TEORÍA DE LA PROBABILIDAD

Convergencia de variables aleatorias

Introducción

Desde los inicios del Cálculo Diferencial e Integral se planteó el problema de expresar una función como una serie de funciones simples (recordemos, por ejemplo, el desarrollo de una función en serie de Taylor). Más adelante, se planteó el problema de expresar una función como una serie trigonométrica. En particular, recordemos que Fourier, en el año 1822, afirmó que una función arbitraria f (con el concepto de función de la época, el cual no era el actual) definida y acotada en el intervalo $[-L, L]$, puede representarse mediante la siguiente serie trigonométrica:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\left[a_n \cos\frac{n\pi x}{L} + b_n sen\frac{n\pi x}{L}\right],$$

donde $a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx$ y, para cada $n \in \mathbb{N}$:

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}dx$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x) sen\frac{n\pi x}{L}dx$$

Esta afirmación de Fourier condujo a investigar a fondo cuándo la serie:

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty}\left[a_n \cos\frac{n\pi x}{L} + b_n sen\frac{n\pi x}{L}\right]$$

con los valores de los coeficientes dados por Fourier, converge efectivamente a la función f.

Las investigaciones alrededor del problema de la convergencia de una serie de funciones produjo la necesidad de profundizar, en general, en el problema de la convergencia de una sucesión de funciones. Surgió en este proceso la definición de la **convergencia uniforme**; esto a raíz de que Cauchy afirmaba que si los términos de una serie son funciones continuas y la serie converge, entonces la función a la cual converge es continua. Abel hizo ver que esta afirmación no es válida en general y Weierstrass, introduciendo el concepto de convergencia uniforme, demostró que la afirmación de Cauchy es válida si la convergencia de la serie es uniforme.

Otro problema relacionado con las series de funciones era el determinar bajo que condiciones se puede integrar término a término una serie convergente de funciones, para obtener la integral de la función a la cual converge.

Al desarrollar Lebesgue su teoría de integración, uno de los aspectos centrales de su planteamiento fue el obtener una definición de integral con la propiedad de que si se tiene una sucesión convergente de funciones integrables, la función límite sea integrable y su integral sea igual al límite de la sucesión formada por las integrales de las funciones que componen la sucesión dada.

La integral que definió Lebesgue tiene la característica de que al modificar los valores de una función integrable en los puntos de un conjunto de medida cero, la función sigue siendo integrable y la integral de la función modificada es igual a la integral de la función original. Es decir, para fines de la integración de funciones, los conjuntos de medida cero son despreciables. En particular, en lo que respecta a la convergencia de una sucesión de funciones, esto lleva a que no es necesario tratar con sucesiones de funciones que converjan en todos los puntos, puede uno limitarse a la convergencia fuera de un conjunto de medida cero. Surgió así el concepto de **convergencia casi en todas partes**.

De esta forma, en el desarrollo de la teoría de integración, incluyendo los trabajos anteriores al de Lebesgue, se fueron encontrando diferentes tipos de convergencia de una sucesión de funciones.

Un tipo de convergencia que surgió en este contexto es la convergencia en medida. Dentro de la teoría de integración la idea surgió del problema de la integración término a término de una serie de funciones, buscando condiciones menos restrictivas que la convergencia uniforme de la serie para asegurar que se puede integrar término a término una serie convergente de funciones para obtener la integral de la función a la cual converge. Sin definirla explicitamente, este tipo de convergencia se encuentra formulada en un artículo de L. Kronecker del año 1878, obviamente sin referirse al concepto de medida, que aún no se había formulado, sino al de contenido. La definición explícita fue dada en el año 1909 por F. Riesz en un artículo titulado *Sur les suites des fonctions mesurables*, donde, entre otras cosas, demostró que si una sucesión de funciones medibles converge en medida a una función medible, entonces existe una subsucesión que converge casi en todas partes.

Sin embargo, la convergencia en medida tiene una historia más antigua; sin definirse explicitamente, era utilizada en el Cálculo de Probabilidades desde la publicación del teorema de Bernoulli en el año 1713, el cual establece la convergencia, que hoy se denomina en medida, de una determinada sucesión de variables aleatorias. Este resultado de Bernoulli marcó la pauta para el desarrollo del CálculoTeoría de Probabilidades hasta principios del siglo XX, cuando se llegó a la formulación general de los llamados teoremas límite, entre los cuales se encuentra la ley débil de los grandes números, la cual es una generalización del resultado de Bernoulli.

En la teoría de la probabilidad, la convergencia casi en todas partes corresponde a la **convergencia con probabilidad 1**, también conocida como **convergencia casi segura**. La convergencia en medida corresponde a la **convergencia en probabilidad**.

En la exposición que sigue, $(\mathbb{E}, \mathcal{E}, \mu)$ será un espacio de medida completo.

Convergencia casi en todas partes

Definición 1. *Diremos que una sucesión de funciones medibles* $\{f_n\}_{n\in\mathbb{N}}$ *converge casi en todas partes a una función medible* f *si* $\lim_{n\to\infty} f_n = f$ *excepto a lo más en un conjunto de medida cero. Si éste es el caso, se escribirá* $f_n \xrightarrow{c.t.p.} f$.

Las siguientes propiedades se siguen inmediatamente de las correspondientes propiedades para las sucesiones de números reales:

1. Si una sucesión $(f_n)_{n\in\mathbb{N}}$ converge casi en todas partes a f, entonces cualquier subsucesión de $(f_n)_{n\in\mathbb{N}}$ también converge casi en todas partes a f.
2. Si $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$ y $f_n \xrightarrow{c.t.p.} g$, entonces $f = g$ casi en todas partes.
3. Si c es una constante y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$, entonces $cf_n \xrightarrow{c.t.p.} cf$.
4. Si $(f_n)_{n\in\mathbb{N}}$ y $(g_n)_{n\in\mathbb{N}}$ son dos sucesiones de funciones medibles tales que $f_n \xrightarrow{c.t.p.} f$ y $g_n \xrightarrow{c.t.p.} g$, entonces $f_n + g_n \xrightarrow{c.t.p.} f + g$ y $f_n g_n \xrightarrow{c.t.p.} fg$.

Los resultados que siguen, con relación a la convergencia casi en todas partes, se entienden mejor si se tienen en mente los siguientes conceptos:

Dada una sucesión $A_1, A_2, \dots$ de subconjuntos de un conjunto $\mathbb{F}$, se define el límite inferior (lím ínf) y el límite superior (lím sup) de esa sucesión de la siguiente manera:

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Obsérvese que $\liminf A_n$ está formado por todos los elementos $x \in \mathbb{F}$ para los cuales existe $N \in \mathbb{N}$ tal que $x \in A_n$ para cualquier $n \geq N$, mientras que $\limsup A_n$ está formado por todos los elementos $x \in \mathbb{F}$ que pertenecen a una infinidad de conjuntos de la sucesión. Así que se tiene siempre $\liminf A_n \subset \limsup A_n$.

Si $\liminf A_n = \limsup A_n$, se dice que la sucesión $(A_n)_{n\in\mathbb{N}}$ converge y al valor común se le llama el límite de la sucesión $(A_n)_{n\in\mathbb{N}}$ y se le denota de la manera usual, $\lim_{n\to\infty} A_n$.

Por ejemplo, si $(A_n)_{n\in\mathbb{N}}$ es una sucesión creciente (resp. decreciente) de subconjuntos de $\mathbb{E}$, entonces converge y $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$ (resp. $\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n$).

Teorema 1. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles. Entonces $f_n \overset{c.t.p.}{\longrightarrow} 0$ si y sólo si para cualquier $\varepsilon > 0$, se tiene:*

$$\mu\left(\limsup \{y \in \mathbb{E} : |f_n(y)| > \varepsilon\}\right) = 0$$

Demostración

Supongamos primero que $\lim_{n\to\infty} f_n = 0$ casi en todas partes. Entonces existe un conjunto $E_0 \subset \mathbb{E}$ de medida 0 tal que si $x \in E_0^c$ entonces $\lim_{n\to\infty} f_n(x) = 0$. Así que, dado $x \in E_0^c$ y $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que $|f_n(x)| \leq \varepsilon$ para cualquier $n \geq N$; esto significa que:

$$x \in \bigcap_{n=N}^{\infty} \{y \in \mathbb{E} : |f_n(y)| \leq \varepsilon\}$$

Dicho de otra forma, si $x \in E_0^c$, entonces, dada cualquier $\varepsilon > 0$:

$$x \in \bigcup_{m=1}^{\infty} \left[\bigcap_{n=m}^{\infty} \{y \in \mathbb{E} : |f_n(y)| \leq \varepsilon\}\right]$$

Así que:

$$E_0^c \subset \bigcup_{m=1}^{\infty} \left[\bigcap_{n=m}^{\infty} \{x \in \mathbb{E} : |f_n(x)| \leq \varepsilon\}\right]$$

Por lo tanto:

$$\mu\left(\bigcap_{m=1}^{\infty} \left[\bigcup_{n=m}^{\infty} \{y \in \mathbb{E} : |f_n(y)| > \varepsilon\}\right]\right) = 0$$

Inversamente, supongamos que, para cualquier $\varepsilon > 0$, se tiene:

$$\mu\left(\bigcap_{m=1}^{\infty} \left[\bigcup_{n=m}^{\infty} \{y \in \mathbb{E} : |f_n(y)| > \varepsilon\}\right]\right) = 0$$

Para cada $r \in \mathbb{N}$, sea:

$$B_r = \bigcap_{m=1}^{\infty} \left[\bigcup_{n=m}^{\infty} \left\{y \in \mathbb{E} : |f_n(y)| > \tfrac{1}{r}\right\}\right]$$

Se tiene $\mu(B_r) = 0$ para cualquier $r \in \mathbb{N}$ y la sucesión de conjuntos $B_1, B_2, \ldots$ es creciente, así que:

$$\mu\left(\bigcup_{r=1}^{\infty} B_r\right) = \lim_{r\to\infty} \mu(B_r) = 0$$

Pero, $B_r^c = \left\{x \in \mathbb{E} : \text{Existe } N(x) \in \mathbb{N} \text{ tal que } |f_n(x)| \leq \frac{1}{r} \text{ para cualquier } n \geq N(x)\right\}$. De manera que si $x \in \bigcap_{r=1}^{\infty} B_r^c$, entonces para cualquier $r \in \mathbb{N}$ existe $N(x) \in \mathbb{N}$ tal que $|f_n(x)| \leq \frac{1}{r}$ para cualquier $n \geq N(x)$. En particular, dada $\varepsilon > 0$ sea $r \in \mathbb{N}$ tal que $\frac{1}{r} < \varepsilon$ y $N(x)$ tal que $|f_n(x)| \leq \frac{1}{r}$ para cualquier $n \geq N(x)$, entonces $|f_n(x)| < \varepsilon$ para cualquier $n \geq N(x)$, lo cual significa que $\lim_{n\to\infty} f_n(x) = 0$. Es decir:

$$\bigcap_{r=1}^{\infty} B_r^c \subset [\lim_{n\to\infty} f_n = 0]$$

Sea $E_0 = \bigcup_{r=1}^{\infty} B_r$. Entonces $\mu(E_0) = 0$ y si $x \in E_0^c = \bigcap_{r=1}^{\infty} B_r^c$, entonces $\lim_{n\to\infty} f_n(x) = 0$. Así que $\lim_{n\to\infty} f_n = 0$ casi en todas partes.

■

Corolario 1. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles. Entonces $f_n \xrightarrow{c.t.p.} f$ si y sólo si para cualquier $\varepsilon > 0$, se tiene:*

$$\mu\left(\limsup \{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\}\right) = 0$$

Teorema 2 (Lema de Borel-Cantelli)**.** *Sea $E_1, E_2, \ldots$ una sucesión de conjuntos medibles tales que $\sum_{n=1}^{\infty} \mu(E_n) < \infty$, entonces:*

$$\mu(\limsup E_n) = 0$$

Demostración

Sea $B = \limsup E_n$.

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} E_n$. Entonces la sucesión de conjuntos B_m es decreciente y $B = \bigcap_{m=1}^{\infty} B_m$, así que:

$$\mu(B) = \mu\left[\bigcap_{m=1}^{\infty} B_m\right] = \lim_{m\to\infty} \mu\left[\bigcup_{n=m}^{\infty} E_n\right] \leq \lim_{m\to\infty} \sum_{n=m}^{\infty} \mu(E_n) = 0$$

■

Corolario 2. *Sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tales que:*

$$\sum_{n=1}^{\infty} \mu\left[|f_n| > \varepsilon\right] < \infty \text{ para cualquier } \varepsilon > 0$$

Entonces $f_n \xrightarrow{c.t.p.} 0$.

Demostración

Dada $\varepsilon > 0$, sea $A(\varepsilon) = \limsup \{y \in \mathbb{F} : |f_n(y)| > \varepsilon\}$.

Por la proposición 2, $\mu[A(\varepsilon)] = 0$ para cualquier $\varepsilon > 0$. Así que el resultado se sigue aplicando el corolario 1.

■

Corolario 3. *Sea f una función medible y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tales que $\sum_{n=1}^{\infty} \mu\left[|f_n - f| > \varepsilon\right] < \infty$ para cualquier $\varepsilon > 0$. Entonces $f_n \xrightarrow{c.t.p.} f$.*

Convergencia en medida

Definición 2. *Diremos que una sucesión de funciones medibles* $(f_n)_{n\in\mathbb{N}}$ *converge en medida si existe una función medible* f *tal que:*

$$\text{lím}_{n\to\to\infty}\,\mu\left(\{y\in\mathbb{E} : |f_n(y)-f(y)|>\varepsilon\}\right)=0$$

para cualquier $\varepsilon>0$. Si éste es el caso, se escribirá $f_n \xrightarrow{\mu} f$.

Obviamente si una sucesión $(f_n)_{n\in\mathbb{N}}$ converge en medida a f, entonces cualquier subsucesión de $(f_n)_{n\in\mathbb{N}}$ también converge en medida a f.

Proposición 1. *Sea* $\{f_n\}$ *una sucesión de funciones medibles tal que* $f_n \xrightarrow{\mu} f$ *y* $f_n \xrightarrow{\mu} g$, *entonces* $f=g$ *casi en todas partes.*

Demostración

Como $|f-g|\leq|f_n-f|+|f_n-g|$, entonces:

$$[|f-g|>\varepsilon]\subset[|f_n-f|+|f_n-g|>\varepsilon]$$

Además, para cualquier $\varepsilon>0$, se tiene:

$$[|f_n-f|+|f_n-g|>\varepsilon]\subset\left[|f_n-f|>\tfrac{\varepsilon}{2}\right]\cup\left[|f_n-g|>\tfrac{\varepsilon}{2}\right]$$

Por lo tanto:

$$\mu\,[|f-g|>\varepsilon]\leq\mu\left[|f_n-f|>\tfrac{\varepsilon}{2}\right]+\mu\left[|f_n-g|>\tfrac{\varepsilon}{2}\right]$$

Así que, tomando límites, se obtiene $\mu\,[|f-g|>\varepsilon]=0$ para cualquier $\varepsilon>0$.

Finalmente, $[|f-g|>0]=\bigcup_{n=1}^{\infty}\left[|f-g|>\frac{1}{n}\right]$, así que:

$$\mu\,[|f-g|>0]\leq\sum_{n=1}^{\infty}\mu\left[|f-g|>\tfrac{1}{n}\right]=0$$

$\blacksquare$

Proposición 2. *Sea* c *una constante y* $(f_n)_{n\in\mathbb{N}}$ *una sucesión de funciones medibles tal que* $f_n \xrightarrow{\mu} f$, *entonces* $cf_n \xrightarrow{\mu} cf$.

Demostración

Si $c=0$, el resultado es obvio.

Si $c\neq0$, se tiene:

$$\lim_{n\to\infty} \mu\left[|cf_n - cf| > \varepsilon\right] = \lim_{n\to\infty} \mu\left[|f_n - f| > \tfrac{\varepsilon}{|c|}\right] = 0$$

■

Proposición 3. *Sean $(f_n)_{n\in\mathbb{N}}$ y $(g_n)_{n\in\mathbb{N}}$ dos sucesiones de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g_n \xrightarrow{\mu} g$, entonces $f_n + g_n \xrightarrow{\mu} f + g$.*

Demostración

Como $|f_n - f + g_n - g| \leq |f_n - f| + |g_n - g|$, se tiene:

$$\left[|f_n - f + g_n - g| > \varepsilon\right] \subset \left[|f_n - f| > \tfrac{\varepsilon}{2}\right] \cup \left[|g_n - g| > \tfrac{\varepsilon}{2}\right]$$

Así que:

$\lim_{n\to\infty} \mu\left[|f_n + g_n - f - g| > \varepsilon\right]$

$\leq \lim_{n\to\infty} \mu\left[|f_n - f| > \tfrac{\varepsilon}{2}\right] + \lim_{n\to\infty} \mu\left[|g_n - g| > \tfrac{\varepsilon}{2}\right] = 0$

■

Proposición 4. *Sean f una función medible, $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \to \mathbb{R}$ una función uniformemente continua, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.*

Demostración

Como g es uniformemente continua en $\mathbb{R}$, dada $\varepsilon > 0$ existe $\delta > 0$ tal que $|g(u) - g(v)| < \varepsilon$ para cualesquiera $u, v \in \mathbb{R}$ tales que $|v - u| < \delta$. Así que:

$$|g \circ f_n(x) - g \circ f(x)| < \varepsilon$$

para cualquier $x \in \mathbb{F}$ tal que $|f_n(x) - f(x)| < \delta$.

Por lo tanto:

$\limsup_{n\to\infty} \mu\left(\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| > \varepsilon\}\right)$

$\leq \limsup_{n\to\infty} \mu\left(\{y \in \mathbb{F} : |g \circ f_n(y) - g \circ f(y)| \geq \varepsilon\}\right)$

$\leq \limsup_{n\to\infty} \mu\left(\{y \in \mathbb{F} : |f_n(y) - f(y)| \geq \delta\}\right)$

$\leq \lim_{n\to\infty} \mu\left(\left\{y \in \mathbb{F} : |f_n(y) - f(y)| > \tfrac{1}{2}\delta\right\}\right) = 0$

Así que $g \circ f_n \xrightarrow{\mu} g \circ f$.

■

Corolario 4. *Sean f una función medible, $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g:\mathbb{R}\to\mathbb{R}$ una función continua nula fuera de un intervalo $[a,b]$, entonces $g\circ f_n \xrightarrow{\mu} g\circ f$.*

Demostración

Como g es uniformemente continua en $[a,b]$ y nula fuera de ese intervalo, es uniformemente continua en $\mathbb{R}$; así que el resultado se sigue de la proposición anterior.

■

Teorema 3. *Supongamos que μ es finita y sean f una función medible, $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$ y $g:\mathbb{R}\to\mathbb{R}$ una función continua, entonces $g\circ f_n \xrightarrow{\mu} g\circ f$.*

Demostración

Como $\sum_{k=0}^{\infty}\mu\left[k\leq|f|<k+1\right]=\mu\left(\mathbb{E}\right)<\infty$, entonces dada $\eta>0$ existe M tal que:

$$\mu\left[|f|>M\right]\leq\mu\left[|f|\geq M\right]=\sum_{k=M}^{\infty}P\left[k\leq|f|<k+1\right]<\tfrac{1}{2}\eta$$

También, como $f_n \xrightarrow{\mu} f$, existe $N\in\mathbb{N}$ tal que, si $n\geq N$, entonces:

$$\mu\left[|f_n-f|>M\right]<\tfrac{1}{2}\eta$$

Sea $D_n=\{y\in\mathbb{E}:|f(y)|\leq M\}\cap\{y\in\mathbb{E}:|f_n(y)-f(y)|\leq M\}$.

Entonces:

$$D_n^c=\{y\in\mathbb{F}:|f(y)|>M\}\cup\{y\in\mathbb{F}:|f_n(y)-f(y)|>M\}$$

Así que:

$$\mu\left(D_n^c\right)<\eta$$

Si $y\in D_n$, entonces:

$$|f_n(y)|\leq|f_n(y)-f(y)|+|f(y)|\leq 2M$$

Definamos:

$$g^{(M)}(u)=\begin{cases} g(u) & \text{si } |u|<2M\\ g(2M) & \text{si } u\geq 2M\\ g(-2M) & \text{si } u\leq -2M\end{cases}$$

$g^{(M)}$ es uniformemente continua en $\mathbb{R}$, así que, dada $\varepsilon>0$ existe $\delta>0$ tal que:

$$\left|g^{(M)}(u)-g^{(M)}(v)\right|<\varepsilon$$

para cualesquiera $u, v \in \mathbb{R}$ tales que $|v-u|<\delta$.

Así que:

$$\left|g^{(M)}\circ f_n(x)-g^{(M)}\circ f(x)\right|<\varepsilon$$

para cualquier $x \in \mathbb{F}$ tal que $|f_n(x)-f(x)|<\delta$.

Por lo tanto, si $n \geq N$, se tiene:

$\mu(\{y \in \mathbb{F} : |g\circ f_n(y)-g\circ f(y)|>\varepsilon\})$

$=\mu(D_n \cap \{y \in \mathbb{F} : |g\circ f_n(y)-g\circ f(y)|>\varepsilon\})+\mu(D_n^c \cap \{y \in \mathbb{F} : |g\circ f_n(y)-g\circ f(y)|>\varepsilon\})$

$\leq \mu\left(D_n \cap \left\{y \in \mathbb{F} : \left|g^{(M)}\circ f_n(y)-g^{(M)}\circ f(y)\right| \geq \varepsilon\right\}\right)+\mu(D_n^c)$

$\leq \mu\left(\left\{y \in \mathbb{F} : \left|g^{(M)}\circ f_n(y)-g^{(M)}\circ f(y)\right| \geq \varepsilon\right\}\right)+\mu(D_n^c)$

$\leq \mu(\{y \in \mathbb{F} : |f_n(y)-f(y)| \geq \delta\})+\mu(D_n^c)$

$\leq \mu\left(\left\{y \in \mathbb{F} : |f_n(y)-f(y)|>\frac{1}{2}\delta\right\}\right)+\eta$

Así que:

$\limsup_{n\to\infty} \mu(\{y \in \mathbb{F} : |g\circ f_n(y)-g\circ f(y)|>\varepsilon\})$

$\leq \lim_{n\to\infty} \mu\left(\left\{y \in \mathbb{F} : |f_n(y)-f(y)|>\frac{1}{2}\delta\right\}\right)+\eta=\eta$

Como lo anterior es válido para cualquier $\eta>0$, se concluye que:

$$\lim_{n\to\infty} \mu(\{y \in \mathbb{E} : |g\circ f_n(y)-g\circ f(y)|>\varepsilon\})=0$$

Por lo tanto, $g\circ f_n \xrightarrow{\mu} g\circ f$.

■

Corolario 5. *Supongamos que μ es finita y sean f y g dos funciones medibles, y $(f_n)_{n\in\mathbb{N}}$ y $(g_n)_{n\in\mathbb{N}}$ dos sucesiones de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g_n \xrightarrow{\mu} g$, entonces $f_n g_n \xrightarrow{\mu} fg$.*

Demostración

Como $f_n g_n = \frac{1}{4}\left[(f_n+g_n)^2-(f_n-g_n)^2\right]$, entonces:

$$f_n g_n \xrightarrow{\mu} \frac{1}{4}\left[(f+g)^2-(f-g)^2\right]=fg$$

■

Teorema 4. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} 0$, entonces $f_n \xrightarrow{\mu} 0$.*

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m(\varepsilon) = \bigcup_{n=m}^{\infty} \{y \in \mathbb{E} : |f_n(y)| > \varepsilon\}$; entonces la sucesión de conjuntos $B_1(\varepsilon), B_2(\varepsilon), \ldots$ es decreciente, así que:

$$\text{lím}_{m\to\infty} \mu\left[B_m(\varepsilon)\right] = \mu\left(\bigcap_{m=1}^{\infty} B_m(\varepsilon)\right) = 0$$

Pero $[|f_m| > \varepsilon] \subset B_m(\varepsilon)$. Por lo tanto:

$$\text{lím}_{m\to\infty} \mu\left[|f_m| > \varepsilon\right] \leq \text{lím}_{m\to\infty} \mu\left[B_m(\varepsilon)\right] = 0$$

$\blacksquare$

Corolario 6. *Supongamos que μ es finita y sea $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$, entonces $f_n \xrightarrow{\mu} f$.*

Si μ no es finita, el corolario anterior no es válido en general.

En efecto, consideremos el ejemplo siguiente:

Ejemplo 1. *Sea $\mathbb{E} = \mathbb{R}$, μ la medida de Lebesgue sobre $\mathbb{R}$ y $(f_n)_{n\in\mathbb{N}}$ la sucesión de funciones definidas por $f_n = I_{(n,n+1)}$ para cualquier $n \in \mathbb{N}$. Entonces la sucesión $(f_n(x))_{n\in\mathbb{N}}$ converge a 0 para cualquier $x \in \mathbb{R}$, pero dada $\varepsilon \in (0,1)$, se tiene:*

$$\mu\left(\{y \in \mathbb{R} : |f_n(y)| > \varepsilon\}\right) = 1$$

para cualquier $n \in \mathbb{N}$.

Así que $(f_n)_{n\in\mathbb{N}}$ no converge en medida a la función idénticamente cero. Más aún, $(f_n)_{n\in\mathbb{N}}$ no converge en medida. En efecto, supongamos que $(f_n)_{n\in\mathbb{N}}$ converge en medida a una función medible f, entonces existe una subsucesión que converge casi en todas partes (este resultado se demuestra más adelante), así que f es la función idénticamente cero.

Como se muestra en el siguiente ejemplo, el inverso del corolario 6 no es válido en general.

Ejemplo 2. *Sea $\mathbb{E} = (0,1]$, μ la medida de Lebesgue sobre $\mathbb{E}$ e $(I_n)_{n\in\mathbb{N}}$ la sucesión de intervalos $J_1 = (0,1]$, $J_2 = (0,\frac{1}{2}]$, $J_3 = (\frac{1}{2},\frac{1}{2}]$, $J_4 = (0,\frac{1}{2^2}]$, $J_5 = (\frac{1}{2^2},\frac{2}{2^2}]$, $J_6 = (\frac{2}{2^2},\frac{3}{2^2}]$, $J_7 = (\frac{3}{2^2},\frac{4}{2^2}]$, $\ldots$; es decir, para $n \in \{0,1,2,\ldots\}$ y $j \in \{0,1,2,\ldots,2^n-1\}$, $J_{2^n+j} = (\frac{j}{2^n},\frac{j+1}{2^n}]$. Para cada $n \in \mathbb{N}$, definamos $f_n = I_{J_n}$; es decir, para $n \in \{0,1,2,\ldots\}$ y $j \in \{0,1,2,\ldots,2^n-1\}$, $f_{2^n+j} = I_{(\frac{j}{2^n},\frac{j+1}{2^n}]}$.*

Si $x \in \mathbb{F}$ y $n \in \{0,1,2,\ldots\}$, entonces existe un único elemento $j_0 \in \{0,1,2,\ldots,2^n-1\}$ tal que $x \in (\frac{j}{2^n}, \frac{j+1}{2^n}]$, así que $f_{2^n+j_0}(x) = 1$ y $f_{2^n+j}(x) = 0$ para cualquier $j \in \{0,1,2,\ldots,2^n-1\} - \{j_0\}$. Por lo tanto, $f_n(x) = 1$ para una infinidad de valores de n y $f_n(x) = 0$ para una infinidad de valores de n. Así que la sucesión $(f_n(x))_{n\in\mathbb{N}}$ no converge para ninguna $x \in \mathbb{F}$.

Sin embargo, para cualquier $n \in \{0,1,2,\ldots\}$ y $j \in \{0,1,2,\ldots,2^n-1\}$, f_{2^n+j} toma únicamente los valores 0 y 1, y $\mu(\{y \in \mathbb{F} : f_{2^n+j}(y) = 1\}) = \frac{1}{2^n}$. Así que $\lim_{n\to\infty} \mu[|f_n| > \varepsilon] = 0$ para cualquier $\varepsilon > 0$. Por lo tanto, $f_n \xrightarrow{\mu} 0$.

De la definición de convergencia en medida podemos ver que, en un sentido, las funciones f_n, para n suficientemente grande, están cercanas a la función límite f. En efecto, si $f_n \xrightarrow{\mu} f$, dada cualquier $\varepsilon > 0$, por pequeña que sea, y dada $\delta > 0$, por pequeña que sea, existe $N \in \mathbb{N}$ tal que:

$$\mu(\{y \in \mathbb{E} : |f_n(y) - f(y)| > \varepsilon\}) < \delta$$

para cualquier $n \geq N$.

Es decir, denotando por A_n al conjunto $\{y \in \mathbb{E} : |f_n(y) - f(y)| > \varepsilon\}$, entonces, para cualquier $n \geq N$, $|f_n(y) - f(y)| \leq \varepsilon$ para cualquier $y \in A_n^c$ y $\mu(A_n) < \delta$.

Lo anterior podría dar la idea de que, fuera de un conjunto de medida pequeña, se tiene convergencia uniforme de la sucesión $(f_n)_{n\in\mathbb{N}}$. Sin embargo, esto no es así ya que el conjunto de medida pequeña (A_n) no es fijo, depende de n. Esto se ve claro en el ejemplo 2 ya que, dada $\varepsilon \in (0,1)$ y $\delta > 0$, digamos $\delta = \frac{1}{2^{n_0}}$, con n_0 grande para asegurar que δ es pequeña, se tiene que, para $n \geq 2^{n_0+1}$, $\mu(A_n) < \delta$, pero:

$A_{2^{n_0+1}} = I_{(\frac{0}{2^{n_0+1}}, \frac{1}{2^{n_0+1}}]}$

$A_{2^{n_0+1}+1} = I_{(\frac{1}{2^{n_0+1}}, \frac{2}{2^{n_0+1}}]}$

$\ldots$

$A_{2^{n_0+2}-1} = I_{(\frac{2^{n_0+1}-1}{2^{n_0+1}}, \frac{2^{n_0+1}}{2^{n_0+1}}]}$

$A_{2^{n_0+2}} = I_{(\frac{0}{2^{n_0+2}}, \frac{1}{2^{n_0+2}}]}$

$A_{2^{n_0+2}+1} = I_{(\frac{1}{2^{n_0+2}}, \frac{2}{2^{n_0+2}}]}$

$\ldots$

$A_{2^{n_0+3}-1} = I_{(\frac{2^{n_0+2}-1}{2^{n_0+2}}, \frac{2^{n_0+2}}{2^{n_0+1}}]}$

$\ldots$

Se tiene la convergencia uniforme fuera de un conjunto fijo de medida pequeña, pero únicamente para alguna subsucesión de $(f_n)_{n\in\mathbb{N}}$, lo cual demostramos más adelante.

Definición 3. *Diremos que una sucesión de funciones medibles* $(f_n)_{n\in\mathbb{N}}$ *es de Cauchy en medida si, para cualquier* $\varepsilon > 0$ *y cualquier* $\delta > 0$ *existe* $N \in \mathbb{N}$ *tal que:*

$$\mu\left(\{y \in \mathbb{F} : |f_n(y) - f_m(y)| > \varepsilon\}\right) < \delta$$

para cualquier par de números naturales n y m mayores o iguales a N.

Teorema 5. *Sea* $(f_n)_{n\in\mathbb{N}}$ *una sucesión de funciones medibles que converge en medida, entonces* $(f_n)_{n\in\mathbb{N}}$ *es de Cauchy en medida.*

Demostración

Sea f el límite en medida de la sucesión $(f_n)_{n\in\mathbb{N}}$.

Dadas $\varepsilon > 0$ y $\delta > 0$, sea $N \in \mathbb{N}$ tal que:

$$\mu\left[|f_n. - f| > \tfrac{1}{2}\varepsilon\right] < \tfrac{1}{2}\delta$$

para cualquier número natural $n \geq N$.

Entonces, si n y m son dos números naturales mayores o iguales que N, se tiene:

$$\mu\left[|f_n. - f_m| > \varepsilon\right] \leq \mu\left[|f_n. - f| > \tfrac{1}{2}\varepsilon\right] + \mu\left[|f_m. - f| > \tfrac{1}{2}\varepsilon\right] < \delta$$

Por lo tanto, $(f_n)_{n\in\mathbb{N}}$ es de Cauchy en medida.

$\blacksquare$

Teorema 6. *Sea* $(f_n)_{n\in\mathbb{N}}$ *una sucesión de funciones medibles y supongamos que* $(f_n)_{n\in\mathbb{N}}$ *es de Cauchy en medida, entonces* $(f_n)_{n\in\mathbb{N}}$ *contiene una subsucesión* $(f_{n_k})_{k\in\mathbb{N}}$ *que converge casi en todas partes a una función medible* f *y tal que, dada* $\delta > 0$*, existe un conjunto medible* A *tal que* $\mu(A) < \delta$ *y la sucesión* $(f_{n_k})_{k\in\mathbb{N}}$ *converge uniformemente a* f *sobre* A^c*.*

Demostración

Para cada $k \in \mathbb{N}$, definamos $\varepsilon_k = \delta_k = \frac{1}{2^k}$. Sabemos que existe $N_k \in \mathbb{N}$ tal que:

$$\mu\left(\{y \in \mathbb{F} : |f_n(y) - f_m(y)| > \varepsilon_k\}\right) < \delta_k$$

para cualquier par de números naturales n y m mayores o iguales a N_k.

Existe entonces una sucesión creciente de números naturales $(m_k)_{k\in\mathbb{N}}$ tal que:

$$\mu\left[\left|f_{m_{k+1}}. - f_{m_k}\right| > \tfrac{1}{2^k}\right] < \tfrac{1}{2^k} \text{ para cualquier } k \in \mathbb{N}$$

Sea $B=\bigcap_{j=1}^{\infty}\left(\bigcup_{i=j}^{\infty}\left[\left|f_{m_{i+1}}.-f_{m_i}\right|>\frac{1}{2^i}\right]\right)$, entonces, como:

$$\sum_{k=1}^{\infty}\mu\left[\left|f_{m_{k+1}}.-f_{m_k}\right|>\frac{1}{2^k}\right]<\infty$$

por el lema de Borel-Cantelli, se tiene que $\mu\left(B\right)=0$.

Para cada $j\in\mathbb{N}$, definamos $B_j=\bigcup_{i=j}^{\infty}\left[\left|f_{m_{i+1}}.-f_{m_i}\right|>\frac{1}{2^i}\right]$.

Para cualquier $i\in\mathbb{N}$ tal que $i\geq j$, se tiene:

$$\left|f_{m_{i+1}}\left(y\right).-f_{m_i}\left(y\right)\right|\leq\frac{1}{2^i}$$

para cualquier $y\in B_j^c$.

Dada $\varepsilon>0$, sea $i_0\in\mathbb{N}$ tal que $i_0\geq N$ y $\sum_{i=i_0}^{\infty}\frac{1}{2^i}<\varepsilon$. Entonces, si $r>s\geq i_0$, se tiene:

$\left|f_{m_r}\left(y\right).-f_{m_s}\left(y\right)\right|\leq\sum_{i=s}^{r-1}\left|f_{m_{i+1}}\left(y\right).-f_{m_i}\left(y\right)\right|$

$\leq\sum_{i=s}^{r-1}\frac{1}{2^i}<\sum_{i=i_0}^{\infty}\frac{1}{2^i}<\varepsilon$

para cualquier $y\in B_j^c$.

Así que la sucesión $\left(f_{m_k}\right)_{k\in\mathbb{N}}$ es uniformemente de Cauchy en B_j^c. Por lo tanto, converge uniformemente en B_j^c.

Si $y\in B^c=\bigcup_{j=1}^{\infty}\left(\bigcap_{i=j}^{\infty}\left[\left|f_{m_{i+1}}.-f_{m_i}\right|\leq\frac{1}{2^i}\right]\right)$, entonces existe $j\in\mathbb{N}$ tal que $y\in B_j^c$, así que la sucesión $\left(f_{m_k}\left(y\right)\right)_{k\in\mathbb{N}}$ es convergente.

Sea $f\left(y\right)=\lim_{k\to\infty}f_{m_k}\left(y\right)$.Entonces $f_{m_k}\xrightarrow{c.t.p.}f$ ya que $\mu\left(B\right)=0$.

Además, para cualquier $j\in\mathbb{N}$, $\left(f_{m_k}\right)_{k\in\mathbb{N}}$ converge uniformemente a f en B_j^c.

Finalmente, como $\lim_{j\to\infty}\mu\left(B_j\right)=0$, dada $\delta>0$, existe $N\in\mathbb{N}$ tal que $\mu\left(B_N\right)<\delta$.

■

Corolario 7. *Sea $\left(f_n\right)_{n\in\mathbb{N}}$ una sucesión de funciones medibles que converge en medida, entonces $\left(f_n\right)_{n\in\mathbb{N}}$ contiene una subsucesión $\left(f_{n_k}\right)_{k\in\mathbb{N}}$ que converge casi en todas partes a una función medible f y tal que, dada $\delta>0$, existe un conjunto medible A tal que $\mu\left(A\right)<\delta$ y la sucesión $\left(f_{n_k}\right)_{k\in\mathbb{N}}$ converge uniformemente a f sobre A^c.*

Teorema 7. *Supongamos que μ es finita y sea $\left(f_n\right)_{n\in\mathbb{N}}$ una sucesión de funciones medibles, de Cauchy en medida, entonces $\left(f_n\right)_{n\in\mathbb{N}}$ converge en medida..*

Demostración

Sea $(f_{m_k})_{k\in\mathbb{N}}$ una subsucesión que converge casi en todas partes a la función f, entonces $f_{m_k} \xrightarrow{\mu} f$.

Dadas $\varepsilon > 0$ y $\delta > 0$, sea $N \in \mathbb{N}$ tal que:

$$\mu\left[|f_{n\cdot} - f_m| > \tfrac{1}{2}\varepsilon\right] < \tfrac{1}{2}\delta$$
$$\mu\left[|f_{m_k\cdot} - f| > \tfrac{1}{2}\varepsilon\right] < \tfrac{1}{2}\delta$$

para cualquier terna de números naturales, n, m y k, mayores o iguales que N.

Sean n y k números naturales mayores o iguales que N, entonces:

$$\mu\left[|f_{n\cdot} - f| > \varepsilon\right] \leq \mu\left[|f_{n\cdot} - f_{m_k}| > \tfrac{1}{2}\varepsilon\right] + \mu\left[|f_{m_k\cdot} - f| > \tfrac{1}{2}\varepsilon\right] < \delta$$

Por lo tanto $f_n \xrightarrow{\mu} f$.

■

Teorema 8. *Si μ es finita y $(f_n)_{n\in\mathbb{N}}$ es una sucesión de funciones medibles que converge casi en todas a la función medible f, entonces, dada $\delta > 0$, existe un conjunto medible A tal que $\mu(A) < \delta$ y la sucesión $(f_n)_{n\in\mathbb{N}}$ converge uniformemente a f sobre A^c.*

Demostración

Para cualquier $\varepsilon > 0$, se tiene:

$$\mu\left(\limsup \{y \in \mathbb{F} : |f_n(y) - f(y)| > \varepsilon\}\right) = 0$$

Para cada $i \in \mathbb{N}$, sea $B^{(i)} = \bigcap_{j=1}^{\infty}\left(\bigcup_{k=j}^{\infty}\left[|f_{k\cdot} - f| > \frac{1}{2^i}\right]\right)$, entonces $\mu\left(B^{(i)}\right) = 0$.

Para cada $j \in \mathbb{N}$, definamos $B_j^{(i)} = \bigcup_{k=j}^{\infty}\left[|f_{k\cdot} - f| > \frac{1}{2^i}\right]$.

Como $\lim_{j\to\infty} \mu\left(B_j^{(i)}\right) = 0$, dada $\delta > 0$, existe $N_i \in \mathbb{N}$ tal que $\mu\left(B_{N_i}^{(i)}\right) < \frac{\delta}{2^i}$.

Definamos $B = \bigcup_{i=1}^{\infty} B_{N_i}^{(i)}$, entonces $\mu(B) < \delta$.

Para cualquier $k \in \mathbb{N}$ tal que $k \geq N_i$, se tiene:

$$|f_k(y) - f(y)| \leq \tfrac{1}{2^i}$$

para cualquier $y \in \left(B_{N_i}^{(i)}\right)^c \supset B^c$.

Dada $\varepsilon > 0$, sea $i \in \mathbb{N}$ tal que $\frac{1}{2^i} < \varepsilon$. Entonces, para cualquier $k \geq N_i$, se tiene:

$$|f_k(y) - f(y)| < \varepsilon$$

para cualquier $y \in B^c$.

Por lo tanto, $(f_n)_{n\in\mathbb{N}}$ converge uniformemente en B^c.

■

Teorema 9. *Sean g una función no negativa e integrable, f una función medible y $(f_n)_{n\in\mathbb{N}}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ para cualquier $n \in \mathbb{N}$ y $f_n \xrightarrow{\mu} f$, entonces* $\lim_{n\to\infty} \int_{\mathbb{E}} |f_n - f|\, d\mu = 0$.

Demostración

Para cada $n \in \mathbb{N}$, definamos:

$$s_n = \int_{\mathbb{F}} |f_n - f|\, d\mu$$

Se tiene entonces:

$$s_n = \int_{\mathbb{F}} |f_n - f|\, d\mu \leq \int_{\mathbb{E}} |f_n|\, d\mu + \int_{\mathbb{E}} |f|\, d\mu \leq 2\int_{\mathbb{E}} g d\mu < \infty$$

Así que la sucesión $(s_n)_{n\in\mathbb{N}}$ está acotada.

Sea $(s_{n_k})_{k\in\mathbb{N}}$ una subsucesión convergente. Como $f_n \xrightarrow{\mu} f$, existe una subsucesión $\left(s_{n_{k_j}}\right)_{j\in\mathbb{N}}$ de $(s_{n_k})_{k\in\mathbb{N}}$ tal que $f_{n_{k_j}} \xrightarrow{c.t.p.} f$. Por el teorema de la convergencia dominada, la sucesión $\left(s_{n_{k_j}}\right)_{j\in\mathbb{N}}$ converge a cero, así que $(s_{n_k})_{k\in\mathbb{N}}$ converge a cero. Por lo tanto, la sucesión $(s_n)_{n\in\mathbb{N}}$ converge a cero.

■

Desigualdad de Chebyshev y ley débil delos grandes números

El gran impulso para el desarrollo de una teoría de la probabilidad, que le haría ganar un lugar dentro de las matemáticas, proviene de los llamados teoremas límite, los cuales se refieren al comportamiento a largo plazo de sucesiones de variables aleatorias. El primero de estos resultados, que para algunos autores marca verdaderamente el inicio de la historia de la teoría de la probabilidad, se debe a Jacques Bernoulli, quien dedicó 20 años de su vida a la búsqueda de una prueba matemática de la relación que existe entre la probabilidad de un evento y la frecuencia relativa con la que éste ocurre en una serie grande de repeticiones del correspondiente experimento aleatorio. El resultado, conocido como teorema de Bernoulli, se publicó en el año 1713, ocho años después de la muerte de su autor.

Puede decirse que, a partir de la publicación del teorema de Bernoulli, el motor de desarrollo de la teoría de la probabilidad fue la búsqueda de resultados que permitieran mejorar y generalizar ese teorema. Vendrían después los teoremas de de Moivre y de Poisson, relativos a la aproximación de una distribución binomial mediante una distribución normal y una distribución Poisson, respectivamente, los cuales fueron publicados en los años 1730 y 1800, respectivamente.

Este proceso continuaría desarrollándose y recibiría un gran impulso, entre 1870 y 1900, con los trabajos de la llamada escuela rusa, representada por Pafnuty Lvovich Chebyshev, Andrei Andreyevich Markov y Aleksandr Mikhailovich Lyapunov, entre otros, los cuales conducirían a la forma general que se dio a los teoremas límite, entre 1900 y 1930, con la formulación de las leyes de los grandes números y el teorema central del límite, tanto en su forma clásica, relativa a la convergencia a la distribución normal, como en su forma moderna, relativa a la convergencia a cualquier otro tipo de distribución, sobresaliendo en este periodo los trabajos de Aleksandr Yakovlevich Khintchine, Andrey Nikolaevich Kolmogorov, J. W. Lindeberg, William Feller y Paul Pierre Lévy, entre otros.

En la exposición que sigue, $(\nleq, \Im, P)$ será un espacio de probabilidad completo.

Como ya lo mencionamos, en la teoría de la probabilidad la convergencia en medida es llamada convergencia en probabilidad.

Si $(X_n)_{n\in N}$ es una sucesión de variables aleatorias que converge a la variable aleatoria X en probabilidad, escribiremos $X_n \overset{P}{\longrightarrow} X$.

La convergencia casi en todas partes es llamada convergencia casi segura o convergencia con probabilidad 1.

Si $(X_n)_{n\in N}$ es una sucesión de variables aleatorias que converge a la variable aleatoria X casi seguramente, escribiremos $X_n \overset{c.s.}{\longrightarrow} X$, o bien:

$$P\left[\text{lím}_{n\to\infty} X_n = X\right] = 1$$

Proposición 5. *Sea X cualquier variable aleatoria y ε cualquier número real positivo, entonces:*

$$P\left[|X| \geq \varepsilon\right] \leq \tfrac{1}{\varepsilon} E\left[|X|\right]$$

Demostración

$$E\left[|X|\right] = \textstyle\int_\Omega |X|\, dP \geq \int_{[|X|\geq\varepsilon]} |X|\, dP \geq \varepsilon P\left[|X| \geq \varepsilon\right]$$

Así que:

$$P\left[|X| \geq \varepsilon\right] \leq \tfrac{1}{\varepsilon} E\left[|X|\right] \qquad \blacksquare$$

Corolario 8. *Sea X cualquier variable aleatoria y ε cualquier número real positivo, entonces:*

$$P\left[|X| \geq \varepsilon\right] \leq \tfrac{1}{\varepsilon^2} E\left[X^2\right]$$

Demostración

$$P\left[|X| \geq \varepsilon\right] = P\left[X^2 \geq \varepsilon^2\right] \leq \tfrac{1}{\varepsilon^2} E\left[X^2\right]$$

■

Corolario 9 (Desigualdad de Chebyshev). *Sea X cualquier variable aleatoria de esperanza finita y ε cualquier número real positivo, entonces:*

$$P\left[|X - E\left[X\right]| \geq \varepsilon\right] \leq \tfrac{1}{\varepsilon^2} Var\left[X\right]$$

Teorema 10 (Ley débil de los grandes números de Chebyshev). *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita. Entonces:*

$$\tfrac{X_1 + \ldots + X_n}{n} \xrightarrow{P} \mu$$

donde μ es la esperanza común de $X_1, X_2, \ldots$.

Demostración

Para cada $n \in \mathbb{N}$, sea $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Entonces Y_n es una variable aleatoria de varianza finita y esperanza μ. De manera que, por la desigualdad de Chebyshev, se tiene:

$$P\left[|Y_n - \mu| > \varepsilon\right] \leq \tfrac{1}{\varepsilon^2} Var\left[Y_n\right] = \tfrac{\sigma^2}{n\varepsilon^2}$$

donde σ^2 es la varianza común de $X_1, X_2, \ldots$. Tomando límites cuando $n \to \infty$ se tiene entonces el resultado.

■

En 1928, Khintchine demostró la ley débil de los grandes números sin la hipótesis de que las variables aleatorias $X_1, X_2, \ldots$ tengan varianza finita. Para ello utilizó un método conocido como el el **método de truncación**, el cual fue introducido por Markov en el año 1913 con relación a un teorema de Lyapunov, el cual generaliza el teorema de de Moivre.

Para demostrar el teorema de Khintchine requerimos de algunos resultados previos.

Lema 1. *Si $f : [0, \infty) \mapsto \mathbb{R}$ es una función decreciente y no negativa tal que $\int_{[0,\infty)} f d\lambda < \infty$ y (a_n) una sucesión creciente de números reales positivos tal que $\lim_{n\to\infty} a_n = \infty$, entonces $\lim_{n\to\infty} a_n f(a_n) = 0$.*

Demostración

La sucesión (s_n), donde $s_n = \sum_{\{k\in\mathbb{N}:k\leq a_n\}} f(k)$, es no decreciente y se tiene:

$$s_n \leq \sum_{\{k\in\mathbb{N}:k\leq a_n\}} \int_{[k-1,k)} f d\lambda \leq \int_{[0,a_n)} f d\lambda \leq \int_{[0,\infty)} f d\lambda < \infty$$

Así que (s_n) converge; por lo tanto, es una sucesión de Cauchy.

Entonces, dada $\varepsilon > 0$ existe un número natural M tal que si $n \geq m \geq M$ entonces $s_n - s_m < \frac{\varepsilon}{2}$, es decir $\sum_{\{k\in\mathbb{N}:a_m<k\leq a_n\}} f(k) < \frac{\varepsilon}{2}$.

Sea ahora $N \in \mathbb{N}$ tal que $a_n > 2(a_M + 1)$ para cualquier $n > N$. Se tiene entonces, para $n > N$, $a_n - 2(a_M + 1) > 0$ y $(a_n - a_M - 1)f(a_n) \leq \sum_{\{k\in\mathbb{N}:a_M<k\leq a_n\}} f(k) < \frac{\varepsilon}{2}$. Así que:

$$a_n f(a_n) < 2(a_n - a_M - 1)f(a_n) < \varepsilon$$

lo cual prueba el resultado.

■

Proposición 6. *Si X es una variable aleatoria de esperanza finita y (a_n) una sucesión creciente de números reales positivos tal que $\lim_{n\to\infty} a_n = \infty$, entonces:*

$$\lim_{n\to\infty} a_n P[X > a_n] = \lim_{n\to\infty} a_n P[X < -a_n] = 0$$

Demostración

Como X tiene esperanza finita, se tiene:

$$\int_0^\infty P[X > x]\, dx = \int_0^\infty [1 - F_X(x)]\, dx < \infty$$

$$\int_0^\infty P[X < -x]\, dx \leq \int_0^\infty P[X \leq -x]\, dx = \int_0^\infty F_X(-x) < \infty$$

Además, las funciones $x \mapsto P[X > x]$ y $x \mapsto P[X < -x]$ son no negativas y decrecientes en el intervalo $[0, \infty)$.

El resultado se sigue entonces del lema 1.

■

Lema 2. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ y (a_n) una sucesión creciente de números reales positivos tal que $\lim_{n\to\infty} a_n = \infty$. Para $n, k \in \mathbb{N}$, definamos:*

$$Y_k^n = \begin{cases} X_k & \text{si } |X_k| \leq a_n \\ 0 & \text{en otro caso} \end{cases}.$$

Entonces, fijando n, las variables aleatorias $Y_1^n, Y_2^n, \ldots$ tienen la misma distribución. Además,si μ_n es la esperanza común de $Y_1^n, Y_2^n, \ldots$, entonces $\lim_{n\to\infty} \mu_n = \mu$.

Demostración

Para cualquier $x \in \mathbb{R}$, se tiene:

$F_{Y_k^n}(x) = P\left[Y_k^n \leq x\right] = P\left[Y_k^n \leq x, |X_k| \leq a_n\right] + P\left[Y_k^n \leq x, |X_k| > a_n\right]$

$= P\left[X_k \leq x, |X_k| \leq a_n\right] + P\left[Y_k^n \leq x, |X_k| > a_n\right]$

$$= \begin{cases} 0 & \text{si } x < -a_n \\ P\left[-a_n \leq X_k \leq x\right] & \text{si } -a_n \leq x < 0 \\ P\left[-a_n \leq X_k \leq x\right] + P\left[|X_k| > a_n\right] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -a_n \\ P\left[-a_n \leq X_k \leq x\right] & \text{si } -a_n \leq x < 0 \\ P\left[X_k \leq x\right] + P\left[X_k > a_n\right] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -a_n \\ F_{X_k}(x) - P\left[X_k < -a_n\right] & \text{si } -a_n \leq x < 0 \\ F_{X_k}(x) + P\left[X_k > a_n\right] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases}$$

De manera que, fijando n, las variables aleatorias $Y_1^n, Y_2^n, \ldots$ tienen la misma distribución. Además:

$\mu_n = E\left[Y_1^n\right] = \int_0^\infty \left[1 - F_{Y_1^n}(x)\right] dx - \int_0^\infty F_{Y_1^n}(-x)dx$

$= \int_0^{a_n} \left[1 - F_{Y_1^n}(x)\right] dx - \int_0^{a_n} F_{Y_1^n}(-x)dx$

$= \int_0^{a_n} \left[1 - F_{X_1}(x) - P\left[X_1 > a_n\right]\right] dx - \int_0^{a_n} \left[F_{X_1}(-x) - P\left[X_1 < -a_n\right]\right] dx$

$= \int_0^{a_n} \left[1 - F_{X_1}(x)\right] dx - a_n P\left[X_1 > a_n\right] - \int_0^{a_n} F_{X_1}(-x)dx + a_n P\left[X_1 < -a_n\right]$

$= \int_0^{a_n} \left[1 - F_{X_1}(x)\right] dx - \int_0^{a_n} F_{X_1}(-x)dx + a_n P\left[X_1 < -a_n\right] - a_n P\left[X_1 > a_n\right]$

Así que, utilizando la proposición 6, se tiener:

$$\lim_{n\to\infty} \mu_n = E\left[X_1\right] = \mu$$

∎

Teorema 11 (Ley débil de los grandes números de Khintchine). *Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ. Entonces:*

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow{P} \mu$$

Demostración

Sea ν es el valor común de $E\left[|X_1|\right]$, $E\left[|X_2|\right], \dots$. Si $\nu = 0$, el resultado es trivial. Supongamos entonces que $\nu > 0$.

Dada $\delta > 0$, definamos, para $n, k \in \mathbb{N}$:

$$a_n = \frac{\delta\varepsilon^2}{8\nu} n \text{ y } Y_k^n = \begin{cases} X_k & \text{si } |X_k| \leq a_n \\ 0 & \text{en otro caso} \end{cases}$$

Por el lema 2, fijando n, las variables aleatorias $Y_1^n, Y_2^n, \dots$ tienen la misma distribución y si μ_n es la esperanza común de $Y_1^n, Y_2^n, \dots$, entonces $\lim_{n\to\infty} \mu_n = \mu$.

Por otra parte, para cualesquiera $n, k \in \mathbb{N}$, se tiene $\left(Y_k^n\right)^2 \leq a_n^2$, así que Y_k^n tiene varianza finita.

Además, $|Y_k^n| \leq |X_k|$ y $|Y_k^n| \leq a_n$, así que, si σ_n^2 es la varianza común de $Y_1^n, Y_2^n, \dots$, se tiene:

$$\sigma_n^2 \leq E\left[\left(Y_k^n\right)^2\right] \leq E\left[a_n |X_k|\right] = a_n E\left[|X_k|\right] = \frac{\delta\varepsilon^2}{8\nu} n E\left[|X_k|\right] = \frac{\delta n \varepsilon^2}{8}$$

Ahora bien, como $\lim_{n\to\infty} \mu_n = \mu$ y $\lim_{n\to\infty} a_n P\left[X_1 > a_n\right] = 0$, existe $N \in \mathbb{N}$ tal que $|\mu_n - \mu| < \frac{\varepsilon}{2}$ y $a_n P\left[X_1 > a_n\right] < \frac{\delta^2}{2}$ para cualquier $n > N$.

Entonces, para $n > N$, se tiene:

$P\left[\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| > \varepsilon\right]$

$\leq P\left[\left|\frac{Y_1^n+\cdots+Y_n^n}{n} - \mu\right| > \varepsilon\right] + P\left[Y_k^n \neq X_k \text{ para alguna } k \leq n\right]$

$\leq P\left[\left|\frac{Y_1^n+\cdots+Y_n^n}{n} - \mu_n\right| > \frac{\varepsilon}{2}\right] + P\left[Y_k^n \neq X_k \text{ para alguna } k \leq n\right]$

Pero, por la desigualdad de Chebyshev, se tiene:

$$P\left[\left|\frac{Y_1^n+\cdots+Y_n^n}{n} - \mu_n\right| > \frac{\varepsilon}{2}\right] \leq \frac{4\sigma_n^2}{n\varepsilon^2} \leq \frac{\delta}{2}$$

Además:

$P\left[Y_k^n \neq X_k \text{ para alguna } k \leq n\right] \leq \sum_{k=1}^n P\left[Y_k^n \neq X_k\right]$

$= \sum_{k=1}^n P\left[|X_k| > a_n\right] = nP\left[X_1 > a_n\right]$

$= \frac{n}{a_n} a_n P[X_1 > a_n] = \frac{1}{\delta} a_n P[X_1 > a_n] < \frac{\delta}{2}$

Así que:

$$P\left[\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| > \varepsilon\right] \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

lo cual prueba el resultado.

■

Desigualdad de Kolmogorov y ley fuerte de los grandes números

Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ. La ley débil de los grandes números establece que $\frac{X_1+\cdots+X_n}{n} \xrightarrow{P} \mu$. En el año 1930 Andrey Nikolaevich Kolmogorov mostró que este resultado puede mejorarse demostrando que la convergencia a μ se da no sólo en probabilidad sino también con probabilidad 1, la cual es un tipo de convergencia más fuerte.

Como vimos antes, si las variables aleatorias tienen varianza finita, la demostración de que la sucesión $Y_n = \frac{X_1+\cdots+X_n}{n}$ converge a μ en probabilidad está basada en la desigualdad de Chebyshev, de la cual se obtiene que $P[|Y_n - \mu| > \varepsilon] \leq \frac{K}{n}$, donde K es una constante. De la proposición 3 puede verse, que para demostrar que la sucesión Y_n converge a μ con probabilidad 1 bastaría con demostrar que $\sum_{n=1}^{\infty} P[|Y_n - \mu| > \varepsilon] < \infty$ para cualquier $\varepsilon > 0$. Para probar esto no basta con aplicar la desigualdad de Chebyshev puesto que ésta únicamente establece que $P[|Y_n - \mu| > \varepsilon] \leq \frac{K}{n}$ y la serie $\sum_{n=1}^{\infty} \frac{1}{n}$ no es convergente.

Kolmogorov logró superar este obstáculo demostrando otra desigualdad.

El resultado de Kolmogorov tiene su origen en el teorema de Borel, publicado en el año 1909.

Teorema 12 (**Desigualdad de Kolmogorov**). *Sean $X_1, \ldots, X_n$ n variables aleatorias independientes de varianza finita y ε cualquier número real positivo, entonces:*

$$P\left[\max_{1\leq j\leq n} |S_j - E[S_j]| > \varepsilon\right] \leq \frac{1}{\varepsilon^2} Var[S_n]$$

donde, para $j \in \{1, \ldots, n\}$, $S_j = \sum_{i=1}^{j} X_i$.

Demostración

Supongamos primero que $E[X_k] = 0$ para cualquier $k \in \{1, \ldots, n\}$. Entonces también se tiene $E[S_k] = 0$ para cualquier $k \in \{1, \ldots, n\}$.

Sea $A = \left\{\omega \in \Omega : \max_{1\leq k\leq n} |S_k(\omega)| > \varepsilon\right\}$ y, para $k \in \{1, \ldots, n\}$:

$$A_k = \left\{ \omega \in A : \max_{1 \leq j \leq k-1} |S_j(\omega)| \leq \varepsilon, |S_k(\omega)| > \varepsilon \right\}$$

donde tomamos $\max\limits_{1 \leq j \leq 0} |S_j(\omega)| \equiv 0$.

Entonces, los eventos $A_1, \ldots, A_n$ son mutuamente excluyentes y $A = \bigcup_{k=1}^n A_k$. Así que:

$E[S_n^2 I_A] = E[S_n^2 \sum_{k=1}^n I_{A_k}] = \sum_{k=1}^n E[S_n^2 I_{A_k}] = \sum_{k=1}^n E\left[(S_k + S_n - S_k)^2 I_{A_k}\right]$

$= \sum_{k=1}^n E\left[\left(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right) I_{A_k}\right]$

$= \sum_{k=1}^n E[S_k^2 I_{A_k}] + 2\sum_{k=1}^n E[S_k(S_n - S_k) I_{A_k}] + \sum_{k=1}^n E\left[(S_n - S_k)^2 I_{A_k}\right]$

Pero, por la proposición **??** y el corolario **??**, $S_k I_{A_k}$ y $S_n - S_k$ son independientes y tienen esperanza finita, de manera que, por la proposición **??**, se tiene:

$$E[S_k I_{A_k}(S_n - S_k)] = E[S_k I_{A_k}] E[S_n - S_k] = E[S_k I_{A_k}] E[S_n - S_k] = 0$$

Por lo tanto:

$Var[S_n] = E[S_n^2] \geq E[S_n^2 I_A] = \sum_{k=1}^n E[S_k^2 I_{A_k}] + \sum_{k=1}^n E\left[(S_n - S_k)^2 I_{A_k}\right]$

$\geq \sum_{k=1}^n E[S_k^2 I_{A_k}] \geq \sum_{k=1}^n \varepsilon^2 E[I_{A_k}] = \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A)$

$= \varepsilon^2 P\left[\max\limits_{1 \leq j \leq n} |S_j - E[S_j]| > \varepsilon\right]$

de lo cual se sigue el resultado.

Para el caso general, sea $Y_k = X_k - E[X_k]$ para $k \in \{1, \ldots, n\}$. Entonces, las variables aleatorias $Y_1, \ldots, Y_n$ son independientes, tienen varianza finita, $\sum_{i=1}^j Y_i = \sum_{i=1}^j (X_i - E[X_i])$ y $E[Y_j] = 0$ para cualquier $j \in \{1, \ldots, n\}$. De manera que si ε es cualquier número real positivo y $S_j = \sum_{i=1}^j X_i$ para cualquier $j \in \{1, \ldots, n\}$, entonces:

$P\left[\max\limits_{1 \leq j \leq n} |S_j - E[S_j]| > \varepsilon\right] = P\left[\max\limits_{1 \leq j \leq n} \left|\sum_{i=1}^j Y_i\right| > \varepsilon\right]$

$\leq \frac{1}{\varepsilon^2} Var\left[\sum_{i=1}^j Y_i\right] = \frac{1}{\varepsilon^2} Var[S_n]$.

■

Teorema 13 (Ley fuerte de los grandes números (1) de Kolmogorov). *Sea* $X_1, X_2, \ldots$ *una sucesión de variables aleatorias, independientes, de varianza finita, esperanza nula y tales que* $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, *donde* σ_n^2 *es la varianza de* X_n. *Entonces:*

$$P\left[\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = 0\right] = 1$$

Demostración

Para cada $n \in \mathbb{N}$ sea $S_n = \sum_{k=1}^{n} X_k$ y, para cada $\varepsilon > 0$, sea:

$$A_\varepsilon = \left\{\omega \in \Omega : \left|\frac{S_n(\omega)}{n}\right| > \varepsilon \text{ para una infinidad de valores de } n\right\}$$

Por la proposición 1, para probar el resultado basta con demostrar que $P(A_\varepsilon) = 0$ para cualquier $\varepsilon > 0$. Para esto definamos:

$$B_{n,\varepsilon} = \left\{\omega \in \Omega : \left|\frac{S_k(\omega)}{k}\right| > \varepsilon \text{ para alguna } k \in \mathbb{N} \text{ tal que } 2^{n-1} < k \leq 2^n\right\}$$

Evidentemente se tiene:

$$A_\varepsilon = \{\omega \in \Omega : \omega \in B_{n,\varepsilon} \text{ para una infinidad de valores de } n\}$$

De manera que, por el lema de Borel-Cantelli, para probar que $P(A_\varepsilon) = 0$ para cualquier $\varepsilon > 0$, basta con demostrar que $\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) < \infty$ para cualquier $\varepsilon > 0$. Pero, utilizando la desigualdad de Kolmogorov, se tiene:

$$P(B_{n,\varepsilon}) = P\left[\max_{2^{n-1}<k\leq 2^n} \left|\frac{S_k}{k}\right| > \varepsilon\right] = P\left[\max_{2^{n-1}<k\leq 2^n} |S_k| > k\varepsilon\right]$$

$$\leq P\left[\max_{2^{n-1}<k\leq 2^n} |S_k| > \varepsilon 2^{n-1}\right] \leq P\left[\max_{1\leq k\leq 2^n} |S_k| > \varepsilon 2^{n-1}\right]$$

$$\leq \frac{1}{\varepsilon^2 2^{2n-2}} Var\left[S_{2^n}\right] = \frac{4}{\varepsilon^2 2^{2n}} \sum_{k=1}^{2^n} \sigma_k^2$$

Así que:

$$\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) \leq \frac{4}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{2^{2n}} \sum_{k=1}^{2^n} \sigma_k^2 = \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}}$$

Sea ahora n_0 el más pequeño número natural tal que $k \leq 2^{n_0}$, entonces:

$$\sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}} = \sum_{n=n_0}^{\infty} \frac{1}{2^{2n}} = \frac{4}{2^{2n_0}} \leq \frac{4}{k^2}$$

Así que:

$$\sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}} \leq 4 \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$$

Por lo tanto:

$$\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) \leq \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n\in\mathbb{N}:k\leq 2^n\}} \frac{1}{2^{2n}} < \infty$$

■

Corolario 10. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes, de varianza finita y tales que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, donde σ_n^2 es la varianza de X_n. Entonces:*

$$P\left[\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (X_k - E[X_k]) = 0\right] = 1$$

Para el caso en que las variables aleatorias $X_1, X_2, \ldots$ sean idénticamente distribuidas se cumple la ley fuerte con la única condición de que la esperanza común de $X_1, X_2, \ldots$ sea finita. La demostración de este resultado se debe también a Kolmogorov y el método de demostración es el de truncación. Se requieren además algunos resultados previos, los cuales se exponen a continuación.

Lema 3. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas de esperanza finita μ. Para $n \in \mathbb{N}$, definamos:*

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Entonces:

1. $\lim_{n\to\infty} E[Y_n] = \mu$.
2. Y_n tiene varianza finita para cualquier $n \in \mathbb{N}$.
3. $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, donde σ_n^2 es la varianza de Y_n.
4. $P[\{\omega \in \Omega : \text{existe } N(\omega) \text{ tal que } Y_n(\omega) = X_n(\omega) \text{ para cualquier } n \geq N(\omega)\}] = 1$.

Demostración

1. Se tiene:

$$F_{Y_n}(x) = \begin{cases} 0 & \text{si } x < -n \\ P[-n \leq X_n \leq x] & \text{si } -n \leq x < 0 \\ P[|X_n| > n] + P[-n \leq X_n \leq x] & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -n \\ P[-n \leq X_n \leq x] & \text{si } -n \leq x < 0 \\ 1 - P[x < X_n \leq n] & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

Así que:

$E[Y_n] = \int_0^{\infty}[1 - F_{Yn}(x)]\,dx - \int_0^{n} F_{Yn}(-x)dx$

$= \int_0^{n} P[x < X_n \leq n]\,dx - \int_0^{n} P[-n \leq X_n \leq -x]\,dx$

$= \int_0^{n} P[x < X_1 \leq n]\,dx - \int_0^{n} P[-n \leq X_1 \leq -x]\,dx$

$= \int_0^{n}[1 - F_{X_1}(x)]\,dx - \int_0^{n} P[X_1 > n]\,dx - \int_0^{n} F_{X_1}(-x)dx. + \int_0^{n} P[X_1 < -n]\,dx$

$= \int_0^{n}[1 - F_{X_1}(x)]\,dx - \int_0^{n} F_{X_1}(-x)dx - nP[X_1 > n] + nP[X_1 < -n]$

Por lo tanto, utilizando la proposición 6, $\lim_{n\to\infty} E[Y_n] = E[X_1] = \mu$.

2. Para cualquier $n \in \mathbb{N}$, se tiene $|Y_n| \leq n$, así que Y_n tiene varianza finita.

3. $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} E[Y_n^2] = \sum_{n=1}^{\infty} \frac{1}{n^2} E\left[X_n^2 I_{[|X_n| \leq n]}\right]$

$= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=1}^{n} E\left[X_n^2 I_{[j-1<|X_n|\leq j]}\right]$

$= E\left[X_1^2 I_{[j-1<|X_1|\leq j]}\right] + \frac{1}{2^2}\left(E\left[X_1^2 I_{[j-1<|X_1|\leq j]}\right] + E\left[X_2^2 I_{[j-1<|X_2|\leq j]}\right]\right)$

$+ \ldots$

$= E\left[X_1^2 I_{[j-1<|X_1|\leq j]}\right]\left(1 + \frac{1}{2^2} + \cdots\right) + E\left[X_2^2 I_{[j-1<|X_2|\leq j]}\right]\left(\frac{1}{2^2} + \frac{1}{3^2} + \cdots\right)$

$+ \cdots$

$= \sum_{j=1}^{\infty} E\left[X_j^2 I_{[j-1<|X_j|\leq j]}\right] \sum_{n=j}^{\infty} \frac{1}{n^2}$

Pero, para cualquier $j \in \{2, 3, \ldots\}$, se tiene:

$$\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \int_{j-1}^{\infty} \frac{1}{x^2} = \frac{1}{j-1} \leq \frac{2}{j}$$

Además,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \sum_{n=2}^{\infty} \frac{1}{n^2} \leq 2$$

Así que, $\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \frac{2}{j}$ para cualquier $j \in \mathbb{N}$.

Además, tomando en cuenta que $X_1, X_2, \ldots$ tienen la misma distribución:

$$E\left[X_j^2 I_{[j-1<|X_j|\leq j]}\right] \leq jE\left[|X_j|\, I_{[j-1<|X_j|\leq j]}\right] = jE\left[|X_1|\, I_{[j-1<|X_1|\leq j]}\right]$$

Por lo tanto:

$\sum_{j=1}^{\infty} E\left[X_j^2 I_{[j-1<|X_j|\leq j]}\right] \sum_{n=j}^{\infty} \frac{1}{n^2} \leq \sum_{j=1}^{\infty} jE\left[|X_1|\, I_{[j-1<|X_1|\leq j]}\right] \frac{4}{j}$

$= 4 \sum_{j=1}^{\infty} E\left[|X_1|\, I_{[j-1<|X_1|\leq j]}\right]$

Sea ahora $Z_n = \sum_{j=1}^{n} |X_1|\, I_{[j-1<|X_1|\leq j]}$, entonces la sucesión de variables aleatorias $Z_1, Z_2, \ldots$ es no decreciente y $\lim_{n\to\infty} Z_n(\omega) = |X_1(\omega)|$ para cualquier $\omega \in \Omega$, así que por el teorema de la convergencia monótona:

$$\sum_{j=1}^{\infty} E\left[|X_1|\, I_{[j-1<|X_1|\leq j]}\right] = \lim_{n\to\infty} E\left[Z_n\right] = E\left[|X_1|\right] < \infty$$

de lo cual se sigue que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$.

4. $P\left[Y_n \neq X_n\right] = P\left[|X_n| > n\right] = P\left[|X_1| > n\right]$.

De manera que, utilizando la proposición **??**:

$$\sum_{n=1}^{\infty} P\left[Y_n \neq X_n\right] = \sum_{n=1}^{\infty} P\left[|X_1| > n\right] \leq \sum_{n=1}^{\infty} P\left[|X_1| \geq n\right] < \infty$$

Así que, por el lema de Borel-Cantelli, si:

$$A = \{\omega \in \Omega : Y_n(\omega) \neq X_n(\omega) \text{ para una infinidad de valores de } n\}$$

entonces $P(A) = 0$.

Sea ahora:

$$B = \{\omega \in \Omega : \text{existe } N(\omega) \text{ talque } Y_n(\omega) = X_n(\omega) \text{ para cualquier } n \geq N(\omega)\}$$

Entonces, $B \supset A^c$, así que, $P(B) \geq P(A) = 1$.

■

Corolario 11. *Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas de esperanza finita. Para $n \in \mathbb{N}$, definamos:*

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Entonces:

$$P\left[\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left[X_k(\omega) - Y_k(\omega)\right] = 0\right\}\right] = 1$$

Demostración

Por la parte *iv* del lema 3, si:

$$B = \{\omega \in \Omega : \text{existe } N(\omega) \text{ talque } Y_n(\omega) = X_n(\omega) \text{ para cualquier } n \geq N(\omega)\}$$

entonces $P(B) = 1$.

Pero si $\omega \in B$, entonces existe $N(\omega)$ tal que $X_n(\omega) - Y_n(\omega) = 0$ para cualquier $n \geq N(\omega)$, así que:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\left[X_k(\omega) - Y_k(\omega)\right] = 0$$

$\blacksquare$

Lema 4. *Sea* (x_n) *una sucesión convergente de números reales y sea* $x = \lim_{n\to\infty} x_n$. *Entonces la sucesión* $z_n = \frac{1}{n}\sum_{k=1}^{n} x_k$ *es convergente y* $\lim_{n\to\infty} z_n = x$.

Demostración

Sea $M > 0$ tal que $|x - x_n| \leq M$ para cualquier $n \in \mathbb{N}$.

Dada $\varepsilon > 0$, sea $m \in \mathbb{N}$ tal que $|x - x_n| < \frac{\varepsilon}{2}$ para cualquier $n \geq m$.

Entonces, para $n > \max\left\{m, \frac{2mM}{\varepsilon}\right\}$, se tiene:

$|z_n - x| = \left|\frac{1}{n}\sum_{k=1}^{n} x_k - x\right| = \left|\frac{1}{n}\sum_{k=1}^{n}(x_k - x)\right| \leq \frac{1}{n}\sum_{k=1}^{n}|x_k - x|$

$= \frac{1}{n}\sum_{k=1}^{m}|x_k - x| + \frac{1}{n}\sum_{k=m+1}^{n}|x_k - x|$

$\leq \frac{mM}{n} + \frac{(n-m)\varepsilon}{2n} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$

lo cual significa que $\lim_{n\to\infty} z_n = x$.

$\blacksquare$

Teorema 14 (Ley fuerte de los grandes números (2) de Kolmogorov). *Sea* $X_1, X_2, \ldots$ *una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita* μ. *Entonces:*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{c.s.} \mu$$

Demostración

Para cada $n \in \mathbb{N}$, sea:

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Por el lema 3, las variables aleatorias $Y_1, Y_2, \ldots$ tienen esperanza finita, $\lim_{n\to\infty} E\left[Y_n\right] = \mu$ y $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, donde σ_n^2 es la varianza de Y_n. De manera que, por el lema 4 y el corolario 10, se tiene:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} E\left[Y_k\right] = \mu$$

$$P\left[\left\{\omega \in \Omega : \lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{k=1}^{n} \left(Y_k(\omega) - E\left[Y_k\right]\right) = 0\right\}\right] = 1$$

de lo cual se obtiene:

$$P\left[\left\{\omega \in \Omega : \lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{k=1}^{n} Y_k(\omega) = \mu\right\}\right] = 1$$

Además, por el corolario 11:

$$P\left[\left\{\omega \in \Omega : \lim_{n\to\infty} \tfrac{1}{n} \textstyle\sum_{k=1}^{n} \left[X_k(\omega) - Y_k(\omega)\right] = 0\right\}\right] = 1$$

de lo cual se obtiene el resultado.

■