Teorema Central del Limite

Jimena Blaiotta
D.N.I. 29905968
jblaiotta@yahoo.com.ar
Guido Spano 933, Lanus
C.P. 1824
L.U.298/02

Pablo Delieutraz
D.N.I. 25227711
pdelieutraz@yahoo.com.ar
M. T. de Alvear 1265, Cap. Fed.
C.P. 1058
L.U. 229/99

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

30 de julio de 2004

Índice

1. Introducción

El Teorema Central de Límite no es un único teorema, sino que consiste en un conjunto de resultados acerca del comportamiento de la distribución de la suma (o promedio) de variables aleatorias.

Con Teorema Central del Límite nos referiremos a todo teorema en el que se afirma, bajo ciertas hipótesis, que la distribución de la suma de un número muy grande de variables aleatorias se aproxima a una distribución normal.

El término "Central", debido a Polyá (1920), significa fundamental, o de "ìmportancia central", este describe el rol que cumple este teorema en la teoría de probabilidades. Su importancia radica en que este conjunto de teoremas desvelan las razones por las cuales, en muchos campos de aplicación, se encuentran en todo momento distribuciones normales, o casi normales.

Un ejemplo típico de este hecho es el caso de los errores de medida. Con respecto a este tema, Laplace propuso una hipótesis que parece ser plausible. Considera el error total como una suma de numerosos errores elementales muy pequeños debidos a causas independientes.

Es casi indudable que varias causas independientes o casi independientes contribuyen al eror total. Así por ejemplo, en las observaciones astronómicas, pequeñas variaciones de temperatura, corrientes irregulares de aire, vibraciones de edificios y hasta el estado de los órganos de los sentidos de un observador, pueden considerarse como algunas pocas de dichas causas numerosas.

El Teorema Central del Límite es obra de muchos grandes matemáticos.

Dentro de la historia del Teorema Central del Límite Laplace ocupa un lugar fundamental: a pesar de que nunca enunció formalmente este resultado, ni lo demostró rigurosamente, a él le debemos este importante descubrimeiento.

2. Definiciones Básicas

Esta sección tiene como objetivo establecer la terminologia y definiciones básicas para evitar interrupciones posteriores y lograr una lectura más fluída.

Definición 1

Una variable aleatoria (v.a.) X se dice Bernoulli si su función de probabilidad esta dada por:

$p(0) = P(X = 0) = 1 - p$
$p(1) = P(X = 1) = p$
para algún $p \in (0, 1)$

Definición 2

Supongamos que se realizan n ensayos independientes, y que en cada uno se tienen dos resultados posibles: éxito y fracaso. Supongamos también que p es la probabilidad de éxito de cada ensayo.

Si X representa el número de éxitos que ocurren en los n ensayos, entonces Xse dice una variable aleatoria binomial (o que tiene distribución binomial) con parámetros (n, p)

Obs. 1: una variable aleatoria Bernoulli es simplemente una binomial de parámetros $(1, p)$

Obs. 2: una v.a. binomial de parámetros (n, p) es la suma de n variables aleatorias Bernoulli con probabilidad $p = P(X_i = 1)$

Obs. 3: si X es una variable aleatoria binomial con parámetros (n, p) notaremos

$$b(k; n, p) = P(X = k)$$

Definición 3

Se llama función de densidad normal a

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

Definición 4

La función de distribución normal es

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

Definición 5

Sea X una v.a., se llama función característica de X a

$$\varphi_X(t) = \mathrm{E}(e^{itX}) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$$

Definición 6

• El momento de orden n es $\mathrm{E}(X^n)$

• El momento absoluto de orden n es $\mathrm{E}(|X|^n)$

• El n-ésimo momento centrado de X es el momento de orden n de $X - \mathrm{E}(X)$, es decir $\mathrm{E}((X - \mathrm{E}(X))^n)$.

3. Aproximación normal de la distribución binomial

3.1. Primera versión del Teorema Central del Límite.

La aproximación normal a la distribución binomial tiene considerable valor teórico y práctico.

Desempeñó un papel importante en el desarrollo de la teoría de probabilidades, ya que condujo al primer teorema del límite.

Esta primer versión del Teorema Central del Límite fue dada por De Moivre en su libro "The Doctrine of Chances" (1733) [1], para el caso especial $p = \frac{1}{2}$. Laplace generalizó al caso p arbitrario y el resultado se enuncia como sigue:

[1] De Moivre escribió su libro "The Doctrine of Chances" en inglés pues en esa época residía en Inglaterra, tras haber escapado de Francia por la persecución de los Protestantes. El título del libro es un sinónimo de "Teoría de la Probabilidad" en concordancia con la frase usada por Bayes en su famosa obra póstuma "An Essay Toward Solving a Problem in the Doctrine of Chances"

Teorema 3.1 (De Moivre-Laplace) *Si $n \to \infty$ y k está restringido a un intervalo $k < K_n$ tal que $K_n^3/n^2 \to 0$, entonces para todo $\epsilon > 0$ y n suficientemente grande se cumple:*

$$1 - \epsilon < \frac{a_k}{h\phi(kh)} < 1 + \epsilon \qquad (1)$$

donde

$$a_k = b(k + m; n, p)$$

donde m es el término central, es decir, el único entero de la forma

$$m = np + \delta \qquad con - q < \delta \leq p$$

$$h = \frac{1}{\sqrt{npq}}$$

Este teorema se puede enunciar de una forma más clara, aunque menos precisa de la siguiente forma[2] :

$$b(k; n, p) \approx \frac{1}{\sqrt{2\pi npq}}\ e^{-\frac{u^2}{2}} \qquad (2)$$

donde $u = \frac{k-np}{\sqrt{npq}}$

Nota: observar que el k al que hace referencia el teorema 3.1 no es el mismo que el de la fórmula (2)

3.2. Uso de la aproximación normal a la binomial.

La aplicación más importante del teorema 3.1 consiste en obtener aproximaciones para las probabilidades de la forma

$$P\big(\alpha \leq X \leq \beta\big) = \sum_{v=1}^{\beta} b(v; n, p) = \sum_{k=\alpha-m}^{\beta-m} a_k$$

donde X es una variable aleatoria con distribución binomial de parámetros n y p, y m es el término central.

Dentro de los límites de aplicación del teorema 3.1 obtenemos una buena aproximación cuando reeplazamos a_k por $h\phi(kh)$. Esta cantidad puede interpretarse como el área de un rectángulo de altura $\phi(kh)$, cuya base es un intervalo de longitud h con centro en kh. Como es habitual, reeemplazamos el área del rectángulo por el área correspondiente situada entre el eje x y la gráfica de ϕ; como es sabido, el error que se comete es despreciable en el límite cuando $h \to 0$. Cuando α y β son enteros, llegamos a la aproximación

$$P\big(\alpha \leq X \leq \beta\big) \approx \Phi\big((\beta - m + \frac{1}{2})h\big) - \Phi\big((\alpha - m - \frac{1}{2})h\big)$$

Sin embargo en la formulación final es preferible reemplazar $z_1 = \frac{\alpha-np}{\sqrt{npq}}$ y $z_2 = \frac{\beta-np}{\sqrt{npq}}$; el error que se comete con esta aproximación tiende obviamente a cero junto con h. Por lo tanto hemos demostrado el siguiente

[2]Esta aproximación es buena cuando $npq \geq 30$

Teorema 3.2 *Para z_1 y z_2 fijos, cuando $n \to \infty$*

$$P\big(np + z_1\sqrt{npq} \leq X \leq np + z_2\sqrt{npq}\big) \to \Phi(z_2) - \Phi(z_1) \tag{3}$$

La relación de límite adquiere una forma más agradable si reemplazamos el término X por

$$X^* = \frac{X - np}{\sqrt{npq}}$$

Así podemos poner (3) como

$$P\big(z_1 \leq X^* \leq z_2\big) \to \Phi(z_2) - \Phi(z_1)$$

Observación: al usar la aproximación normal a la distribución binomial, estamos aproximando la distribución de una variable aleatoria discreta con la distribución de una variable aleatoria continua. Por lo tanto hay que tener algún cuidado con los puntos extremos del intervalo considerado. Por ejemplo, para una variable aleatoria continua, $P(X = 3) = 0$, mientras que para una variable aleatoria discreta esta probabilidad puede ser positiva.

Se ha encontrado que la siguiente *corrección para continuidad* mejora la aproximación anterior:

- $P(X = k) \simeq P(k - \frac{1}{2} \leq X \leq k + \frac{1}{2})$
- $P(a \leq X \leq b) \simeq P(a - \frac{1}{2} \leq X \leq b + \frac{1}{2})$

APROXIMAR!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

4. Teorema Central del Límite de Laplace

En esta sección daremos una idea de la forma en que trabajó Laplace.

El resultado obtenido por Laplace alrededor de 1810, en notación moderna, es el siguiente[3] :

Teorema 4.1 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes, identicamente distribuidas (iid) y acotadas, discretas o absolutamente continuas. Y sean $\mu = \mathrm{E}(X_1)$ y $\sigma^2 = \mathrm{Var}(X_1)$. Si n es suficientemente grande, vale la siguiente aproximación:*

$$P(n\mu + r_1\sqrt{n} \leq X_1 + \ldots + X_n \leq n\mu + r_2\sqrt{n}) \approx \frac{1}{\sqrt{2\pi}} \int_{r_1}^{r_2} \frac{1}{\sigma} e^{-\frac{t^2}{2\sigma^2}} \, dt$$

Observar que el teorema de De Moivre-Laplace es un caso muy particular del teorema 4.1 si pensamos a la variable aleatoria binomial X como suma de variables aleatorias independientes Bernoulli X_i.

[3] A menudo nos referiremos a este teorema como el T.C.L. de Laplace, ya que, a pesar de no haber establecido explícitamente las hipótesis, como tampoco haber dado una demostración totalmente rigurosa, a él se debe este importante descubrimiento

4.1. Suma de variables aleatorias independientes

La suma de variables aleatorias independientes jugó un papel importante en el trabajo probabilístico de Laplace desde sus más tempranos comienzos. En uno de sus primeros papers publicados, Laplace (1776) apuntó a calcular la probabilidad de que la suma de los ángulos de inclinación de las órbitas de los cometas no sobrepase un cierto valor o, que la media aritmética de esos ángulos esté entre dos límites dados. Él supuso que la distribución de las medidas de los ángulos era aleatoria como así también supuso tácitamente que todos los ángulos eran estocásticamente independientes. Laplace tuvo éxito en calcular estas probabilidades para un número arbitrario de cometas.

En el caso de un número considerable de cuerpos celestes las fórmulas obtenidas eran demasiado complicadas para una evaluación numérica directa. En esta etapa de su trabajo matemático Laplace no pudo desarrollar aproximaciones usables.

4.2. Herramientas utilizadas por Laplace

4.2.1. Método de Laplace para aproximar integrales

En su trabajo "Mémoire sur la probabilité des causes"(1774), Laplace desarrolló técnicas para aproximar integrales que dependen de un "gran parámetro", como por ejemplo la función Gamma

$$\Gamma(s+1) = \int_0^\infty e^{-x} x^s dx$$

donde s es el "gran parámetro".

La idea básica del "método de aproximación laplaciano"es la siguiente:

Sea f el integrando que depende de un "gran parámetro", y supongamos que f tiene un único máximo muy pronunuciado de manera tal que la integral sobre un intervalo pequeño alrededor de este máximo contribuye considerablemente al resultado de la integral. Entonces se puede esperar que la función f sea asintóticamente igual a una función de la forma

$$f(a)e^{-\alpha(x-a)^{2k} \pm \ldots}$$

si f alcanza su máximo en $x = a$. Basado en esta idea, el método de Laplace consiste en desarrollar en series apropiadas alrededor de la abscisa del máximo.

En el caso de muchas fórmulas probabilísticas este método de aproximación funcionaba bastante bien.

En su artículo de 1774, Laplace trató problemas de aproximación con un estilo analítico muy parecido al de Euler.

En su trabajo posterior dejó el estilo euleriano y desarrolló, influenciado por el análisis algebraico de Lagrange, un estilo especial algoritmico-algebraico trabajando principalmente con desarrollos en series formales. La deducción de Laplace del T.C.L. fue escrita en este estilo.

4.2.2. Funciones características

La herramienta de Laplace que fue crucial para su éxito en aproximar distribuciones de sumas de variables aleatorias independientes por distribuciones normales fue su modificación de las funciones generatrices.

Mostraremos la esencia de su procedimiento, en el caso especial de variables aleatorias idéntica y simetricamente distribuidas $X_1, \dots, X_n$ las cuales toman los valores $\frac{k}{m}$ (m es un número natural dado y $k = -m, \dots, m$) con las respectivas probabilidades $p_k = p_{-k}$.

Para calcular la probabilidad

$$P_j = P\Big(\sum_{i=1}^{n} X_i = \frac{j}{m}\Big) \qquad\qquad (-nm \leq j \leq nm)$$

Laplace utilizó la función generatriz $T(t) = \sum_{k=-m}^{m} p_k t^k$. Debido a la independencia de las variables X_i (presupuesta tacitamente por Laplace) P_j es igual al coeficiente de t^j en $[T(t)]^n$ después de desarrollar el producto. La ejecución directa de este método es complicada. Laplace introdujo el truco de sustituir la variable t por e^{ix} (donde $i^2 = -1$). De esta forma surgen, para un caso particular, lo que hoy llamamos funciones características.

Usando la igualdad

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx} e^{isx} dx = \delta_{ts} \qquad (t, s \in \mathbb{Z})$$

se deduce

$$P_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijx}\Big[\sum_{k=-m}^{m} p_k e^{ikx}\Big]^n dx$$

4.3. La deducción de la aproximación a la distribución normal

La última integral era, al menos formalmente, accesible para aplicar los métodos de aproximación de integrales de Laplace. Sin embargo era necesario una modificación, ya que Laplace no consideró el desarrollo de todo el integrando alrededor de su máximo en $x = 0$, sino que solo lo hizo para el factor

$$\big[\sum_{k=-m}^{m} p_k e^{ikx}\big]^n$$

que es la función característica de $X_1 + \cdots + X_n$

Desarrollando e^{ikx} en series de potencias se obtiene:

$$P_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijx}\Big[\sum_{k=-m}^{m} p_k e^{ikx}\Big]^n dx =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijx}\Big[\sum_{k=-m}^{m} p_k(1 - k^2x^2 - \frac{ik^3x^3}{6} + \cdots\Big]^n dx$$

Teniendo en cuenta que $p_k = p_{-k}$ y con la abreviación $m^2\sigma^2 = \sum_{k=-m}^{m} p_k k^2$ obtenemos:

$$P_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijx}\Big[1 - \frac{m^2\sigma^2x^2}{2} + \cdots\Big]^n dx$$

La expansión formal de

$$\log\Big[1-\frac{m^2\sigma^2x^2}{2}+\cdots\Big]^n =: \log z$$

en una serie de potencias de x, da lugar a

$$\log(z) = -\frac{m^2\sigma^2nx^2}{2}-\frac{m^4\sigma^4nx^4}{8}+\cdots,$$

y por lo tanto

$$z = e^{-\frac{m^2\sigma^2nx^2}{2}-\frac{m^4\sigma^4nx^4}{8}+\cdots} = e^{-\frac{m^2\sigma^2nx^2}{2}}\big(1-\frac{m^4\sigma^4nx^4}{8}+\cdots\big)$$

Después de la sustisución $x=\frac{y}{\sqrt{n}}$ resulta

$$P_j = \frac{1}{2\pi\sqrt{n}}\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} e^{-ij\frac{y}{\sqrt{n}}}e^{-\frac{m^2\sigma^2nx^2}{2}}\big(1-\frac{m^4\sigma^4nx^4}{8}+\cdots\big)dy$$

Para una aproximación con n grande, ignoramos, al igual que Laplace algunos términos, y al mismo tiempo cambiamos los límites de integración, obteniendo:

$$P_j \approx \frac{1}{2\pi\sqrt{n}}\int_{-\infty}^{\infty} e^{-ij\frac{y}{\sqrt{n}}}e^{-\frac{m^2\sigma^2nx^2}{2}}dy$$

esta última integral es igual a

$$\frac{1}{m\sigma\sqrt{2\pi n}}e^{-\frac{j^2}{2m^2\sigma^2n}}$$

La suma de esta última expresión desde $r_1\sqrt{n}$ hasta $r_2\sqrt{n}$, que puede ser aproximada por una integral, da lugar a una aproximación para

$$P\big(r_1\sqrt{n}\leq\sum X_i\leq r_2\sqrt{n}\big)$$

que se corresponde con el enunciado del teorema 4.1.

Según Laplace es posible generalizar a variables absolutamente continuas suponiendo que m es "infinitamente grande". Su intento por demostrar esta importante proposición no resiste un análisis riguroso moderno y, por otra parte, no se puede hacer esta demostración en forma rigurosa.

5. El Teorema Central del Límite durante el período 1810-1853

La versión del T.C.L. de Laplace sirvió principalmente como una herramienta de "sentido común", y su importancia fue determinada por un campo fuera de las matemáticas.

A mediados del siglo XIX el T.C.L. se hizo parte de la propia matemática a través de las contribuciones de Dirichlet y Cauchy.

A pesar de los intentos realizados por diversos matemáticos, durante este período no se encuentran demostraciones rigurosas del T.C.L. de Laplace. Esto solo se lograría después de muchos años.

5.1. Siméon Denis Poisson (1824)

Poisson generalizó el T.C.L. de Laplace a sumas y combinaciones lineales de errores observacionales con diferentes distribuciones (no necesariamente simétricas). Señaló la importancia de las presuposiciones, a diferencia de Laplace, que nunca las enunció explicitamnete.

Poisson trabajó principalmente con contraejemplos (de un teorema no formulado claramente por Laplace), el más importante de ellos es el siguiente:

La suma de variables aleatorias X_i independientes, identicamente distribuídas (i.i.d.) con función de densidad

$$f(x) = \frac{1}{\pi(1+x^2)}$$

no se aproxima a una distribución normal, aún cuando la cantidad de sumandos sea muy grande. Obviamente, alguna de las hipótesis del que hoy conocemos como teorema de Laplace (teorema 4.1), no se cumple. Se puede demostrar facilmente que la $\mathrm{E}(X_i)$ no existe.

Poisson, también observó que estos casos no se encuentran en la práctica. Las variables aleatorias con función de densidad f jugarían un rol muy importante en el trabajo posterior de Cauchy, es por este motivo que hoy lleva su nombre.

5.2. Peter Gustav Lejeune Dirichlet (1846)

Dirichlet es conocido por sus contibuciones pioneras a la física matemática y a la teoría de números. En el campo de la teoría de las probabilidades solamente se pueden encontrar unas breves notas, aunque durante su período en Berlín (1828-1855) frecuentemente dictó cursos en teoría de las probabilidades o teoría del error, donde presentó nuevas y originales ideas.

El interés principal de Dirichlet no era el tratramiento de las aplicaciones o fundamentos probabilísticos, sino mas bien, la discusión de los problemas análiticos generados.

Él consideró estos problemas como aplicaciones de la teoría de integrales definidas, y por lo tanto a varios de sus cursos concernientes a estos temas los llamó "Anwendungen der Integralrechnung". Por ejemplo, la "demostración" del T.C.L. fue presentada en 1 hora extra de un curso de 4 hs. sobre integrales definidas.

5.3. Augustin Louis Cauchy (1853)

El último artículo de Cauchy, en un total de 8 papers, contien una interesante discusión sobre la aproximación a la distribución normal de combinaciones lineales de errores aleatorios. Básicamente, su línea de argumentación análitica es similar a la de Dirichlet y muestra métodos que aún son utilizados en el tratamiento moderno del T.C.L.. Cauchy utilizó funciones características y teoremas de inversión para demostrar el caso de variables aleatorias continuas i.i.d. con función de densidad simétrica y soporte compacto.

6. El Teorema Central del Límite y sus primeras demostraciones rigurosas

6.1. Teorema Central del Límite de Liapunov

Las primeras demostraciones realmente rigurosas de T.C.L. son el resultado de la labor de tres grandes matemáticos rusos: Tshebyshev (1887), Markov (1898) y Liapunov (1900-1901).

Tshbyshev y Markov lo hicieron utilizando el método de los momentos. Liapunov fue el primero en utilizar el método de las funciones características. Él, ha demostrado por este método, que el T.C.L. era aplicable con hipótesis mucho más generales que los de Tshbyshev y Markov [4]además su método de demostración tiene la ventaja de la simplicidad.

El resultado es el siguiente:

Teorema 6.1 (Liapunov) *Sean $(X_k)_{k\in\mathbb{N}}$ v.a. independientes cuyos tres primeros momentos existen, y sean: $m_k = E(X_k)$, $\sigma^2 = Var(X_k) < \infty$*

$a_k = E((X_k - m_k)^3)$momento centrado de tercer orden

$b_k = E(|X_k - m_k|^3)$momento centrado absoluto de tercer orden

Llamemos

$$s_n := \sqrt{\sum_{k=1}^{n} \sigma_k^2} \qquad B_n := \sqrt[3]{\sum_{k=1}^{n} b_k}$$

Si se verifica la siguiente condición (condición de Liapunov):

$$\lim_{n\to\infty} \frac{B_n}{s_n} = 0 \tag{4}$$

entonces

$$\lim_{n\to\infty} F_n(x) = \Phi(x)$$

donde F_n es la función de distribución de

$$S_n^* = \frac{\sum_{k=1}^{n}(X_k - m_k)}{s_n}$$

Observación: si las v.a. $(X_k)_{k\in\mathbb{N}}$ son identicamente distribuídas, entonces la condición de Liapunov se cumple. En efecto, en este caso son iguales todos los desvíos σ_k (llamemos $\sigma = \sigma_k$), y también son iguales todos los momentos b_k (llamemos $b = b_k$), entonces resulta

$$s_n = \sigma\sqrt{n}$$

$$B_n = \sqrt[3]{b}\sqrt[3]{n}$$

Luego

$$\lim_{n\to\infty} \frac{B_n}{s_n} = \frac{\sqrt[3]{b}}{\sigma} \lim_{n\to\infty} \frac{1}{\sqrt[6]{n}} = 0$$

[4]Markov señaló a continuación que se puede demostrar también el teorema de Liapunov por el método de los momentos.

Aunque para variables identicamente distribuídas la condición de Liapunov se cumple, observemos que no se puede deducir el teorema de Laplace a partir del teorema de Liapunov, ya que puede existir

$$\int_{-\infty}^{+\infty} x^2 dF(x)$$

pero

$$\int_{-\infty}^{+\infty} |x|^3 dF(x) = +\infty$$

e incluso que sea

$$\int_{-\infty}^{+\infty} |x|^{2+\delta} dF(x) = +\infty \qquad \forall\ \delta > 0.$$

Liapunov también ha demostrado el T.C.L. con hipótesis más débiles. Para esto supuso la existencia, en lugar del tercer momento, del momento de orden $2+\delta$ $(\delta > 0)$ y en lugar de la condición de Liapunov exigió

$$\lim_{n\to\infty} \frac{B_n(\delta)}{s_n} = 0 \tag{5}$$

donde

$$B_n(\delta) = \left(\sum_{k=1}^{n} \mathrm{E}\big(|X_k - m_k|^{2+\delta}\big) \right)^{\frac{1}{2+\delta}}.$$

6.2. Teorema Central del Límite de Lindeberg

Lindeberg propuso una condición aún más general que la condición (4) la cual permite demostrar el T.C.L. Esta condición es también, en cierto sentido necesaria.

Teorema 6.2 (Lindeberg) *Sean $X_1, X_2, \ldots, X_k, \ldots$ variables aleatorias independientes, cuyas medias μ_k =$E(X_k)$ y desvíos $\sigma_k = \sqrt{Var(X_k)}$ existen. Llamemos*

$$s_n := \sqrt{\sum_{k=1}^{n} \sigma_k^2}$$

y $F_k(x)$ a la función de distribución de $X_k - \mu_k$. Si $\forall\ \epsilon > 0$, se cumple la siguinte condición (condición de Lindeberg)

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1} n \int_{|x|>\epsilon s_n} x^2 dF_k(x) = 0 \tag{6}$$

entonces

$$\lim_{n\to\infty} G_n(x) = \Phi(x)$$

donde G_n es la función de distribución de

$$S_n^* = \frac{\sum_{k=1}^{n} (X_k - \mu_k)}{s_n}$$

En los libros modernos de probabilidad, este teorema se demuestra usando funciones características. Lindeberg en su demostración no las utilizó.

Observación: de la condición de Liapunov (4) se deduce la condición de Lindeberg (6), en efecto:

$$\frac{1}{s_n^2}\sum_{k=1} n\int_{|x|>\epsilon s_n} x^2 dF_k(x) \leq \frac{1}{\epsilon s_n^2}\sum_{k=1} n\int_{-\infty}^{+\infty} |x|^3 dF_k(x) = \frac{1}{\epsilon}\left(\frac{B_n}{s_n}\right)^3$$

La condición de Lindeberg también se puede deducir a partir de (5).

Por lo tanto para demostrar el teorema de Liapunov, basta demostrar el teorema de Lindeberg (para una demostración ver Rényi [7]).

6.3. Teorema Central del Límite de Lindeberg-Lévy

Teorema 6.3 (Lindeberg-Lévy) *Sean $X_1, X_2, \ldots, X_k, \ldots$ variables aleatorias i.i.d. con $\mu = E(X_k) < \infty$ y varianza positiva $\sigma^2 = Var(X_k) < \infty$. Y sean*

$$S_n^* = \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}}$$

y F_n las funciones de distribución de S_n^*.

Entonces

$$\lim_{n\to\infty} F_n(x) = \Phi(x)$$

Para la demostración de este teorema, necesitamos el siguiente lema previo que enunciamos sin demostración.

Lema: Si la sucesión de funciones características $(\varphi_n(t))$ converge en cada punto t a una función $\varphi(t)$ continua en algún intervalo $|t| < \tau$, entonces la sucesión $F_n(x)$ de las correspondientes funciones de distribución converge a la función de distribución $F(x)$ que corresponde a la función característica $\varphi(t)$.

Demostración del Teorema 6.3. Todas las variables $X_k - \mu$ tienen la misma distribución, por lo tanto también tienen la misma función característica $\varphi_X(t)$. Además:

$$\mathrm{E}(X_k - \mu) = 0 \qquad \text{y} \qquad \mathrm{Var}(X_k - \mu) = \sigma^2$$

Debido a las dos siguientes propiedades:

1. Si X e Y son variables aleatorias independientes, entonces:

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

2. $\varphi_{aX}(t) = \varphi_X(at)$

de acuerdo a esto, la función característica $\varphi_{S_n^*}$ de S_n^* es

$$\varphi_{S_n^*}(t) = \left[\varphi_X\Big(\frac{t}{\sigma\sqrt{n}}\Big)\right]^n \tag{7}$$

El desarrollo en serie de Taylor de $\varphi_X(t)$ alrededor de $t = 0$ es:

$$\varphi_X(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2) \tag{8}$$

Reemplazando la expresión (8) en la fórmula (7) se obtiene:

$$\varphi_{S_n^*}(t) = \left[1 - \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\right]^n$$

donde $\forall\ t$ tenemos:

$$\lim_{n\to\infty} no\Big(\frac{t^2}{n}\Big) = 0 \tag{9}$$

Sea

$$u = -\frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)$$

Entonces

$$\log\varphi_{S_n^*}(t) = n\log(1+u) = n\Big[-\frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big] = -\frac{t^2}{2} + no\Big(\frac{t^2}{n}\Big)$$

De la relación (9) tenemos que

$$\lim_{n\to\infty} \varphi_{S_n^*} = e^{-\frac{t^2}{2}}$$

Usando el lema previo, queda demostrado el teorema. ■

Lindeberg (1922) dió la primera demostración rigurosa de un teorema más general (teorema 6.2) que contiene al teorema 6.3, en efecto: si todas las variables X_k tienen la misma distribución, con desvío finito σ, tenemos

$$F(x) = F_k(x), \qquad \int_{-\infty}^{+\infty} x^2 dF(x) = \sigma^2 \qquad \text{y } s_n = \sigma\sqrt{n}$$

De donde

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x|>\epsilon s_n} x^2 dF_k(x) = \frac{1}{\sigma^2} \lim_{n\to\infty} \int_{|x|>\epsilon s_n} x^2 dF_k(x) = 0$$

El teorema 6.3 es básicamente una reformulación del T.C.L. de Laplace (quien no dió una demostración rigurosa así como tampoco estableció claramente las hipótesis del teorema).

Lévy en su demostración del teorema 6.3 (1930-1935) reinventó la técnica de las funciones características. Existen razones históricas de que Lévy no conocía el trabajo de Laplace: los libros franceses de esa época (Bertrand (1889-1907), Poincaré (1896-1912), Borel (1909-1924)) no mencionaban a Laplace ni a las funciones características, de hecho Lévy dudó en usarlas en su trabajo debido a las críticas de Borel.

6.4. Teorema Central del Límite de Lindeberg-Feller

Feller demostró que la condición de Lindeberg es necesaria en el siguiente sentido:

Si valen las dos condiciones siguientes:

$$\lim_{n\to\infty} \max_{1\le k\le n} \Big(\frac{\sigma_k}{s_n}\Big) = 0 \qquad \lim_{n\to\infty} F_n(x) = \Phi(x)$$

entonces se cumple la condición de Lindeberg, es decir

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1} n \int_{|x|>\epsilon s_n} (x)^2 dF_k(x) = 0$$

Más aún, tenemos el siguiente resultado:

Teorema 6.4 (Lindeberg-Feller) *Sean $X_1, X_2, \ldots, X_k, \ldots$ variables aleatorias independientes cuyas varianzas existen, y sea $G_k(x)$ la función de distribución de de X_k, $\mu = E(X_k)$, $\sigma_k = \sqrt{Var(X_k)}$, llamemos $F_n(x)$ a la función de distribución de*

$$S_n^* = \frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n} \qquad \text{donde } s_n = \sum_{k=1}^n \sigma^2$$

Entonces las relaciones

$$\lim_{n\to\infty} max_{1\le k\le n}(\frac{\sigma_k}{s_n}) = 0 \qquad \lim_{n\to\infty} F_n(x) = \Phi(x)$$

valen, si y solo si, $\forall \epsilon > 0$

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dG_k(x) = 0$$

De este último teorema se puede deducir otra condición necesaria y suficiente:

Teorema 6.5 (Teorema) *Sean $X_1, X_2, \ldots, X_k, \ldots$ variables aleatorias independientes uniformemente acotadas, es decir: existe una constante $a > 0$ tal que $\forall k$*

$$P(|X_k| \le a) = 1$$

y supongamos que $Var(X_k) \neq 0$ $\forall k$.

Entonces

$$\lim_{n\to\infty} F_n(x) \Leftrightarrow \lim_{n\to\infty} s_n^2 = \infty$$

7. Generalizaciones del Teorema Central del Límite

7.1. El Teorema Central del Límite para variables dependientes

Markov fue el primero en intentar relajar la condición de independencia en el T.C.L. y consideró variables aleatorias dependientes, las cuales dan paso a su teorema “Cadenas de Markov”.

El trabajo de Bernstein (1926) sobre variables dependientes puede considerarse como una extraordinaria contribución moderna a la Teoría de la Probabilidad, sin embargo, las condiciones establecidas por Bernstein para la validez de los teoremas son bastante complicadas.

7.2. El Teorema Central del Límite para vectores aleatorios

El Teorema Central del Límite para variables aleatorias puede ser generalizado para vectores aleatorios. La primer demostración rigurosa del TCL para el caso de vectores aletorios independientes fue publicada por Bernstein en 1926. Su demostración sigue la línea de la demostración de Liapunov para sumas de variables aleatorias independientes. Más aún, Bernstein mostr'0 que el TCL es válido en el caso de vectores aleatorios dependientes cuando se cumplen cieras condiciones adicionales. Laplace en 1810 ya había señalado la posibilidad de tales extenciones.

Referencias

[1] Julie Xiuyu Cong: *Historical Development of Central Limit Theorem*; 2003

[2] W. Feller: *Introducción a la teoría de las probabilidades y sus aplicaciones*; Limusa, 1996

[3] H. Fischer: *The Central Limit Theorem from Laplace to Cauchy*; 2000

[4] M. Fisz: *Probability Theory and Mathematical Statistics*; John Wiley & Sons, Inc., 1963

[5] P. Meyer: *Probabilidad y aplicaiones estadísticas*; Fondo Educativo Interamericano, S.A., 1963

[6] F. Mosteller; R. Rourke; G. B. Thomas, Jr: *Probability and Statistics*; Addison-Wesley Publishing Company, Inc., 1961

[7] A. Rényi: *Cálculo de Probabilidades*; Editorial Reverté, S.A., 1976

[8] S. Ross: *A First Course in Probability*; Prentice Hall, Inc., cuarta ed., 1994

[9] J. V. Uspensky: *Introduction to Mathematical Probability*; Mc Graw-Hill Book Company London, 1937