrm{Var}(X_k) = k^{2\alpha}$. En consecuencia

$$s_n^2 = \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} k^{2\alpha} \sim \int_0^n x^{2\alpha}\, dx = \frac{n^{2\alpha+1}}{2\alpha + 1}.$$

Ahora podemos verificar si (5.22) vale:

$$\frac{1}{n^2}s_n^2 \sim \frac{n^{2\alpha-1}}{2\alpha+1}$$

y cuando $n \to \infty$ esto va a 0 si $\alpha < 1/2$. Por lo tanto la LDGN se satisface si $\alpha < 1/2$.

Para el TCL necesitamos verificar si la condición de Lindeberg vale y para esto usaremos la versión que nos da la ecuación (5.28). Sabemos que $Y_i = 0$ siempre que $|X_i| = i^\alpha < \varepsilon s_n \sim \big(\varepsilon n^{2\alpha+1}/(1+2\alpha)\big)^{1/2}$. Llamemos $n_0 = \big(\varepsilon n^{2\alpha+1}/(1+2\alpha)\big)^{1/2}$ entonces

$$\frac{n_0}{n} \sim \frac{\varepsilon^{1/2} n^{1+1/2\alpha}}{(1+2\alpha)^{1/2\alpha} n} \to \infty$$

Por lo tanto, para n suficientemente grande, $i \le n < \varepsilon s_n$ y $Y_i = 0$. Esto muestra que la condición (5.28) se satisface y el TCL vale.

▲

Por último presentamos el teorema de Berry-Esseen, que nos da información sobre la distancia entre la distribución de la suma estandarizada y la función de distribución normal típica.

Teorema 5.12 (Berry-Esseen) *Sea $X_n, n \ge 1$ una sucesión de v.a.i.i.d. con media 0, varianza $\sigma^2 > 0$ y $\mathrm{E}(|X_i^3|) = \rho < \infty$. Sea F_n la función de distribución de $S_n/\sqrt{n}\sigma$. Entonces, para todo $x \in \mathbb{R}$,*

$$\big|F_n(x) - \Phi(x)\big| \le \frac{3\rho}{\sigma^3\sqrt{n}}.$$

5.4.1. Aplicaciones

Ejemplo 5.20 (Error de redondeo)
En este ejemplo retomamos el problema del error de redondeo. Supongamos que redondeamos una serie de números al entero más cercano. El error que se comete al redondear el k-ésimo número es una variable X_k que tiene distribución uniforme $\mathcal{U}(-0.5, 0.5)$. Si sumamos n de estos números, el error que cometemos será $S_n = X_1 + X_2 + \cdots + X_n$, cuya distribución no es sencilla de determinar. Sin embargo, podemos recurrir al Teorema Central de Límite para aproximar la distribución de esta variable. En los ejemplos 5.1 vimos que $\mu = \mathrm{E}(X_k) = 0$ y $\sigma^2 = \mathrm{Var}(X_k) = 1/12$.

Si queremos hallar la probabilidad de que el error cometido sea mayor que α tenemos

$$\begin{aligned} P(S_n > \alpha) &= P\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{\alpha - n\mu}{\sigma\sqrt{n}}\Big) \\ &= P\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{\sqrt{12}\alpha}{\sqrt{n}}\Big) \\ &\approx 1 - \Phi\Big(\frac{\sqrt{12}\alpha}{\sqrt{n}}\Big) \end{aligned}$$

Para ver un caso concreto, supongamos que sumamos 50 números redondeados al entero más cercano y queremos hallar la probabilidad de que el error que cometemos sea menor o igual que 5, es decir, queremos hallar

$$p = P(|S_{50}| \le 5) \approx \Phi\Big(\frac{5\sqrt{12}}{\sqrt{50}}\Big) = 0.9928$$