Capítulo 5

Sucesiones de Variables Aleatorias

5.1. Introducción

La introducción de los conceptos de medida e integral nos va a permitir considerar nuevos modos de convergencia para sucesiones de funciones. Vamos a comenzar por recordar dos conceptos clásicos de convergencia.

Definición 5.1 Sea $X_n,\ n \geq 1$ una sucesión de funciones medibles a valores reales definidas sobre el espacio de medida $(\Omega, \mathcal{F}, \mu)$ y X una función real y medible, definida sobre el mismo espacio. Decimos que la sucesión (X_n) *converge (puntualmente)* a X si para todo $\omega \in \Omega$ se tiene que

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

es decir, dados $\omega \in \Omega$ y $\varepsilon > 0$ existe $N = N(\omega, \varepsilon) \in \mathbb{N}$ tal que

$$n \geq N \Rightarrow |X_n(\omega) - X(\omega)| < \varepsilon$$

Usamos la notación

$$X_n \to X$$

para denotar la convergencia puntual.

Definición 5.2 Sean $X_n,\ n \geq 1$ y X como en la definición anterior. Decimos que la sucesión (X_n) *converge uniformemente* a X si dado $\varepsilon > 0$ existe $N = N(\varepsilon) \in \mathbb{N}$ tal que, para cualquier punto ω en Ω se tiene que

$$n \geq N \Rightarrow |X_n(\omega) - X(\omega)| < \varepsilon$$

Usamos la notación

$$X_n \xrightarrow{U} X$$

para denotar la convergencia uniforme.

Observación 5.1

- La diferencia entre las dos definiciones radica en que en el primer caso N depende del punto ω en el cual estamos viendo la convergencia, mientras que en el segundo caso el mismo N sirve para todos los puntos del espacio.
- La segunda definición dice que si consideramos una banda de ancho 2ε alrededor de la función X, existe un N que depende únicamente de ε tal que, si $n \geq N$ la función X_n está dentro de esta banda.
- Es claro que convergencia uniforme implica convergencia puntual pero el recíproco es falso.

5.2. Convergencia Casi Segura

Definición 5.3 Sean $X_n,\ n \geq 1$ y X como en la definición 5.1. Decimos que la sucesión (X_n) *converge casi seguramente* a X si existe un conjunto nulo $N \in \mathcal{F}$ tal que, para cualquier punto $\omega \notin N$ se tiene que

$$X_n(\omega) \to X(\omega).$$

Usamos las notaciones

$$X_n \xrightarrow{c.s.} X \quad o \quad X_n \to X,\ c.s.$$

para denotar la convergencia casi segura. Si $(\Omega, \mathcal{F}, \mu)$ es un espacio de probabilidad este tipo de convergencia se conoce como *convergencia con probabilidad 1* y se denota

$$X_n \xrightarrow{c.p.1} X \quad o \quad X_n \to X,\ c.p.1$$

Es obvio que convergencia puntual implica convergencia c.s. y el siguiente ejemplo muestra que el recíproco no es cierto.

Ejemplo 5.1
Sea $(\Omega, \mathcal{F}, \mu)$ el espacio de Lebesgue $([0,1], \mathcal{B}, \lambda)$ y $X_n(\omega) = \omega^n,\ X(\omega) = 0$ para $\omega \in [0,1]$. Entonces

$$X_n(\omega) \to 0 \quad \text{para } \omega \in [0,1),$$

es decir, la sucesión X_n converge a X salvo en el punto 1, de modo que el conjunto donde no hay convergencia es un conjunto nulo.

Proposición 5.1 *Si $(\Omega, \mathcal{F}, \mu)$ es un espacio completo, $X_n \xrightarrow{c.s.} X$ y las funciones X_n son medibles, entonces X también es medible.*

Demostración. Sabemos que existe un conjunto nulo (y por lo tanto medible) N tal que X_n converge a X fuera de N, es decir, si $\omega \notin N$

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

Para $c \in \mathbb{R}$ los conjuntos

$$A_1 = \{\omega : X(\omega) > c\}, \qquad A_2 = \{\omega : \liminf X_n(\omega) > c\}$$

no necesariamente coinciden, pero cualquier punto que esté en uno y no esté en otro debe estar en N. Además, sabemos que A_2 es medible porque $\liminf X_n$ es medible. Sea

$$B_1 = A_1 - A_2, \qquad B_2 = A_2 - A_1.$$

Ambos conjuntos son subconjuntos de N y como el espacio es completo, son conjuntos medible. Finalmente observamos que

$$A_1 = (A_2 \cup B_1) \cap B_2^c$$

de modo que F_1 es medible. ■

Definición 5.4 Sea $X_n,\ n \geq 1$ como en la definición 5.1. Decimos que la sucesión (X_n) *es de Cauchy casi seguramente* si existe un conjunto nulo $N \in \mathcal{F}$ tal que, para cualquier punto $\omega \notin N$ la sucesión $X_n(\omega)$ es de Cauchy.

Proposición 5.2 *Una sucesión de funciones medibles (X_n) converge c.s. si y sólo si es de Cauchy c.s.*

La demostración queda como ejercicio.

5.3. Convergencia en Medida

Definición 5.5 Sean $X_n,\ n \geq 1$ y X como en la definición 5.1. Decimos que la sucesión (X_n) *converge en medida* a X si para cualquier $\varepsilon > 0$ se tiene que

$$\lim_{n\to\infty} \mu(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

Usamos las notaciones

$$X_n \xrightarrow{\mu} X \quad o \quad X_n \to X, \text{ en medida}$$

para denotar la convergencia en medida. Si $(\Omega, \mathcal{F}, \mu)$ es un espacio de probabilidad este tipo de convergencia se conoce como *convergencia en probabilidad* y se denota

$$X_n \xrightarrow{P} X \quad o \quad X_n \to X, \text{ en probabilidad.}$$

Ese tipo de convergencia es más débil que la convergencia casi segura, como lo muestran la siguiente la proposición y el ejemplo que le sigue.

Proposición 5.3 *Sean $X_n,\ n \geq 1$ y X como en la definición 5.1 y supongamos que $\mu(\Omega) < \infty$. Si X_n converge a X c.s. entonces también converge en medida.*

Demostración. Si $X_n \to X$ c.s. entonces para cualquier $\varepsilon > 0$

$$\begin{aligned} 0 &= \mu(\{|X_n - X| > \varepsilon\} \text{ para infinitos } n) \\ &= \mu(\limsup \{|X_n - X| > \varepsilon\}) \\ &= \lim_{n\to\infty} \mu(\cup_{k\geq n}\{|X_k - X| > \varepsilon\}) \\ &\geq \lim_{n\to\infty} \mu(\{|X_n - X| > \varepsilon\}). \end{aligned}$$

■

Para ver que el recíproco no es cierto tenemos el siguiente ejemplo.

Ejemplo 5.2
Sea $(\Omega, \mathcal{F}, \mu) = ([0,1], \mathcal{B}, \lambda)$ el espacio de Lebesgue y definimos la siguiente sucesión de variables aleatorias:

$$\begin{array}{lll} X_1(\omega) = \mathbf{1}_{[0,1]}(\omega), & & \\ X_2(\omega) = \mathbf{1}_{[0,1/2)}(\omega), & X_3(\omega) = \mathbf{1}_{[1/2,1]}(\omega) & \\ X_4(\omega) = \mathbf{1}_{[0,1/3}(\omega), & X_5(\omega) = \mathbf{1}_{[1/3,2/3)}(\omega), & X_6(\omega) = \mathbf{1}_{[2/3,1]}(\omega), \\ \vdots & & \end{array}$$

Observamos que como las variables X_n son funciones indicadoras de intervalos, sólo son distintas de 0 en esos intervalos, cuya longitud tiende a 0 cuando $n \to \infty$. Esto muestra que $X_n \to 0$ en probabilidad. En cambio, para cualquier $\omega \in [0,1]$ tenemos que $X_n(\omega) = 1$ para infinitos valores de n, y $X_n(\omega) = 0$ para el resto, que también son infinitos. Por lo tanto la sucesión X_n no converge puntualmente en ningún punto.

Ejemplo 5.3
Para ver que $\mu(\Omega) < \infty$ es una condición necesaria en el teorema anterior consideramos el siguiente ejemplo: Sea $X_n = \mathbf{1}_{(-n,n)^c}$, $X = 0$, entonces observamos que $X_n(\omega) \to X(\omega)$ para todo ω pero para cualquier $n \in \mathbb{N}$ y cualquier $\varepsilon > 0$,

$$\mu(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = \infty$$

y por lo tanto no hay convergencia en medida.

5.3.1. Aplicaciones Estadísticas

Supongamos que tenemos una familia de modelos $(\Omega, \mathcal{F}, P_\theta), \theta \in \Theta$, y observamos una muestra aleatoria simple $X_1, \dots, X_n$. Con base en estas observaciones queremos estimar el valor del parámetro θ, es decir, deseamos seleccionar el modelo correcto.

En esta situación lo usual es usar un *estadístico* $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ para estimar el valor de θ. Este estadístico es una variable aleatoria definida sobre el mismo espacio de probabilidad. Decimos que el estimador es *débilmente consistente* si para todo $\theta \in \Theta$,

$$P_\theta(|\hat{\theta}_n - \theta| > \varepsilon) \to 0, \qquad n \to \infty,$$

es decir, si $\hat{\theta}_n$ converge en probabilidad al verdadero valor de θ. El estimador es *(fuertemente) consistente* si esta convergencia se da con probabilidad 1.

5.3.2. Consecuencias de la Convergencia en Medida

Aún cuando hemos visto que convergencia en medida es más débil que convergencia c.s., el siguiente teorema, debido a Riesz, muestra que si tenemos convergencia en medida, siempre hay una subsucesión que converge c.s. al mismo límite.

Teorema 5.1 (Riesz) *Si la sucesión X_n converge en medida a X en Ω, existe una subsucesión (X_{n_k}) que converge c.s. a X.*

Demostración. Como la sucesión de funciones converge c.s. podemos hallar $n_1 \in \mathbb{N}$ tal que

$$\mu(\{\omega \in \Omega : |X_{n_1}(\omega) - X(\omega)| \geq 2^{-1}\}) < 2^{-1}.$$

Para cada $k > 1$ hallamos $n_k > n_{k-1}$ tal que

$$\mu(\{\omega \in \Omega : |X_{n_k}(\omega) - X(\omega)| \geq 2^{-k}\}) < 2^{-k}.$$

Veamos que la subsucesión (X_{n_k}) converge c.s. a X. Sea

$$S_k = \bigcup_{i \geq k} \{\omega \in \Omega : |X_{n_i}(\omega) - X(\omega)| \geq 2^{-i}\}.$$

Observamos que S_i es una sucesión decreciente de conjuntos. Sea

$$S = \bigcap_{k \geq 1} S_k.$$

Por sub-aditividad tenemos que la medida de S_k está acotada por

$$\mu(S_k) \leq 2^{-k} + 2^{-k-1} + 2^{-k-2} + \cdots = 2^{1-k}$$

y en consecuencia la medida de S es

$$\mu(S) = \lim_{k \to \infty} \mu(S_k) = 0.$$

Veamos que si $\omega \in \Omega - S$ entonces $X_{n_k}(\omega) \to X(\omega)$. Sea $\varepsilon > 0$, escogemos K de modo que $2^{-K} < \varepsilon$ y $\omega \notin S_K$. Entonces para todo $k \geq K$, $\omega \notin S_k$ y en consecuencia

$$|X_{n_k}(\omega) - X(\omega)| < 2^{-k} < \varepsilon.$$

■

Corolario 5.1 *X_n converge en medida a X si y sólo si cada subsucesión (X_{n_k}) contiene una subsucesión $(X_{n_{k_i}})$ que converge a X c.s.*

Demostración. Si $X_n \xrightarrow{\mu} X$ y (X_{n_k}) es una subsucesión, entonces $X_{n_k} \xrightarrow{\mu} X$ y por el teorema anterior existe una subsucesión $(X_{n_{k_i}})$ que converge a X c.s.

Recíprocamente, supongamos que toda subsucesión tiene una subsucesión que converge c.s. a X y supongamos que X_n no converge a X en probabilidad para obtener una contradicción.

Si X_n no converge en probabilidad a X, existen una subsucesión (X_{n_k}), $\delta > 0$ y $\varepsilon > 0$ tales que

$$\mu(|X_{n_k} - X| > \varepsilon) \geq \delta. \tag{5.1}$$

Pero esta subsucesión (X_{n_k}) debería tener una subsucesión que converge c.s. a X y por lo tanto en probabilidad. Esto contradice a (5.1). ∎

Corolario 5.2 *a) Si $X_n \to X$ c.s. y $g : \mathbb{R} \to \mathbb{R}$ es continua entonces $g(X_n) \to g(X)$ c.s.*

b) Si $X_n \xrightarrow{\mu} X$ y $g : \mathbb{R} \to \mathbb{R}$ es continua entonces $g(X_n) \xrightarrow{\mu} g(X)$.

Demostración. (a) Existe un conjunto nulo $N \in \mathcal{B}$ tal que si $\omega \notin N$ entonces $X_n(\omega) \to X(\omega)$ en $\mathbb{R}$. Por continuidad de g, si $\omega \notin N$,

$$g(X_n(\omega)) \to g(X(\omega)),$$

es decir, $(g(X_n))$ converge c.s. a $g(X)$.

(b) Sea $(g(X_{n_k}))$ una subsucesión de $(g(X_n))$. Basta hallar una subsucesión $(g(X_{n_{k_i}}))$ que sea c.s. convergente. Sabemos que (X_{n_k}) tiene una subsucesión c.s. convergente $(X_{n_{k_i}})$ que converge a X. Por lo tanto $g(X_{n_{k_i}}) \to g(X)$ c.s. ∎

Corolario 5.3 (de Convergencia Dominada de Lebesgue) *Si $X_n \xrightarrow{\mu} X$ y existe una función medible $Y \in L^1$ tal que $|X_n| \leq Y$, entonces*

$$\int X_n \, d\mu \to \int X \, d\mu.$$

Demostración. Basta demostrar que toda subsucesión convergente de $\int X_n \, d\mu$ converge a $\int X \, d\mu$.

Supongamos que $\int X_{n_k} \, d\mu$ converge. Por hipótesis tenemos convergencia en probabilidad y en consecuencia (X_{n_k}) tiene una subsucesión $(X_{n_{k_i}})$ que converge c.s. a X. Por el TCD tenemos que

$$\int X_{n_{k_i}} \, d\mu \to \int X \, d\mu$$

En consecuencia $\int X_{n_k} \, d\mu \to \int X \, d\mu$. ∎

Proposición 5.4 (Propiedades) *Si $X_n \xrightarrow{\mu} X$ y $Y_n \xrightarrow{\mu} Y$*

1. *$X_n + Y_n \xrightarrow{\mu} X + Y$.*

2. *$X_n Y_n \xrightarrow{\mu} XY$.*

Demostración. (1) Observamos que

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

Tomando medidas, usando subaditividad y haciendo $n\to\infty$ se obtiene el resultado.

(2) Observamos que basta demostrar que para toda subsucesión (n_k) existe una subsucesión n_{k_i} tal que

$$X_{n_{k_i}}Y_{n_{k_i}}\xrightarrow{c.s.}XY.$$

Como $X_{n_k}\xrightarrow{\mu}X$, existe una subsucesión (n'_k) tal que

$$X_{n'_k}\xrightarrow{c.s.}X.$$

Como $Y_n\xrightarrow{\mu}Y$, dada la subsucesión (n'_k), existe una subsucesión (n'_{k_i}) tal que

$$X_{n'_{k_i}}\xrightarrow{c.s.}X,\qquad Y_{n'_{k_i}}\xrightarrow{c.s.}Y,$$

y en consecuencia tenemos

$$X_{n'_{k_i}}Y_{n'_{k_i}}\xrightarrow{c.s.}XY.$$

Por lo tanto, toda subsucesión de (X_nY_n) tiene una subsucesión que converge c.s. ■

Proposición 5.5 (Ley Débil de Grandes Números) *Si* $(X_n,\ n\geq 1)$ *son i.i.d. con* $\mathrm{E}(X_n)=\mu$ *y* $\mathrm{Var}(X_n)=\sigma^2$*, entonces*

$$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu$$

Demostración. Basta usar la desigualdad de Chebyshef. ■

5.4. Convergencia en L^p

En esta sección estudiamos un nuevo tipo de convergencia que tiene gran importancia en el análisis funcional. Recordemos que una función medible X pertenece al espacio L^p si $\int|X|^p\,d\mu<\infty$. Para $p\geq 1$ y $X,Y\in L^p$ definimos una distancia en este espacio por

$$d_p(X,Y)=\Big(\int|X-Y|^p d\mu\Big)^{1/p}.$$

Para ver que d_p es una distancia demostraremos más adelante la desigualdad triangular, que se conoce como la desigualdad de Minkowski. Por otro lado, es claro que $d_p(X,Y)=d_p(Y,X)$. Además si $X=Y$ tenemos que $d_p(X,Y)=0$ pero el recíproco no es cierto: Si $d_p(X,Y)=0$ sólo podemos concluir que $X\overset{c.s.}{=}Y$. Por lo tanto d_p sólo será una *pseudo-distancia*.

Para tener una distancia podemos tomar la relación de equivalencia $X\overset{c.s.}{=}Y$ sobre el espacio L^p y en lugar de considerar las funciones medibles X consideramos las clases de equivalencia $\widetilde{X}$. El espacio cociente, formado por las clases de equivalencia, lo denotamos $\mathcal{L}^p$ y d_p si será una distancia sobre este espacio. En lo que sigue vamos a obviar este procedimiento y hablaremos de d_p como una distancia sobre el espacio L^p.

Hay una norma que induce esta distancia:

$$||X||_p=\Big(\int(|X|^p)\Big)^{1/p}.$$

Definición 5.6 Decimos que la sucesión $(X_n,\ n \geq 1)$ converge a X en L^p si todas las variables están en L^p y

$$\int |X_n - X|^p \, d\mu \to 0$$

cuando $n \to \infty$. Usamos la notación

$$X_n \xrightarrow{L^p} X.$$

Proposición 5.6 *Convergencia en L^p implica convergencia en medida.*

Demostración. Tenemos

$$\begin{aligned}
\mu(|X_n - X| \geq \varepsilon) &= \int \mathbf{1}_{\{|X_n - X| > \varepsilon\}} d\mu \\
&= \int \mathbf{1}_{\{|X_n - X|^p > \varepsilon^p\}} d\mu \\
&\leq \int \frac{|X_n - X|^p}{\varepsilon^p} \mathbf{1}_{\{|X_n - X|^p > \varepsilon^p\}} d\mu \\
&\leq \frac{1}{\varepsilon^p} \int |X_n - X|^p \, d\mu \to 0
\end{aligned}$$

cuando $n \to \infty$. ■

El siguiente ejemplo muestra que el recíproco no es cierto.

Ejemplo 5.4
De nuevo, el espacio de medida es el espacio de probabilidad de Lebesgue $([0,1], \mathcal{B}, \lambda)$ y definimos

$$X_n = 2^n \mathbf{1}_{(0, \frac{1}{n})}.$$

Entonces

$$P(|X_n| > \varepsilon) = \lambda((0, \frac{1}{n})) = \frac{1}{n} \to 0$$

pero

$$\mathrm{E}(|X_n|^p) = 2^{np} \frac{1}{n} \to \infty.$$

Convergencia en L^p no implica convergencia c.s., como lo muestra el ejemplo 5.2. Para cualquier $p > 0$, como la variable X_n sólo toma valores 0 ó 1, el valor esperado $\mathrm{E}(|X_n|^p)$ es igual a la longitud del intervalo correspondiente a X_n, y esta longitud tiende a 0.

Convergencia casi segura tampoco implica convergencia en L^p, como lo muestra el ejemplo 5.3. Aún cuando la medida del espacio sea finita, es posible dar ejemplo de sucesiones que convergen c.s. pero no convergen en L^p.

Desigualdades

Teorema 5.2 (Desigualdad de Hölder) *Sea p, q números reales tales que $p, q > 1$ y*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{5.2}$$

Sean $X \in L^p$, $Y \in L^q$, entonces el producto XY es integrable y

$$\int |XY| \, d\mu \leq \Big(\int |X|^p \, d\mu \Big)^{1/p} \Big(\int |Y|^q \, d\mu \Big)^{1/q}. \tag{5.3}$$

En términos de las normas la desigualdad dice que

$$||XY||_1 \leq ||X||_p ||Y||_q.$$

Si p y q satisfacen (5.2) decimos que son *conjugados*.

Demostración. Observamos inicialmente que si $\int |X|^p \, d\mu = 0$ entonces $X = 0$ *c.s.* y en consecuencia $\int |XY| \, d\mu = 0$, de modo que la desigualdad es cierta. Algo similar ocurre si $\int |Y|^q \, d\mu = 0$, así que podemos suponer que el lado derecho de (5.3) es estrictamente positivo.

Dados $a > 0, b > 0$, existen $s, t \in \mathbb{R}$ tales que

$$a = \exp\{s/p\}, \qquad b = \exp\{t/q\}.$$

Como la función exponencial es convexa y $p^{-1} + q^{-1} = 1$ tenemos que

$$\exp\{p^{-1}s + q^{-1}t\} \leq p^{-1}\exp\{s\} + q^{-1}\exp\{t\},$$

y usando la definición de s, t,

$$ab \leq p^{-1}a^p + q^{-1}b^q.$$

Ahora reemplazamos a por $|X|/||X||_p$ y b por $|Y|/||Y||_q$ y obtenemos

$$\frac{|XY|}{||X||_p ||Y||_q} \leq p^{-1}\Big(\frac{|X|}{||X||_p}\Big)^p + q^{-1}\Big(\frac{|Y|}{||Y||_q}\Big)^q$$

Finalmente, tomando integrales,

$$\frac{1}{||X||_p ||Y||_q} \int |XY| \, d\mu \leq p^{-1} + q^{-1} = 1$$

■

Teorema 5.3 (Desigualdad de Minkowski) *Supongamos que $X, Y \in L^p$ para $1 \leq p < \infty$. Entonces $X + Y \in L^p$ y*

$$||X + Y||_p \leq ||X||_p + ||Y||_p.$$

Demostración. La desigualdad

$$|X + Y|^p \leq 2(|X|^p \vee |Y|^p) \leq 2(|X|^p + |Y|^p) \in L^p$$

muestra que los espacios L^p, $p \geq 1$ son cerrados aditivamente. So $p = 1$ la desigualdad de Minkowski sigue de la desigualdad triangular usual. Sea $1 < p < \infty$ y escojamos q conjugado de p de modo que $p^{-1} + q^{-1} = 1$, es decir, $p - 1 = p/q$. Usando la desigualdad de Hölder,

$$\begin{aligned}
||X + Y||_p^p &= \int |X + Y|^p \, d\mu = \int |X + Y||X + Y|^{p/q} \, d\mu \\
&\leq \int |X||X + Y|^{p/q} \, d\mu + \int |Y||X + Y|^{p/q} \, d\mu \\
&\leq ||X||_p || \, |X + Y|^{p/q}||_q + ||Y||_p || \, |X + Y|^{p/q}||_q \\
&= (||X||_p + ||Y||_p) || \, |X + Y|^{p/q}||_q \\
&= (||X||_p + ||Y||_p) \Big(\int |X + Y|^p \, d\mu \Big)^{1/q} \\
&= (||X||_p + ||Y||_p) ||X + Y||_p^{p/q} \\
&= (||X||_p + ||Y||_p) ||X + Y||_p^{p-1}
\end{aligned}$$

Si el último factor es distinto de 0, podemos dividir por él para obtener el resultado. Si es igual a 0, ambos lados de la desigualdad valen 0. ■

La desigualdad anterior es la desigualdad triangular para $|| \cdot ||_p$.

Teorema 5.4 (Desigualdad de Jensen) *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, $g : \mathbb{R} \to \mathbb{R}$ una función convexa y X una variable aleatoria integrable tal que $g(X)$ también es integrable. Entonces*

$$\mathrm{E}(g(X)) \geq g(\mathrm{E}(X)).$$

Demostración. Sea g una función convexa, sabemos que dado cualquier punto de la gráfica de g podemos hallar (al menos) una recta que es tangente a la curva en ese punto y tal que la recta siempre está por debajo de la curva que representa a g. Esta recta (que en general no es única) se conoce como la recta de soporte de g. Sea $r(x) = ax + b$ la recta de soporte de g en el punto $(\mathrm{E}(X), g(\mathrm{E}(X)))$ sobre la gráfica de función g. Entonces

$$aX(\omega) + b \leq g(X(\omega)).$$

Tomando esperanza obtenemos

$$a\,\mathrm{E}(X) + b \leq \mathrm{E}(g(X)).$$

Pero como la recta $r(x)$ es tangente a la función g en $(\mathrm{E}(X), g(\mathrm{E}(X)))$, $a\,\mathrm{E}(X) + b = \mathrm{E}(g(X))$. ■

Ejemplo 5.5
Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, X una variable aleatoria y $0 < \alpha < \beta$. Definimos

$$r = \frac{\beta}{\alpha} > 1, \qquad s = \frac{\beta}{\beta - \alpha}.$$

Entonces

$$\frac{1}{r} + \frac{1}{s} = \frac{\alpha}{\beta} + \frac{\beta - \alpha}{\beta} = 1.$$

Ahora ponemos $Z = |X|^\alpha$, $\quad Y = 1$ usamos la desigualdad de Hölder:

$$\mathrm{E}(|ZY|) \leq (\mathrm{E}\,|Z|^r)^{1/r} (\mathrm{E}(|Y|^s)^{1/s},$$

es decir

$$\mathrm{E}(|X|^\alpha) \leq (\mathrm{E}\,|X|^{r\alpha})^{1/r} 1 = (\mathrm{E}\,|X|^\beta)^{\alpha/\beta}$$

de modo que

$$(\mathrm{E}(|X|^\alpha))^{1/\alpha} \leq (\mathrm{E}\,|X|^\beta)^{1/\beta}$$

y

$$||X||_\alpha \leq ||X||_\beta.$$

Concluimos que $X \in L^\beta \Rightarrow X \in L^\alpha$ si $\alpha < \beta$. Más aún, $||X||_\alpha = (\mathrm{E}(|X|^\alpha))^{1/\alpha}$ es no-decreciente en α. Como consecuencia tenemos que si $X_n \xrightarrow{L^p} X$ y $r < p$, $X_n \xrightarrow{L^r} X$.

Teorema 5.5 *Para $p \geq 1$ el espacio L^p es completo, es decir, si (X_n) es una sucesión de Cauchy en L^p, existe $X \in L^p$ tal que $X_n \xrightarrow{L^p} X$.*

Demostración. Dado $\varepsilon > 0$ sea $N = N(\varepsilon) \in \mathbb{N}$ tal que si $n, m \geq N$,

$$\int |X_n - X_m|^p \, d\mu < \varepsilon^{p+1}. \tag{5.4}$$

Llamemos $N_k = N(\varepsilon 2^{-k})$ y supongamos que $N_{k+1} > N_k$ para todo k. Definimos los conjuntos

$$A(\varepsilon, m, n) = \{\omega : |X_m(\omega) - X_n(\omega)| \geq \varepsilon\}$$

entonces

$$\int |X_n - X_m|^p \, d\mu \geq \int_{A(\varepsilon,m,n)} |X_n - X_m|^p \geq \varepsilon^p \mu(A(\varepsilon, m, n)) \tag{5.5}$$

y combinando (5.4) y (5.5) obtenemos $\mu(A(\varepsilon, m, n)) < \varepsilon$ si $m, n \geq N(\varepsilon)$.

Sea ahora $A_k = A(\varepsilon 2^{-k}, N_{k+1}, N_k)$, $B_k = \cup_{i \geq k} A_i$, tenemos $\mu(A_k) < 2^{-k}\varepsilon$, $\mu(B_k) < 2^{1-k}\varepsilon$ y si $\omega \notin B_k$

$$|X_{N_{i+1}}(\omega) - X_{N_i}(\omega)| < \varepsilon 2^{-i} \quad \text{para } i \geq k.$$

Por lo tanto la serie $\sum_i (X_{N_{i+1}} - X_{N_i})$ converge fuera de $B = \cap_{k \geq 1} B_k$ y $\mu(B) = 0$. En consecuencia existe X tal que $X_{N_i} \to X$ c.s.

Para un entero fijo r ponemos $Y_i = |X_{N_i} - X_r|^p$, $Y = |X - X_r|^p$ y obtenemos una sucesión Y_i de funciones medibles no-negativas con $\liminf Y_i = \lim Y_i = Y$ c.s. Por el lema de Fatou tenemos

$$\int Y \, d\mu \leq \liminf_{i \to \infty} \int |X_{N_i} - X_r|^p \, d\mu < \varepsilon$$

si $r > N(\varepsilon)$. En consecuencia Y es integrable, es decir $(X - X_r) \in L^p$, lo cual implica que $X \in L^p$. Además probamos que

$$\int |X - X_r|^p \, d\mu < \varepsilon \quad \text{si } r > N(\varepsilon)$$

de modo que $X_r \xrightarrow{L^p} X$. ■

Observación 5.2 Es importante resaltar que el resultado anterior es falso para la integral de Riemann sobre intervalos finitos. No es difícil construir ejemplos de sucesiones de funciones cuyas potencias de orden p son integrables según Riemann, que son de Cauchy en L^p pero cuyo límite es discontinuo en un conjunto de medida positiva y por lo tanto no pueden ser integrables según Riemann.

Como consecuencia del teorema anterior observamos que los espacios L^p son espacios normados que son completos respecto a la métrica inducida por la norma. Un espacio con estas propiedades se conoce como un *espacio de Banach*.

Si $p > 1$ y $q > 1$ son conjugados decimos que los espacios L^p y L^q son conjugados. Por la desigualdad de Hölder sabemos que si $X \in L^p$ y $Y \in L^q$ entones el producto XY es integrable. Observamos que el conjugado de $p = 2$ es $q = 2$, es decir que L^2 es su propio espacio conjugado, y además es el único caso en el que esto ocurre. Esto quiere decir que si tomamos dos funciones de L^2, su producto es integrable, y por lo tanto podemos definir un producto interno en L^2: Para $X, Y \in L^2$,

$$\langle X, Y \rangle = \int XY \, d\mu.$$

Es fácil ver que esta definición satisface las condiciones de un producto interno (pasando al espacio cociente $\mathcal{L}^2$ si es necesario). Además, la norma asociada a este producto interno es la norma $||\cdot||_2$, que definimos anteriormente, ya que

$$\langle X, X \rangle = \int |X|^2 \, d\mu = ||X||_2^2.$$

Por lo tanto L^2 es un espacio de Banach que tiene un producto interno que induce la norma del espacio. Un espacio con estas propiedades se conoce como un *espacio de Hilbert*.

5.5. Convergencia en Distribución

Definición 5.7 Si F es una f.d. definimos el *conjunto de puntos de continuidad* o *conjunto de continuidad* de F por

$$\mathcal{C}(F) = \{x \in \mathbb{R} : F \text{ es continua en } x\}$$

Un intervalo finito I con extremos $a < b$ es un *intervalo de continuidad* para F si $a, b \in \mathcal{C}(F)$.

Definición 5.8 Sea $(X_n,\ n \geq 1)$ una sucesión de v.a. con f.d. $F_{X_n},\ n \geq 1$. X_n *converge en distribución* a la v.a. X cuando $n \to \infty$ si

$$F_{X_n}(x) \to F_X(x) \quad \text{cuando } n \to \infty,\ \forall x \in \mathcal{C}(F_X).$$

Notación: $X_n \xrightarrow{d} X$ o $X_n \xrightarrow{w} X$ cuando $n \to \infty$.

Observación 5.3 Como en esta definición sólo intervienen las funciones de distribuciones, las variables no necesariamente están definidas en un mismo espacio de probabilidad.

Abusando la notación, escribiremos $X_n \xrightarrow{w} \mathcal{N}(0,1)$ en lugar de $X_n \xrightarrow{w} X$ cuando $n \to \infty$ donde $X \sim \mathcal{N}(0,1)$.

Ejemplo 5.6
Sea $X_n \sim \delta_{1/n}$, es decir, la delta de Dirac concentrada en el punto $1/n$. Si la definición 5.8 tiene sentido deberíamos tener $X_n \xrightarrow{d} \delta_0$ cuando $n \to \infty$. Tenemos

$$F_{X_n}(x) = \begin{cases} 0, & \text{si } x < 1/n, \\ 1, & \text{si } x \geq 1/n, \end{cases} \quad \to \quad \begin{cases} 0, & \text{si } x \leq 0, \\ 1, & \text{si } x > 1/n. \end{cases}$$

Por lo tanto $F_{X_n} \to F_{\delta_0}(x)$ cuando $n \to \infty$ para todo $x \in \mathcal{C}(F_{\delta_0})$ pero no para todo x. Si en cambio tuviéramos $Y_n \sim \delta_{-1/n}$ entonces

$$F_{Y_n}(x) = \begin{cases} 0, & \text{si } x < -1/n, \\ 1, & \text{si } x \geq -1/n, \end{cases} \quad \to \quad \begin{cases} 0, & \text{si } x < 0, \\ 1, & \text{si } x \geq 1/n. \end{cases}$$

y en este caso si tenemos convergencia para todo x.

Teorema 5.6 *Sean X y $X_n,\ n \geq 1$ v.a. Si $X_n \xrightarrow{P} X$ entonces $X_n \xrightarrow{d} X$.*

Demostración. Sea $\varepsilon > 0$, entonces

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

y por la convergencia en probabilidad

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

De manera similar, cambiando X_n por X, x por $x - \varepsilon$, X por X_n y $x + \varepsilon$ por x, sigue que

$$\liminf_{n\to\infty} F_{X_n}(x) \geq F_X(x - \varepsilon).$$

Estas dos relaciones valen para todo x y todo $\varepsilon > 0$. Suponiendo ahora que $x \in \mathcal{C}(F_X)$ y haciendo $\varepsilon \to 0$ obtenemos que

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

■

Observación 5.4 Si F_X tiene un salto en x, sólo podemos concluir que

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

Como $F_X(x) - F_X(x^-)$ es el tamaño del salto no es posible obtener convergencia en un salto.

Ejemplo 5.7
Sea $N \sim \mathcal{N}(0,1)$, de modo que la distribución es simétrica. Definimos $X_n = (-1)^n N$ para $n \geq 1$, entonces $X_n \stackrel{d}{=} N$ de modo que $X_n \xrightarrow{d} N$, pero (X_n no converge ni c.s. ni en probabilidad.

Ejemplo 5.8
Para $\alpha > 0$ sea $X_n,\ n \geq 1$ una sucesión de v.a.i. tales que

$$P(X_n = 0) = 1 - \frac{1}{n^\alpha} \qquad \text{y} \qquad P(X_n = n) = \frac{1}{n^\alpha}$$

Los siguientes resultados son ciertos:

$$\begin{aligned} X_n &\xrightarrow{P} 0 \quad (n\to\infty) && \text{aún sin independencia} \\ X_n &\xrightarrow{c.s.} 0 \quad (n\to\infty) && \text{sii } \alpha > 1 \\ X_n &\xrightarrow{L^p} 0 \quad (n\to\infty) && \text{sii } \alpha > p \end{aligned}$$

La convergencia en probabilidad es consecuencia de:

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

La convergencia con probabilidad 1 sigue del Lema de Borel-Cantelli ya que

$$\sum_{n=1}^{\infty} P(|X_n| > \varepsilon) \begin{cases} < \infty & \text{cuando } \alpha > 1, \\ = \infty & \text{cuando } \alpha \leq 1, \end{cases}$$

En cuanto a la convergencia en L^p

$$\mathrm{E}\,|X_n|^p = 0^p \cdot \big(1 - \frac{1}{n^\alpha}\big) + n^p \cdot \frac{1}{n^\alpha} = n^{p-\alpha} \begin{cases} \to 0, & \text{para } p < \alpha, \\ = 1, & \text{para } p = \alpha, \quad \text{cuando } n \to \infty, \\ \to \infty, & \text{para } p > \alpha. \end{cases}$$

Observamos que $\mathrm{E}\,|X|^p$ ni converge a 0 ni diverge a infinito cuando $p = \alpha$, sino que es igual a 1.

Teorema 5.7 *Sea* $(X_n,\ n \geq 1)$ *una sucesión de v.a.i. y* c *una constante, entonces*

$$X_n \xrightarrow{d} \delta_c \quad (n\to\infty) \quad \Leftrightarrow \quad X_n \xrightarrow{P} c \quad (n \to \infty).$$

Demostración. Supongamos que $X_n \xrightarrow{d} \delta_c$ cuando $n \to \infty$ y sea $\varepsilon > 0$. Entonces

$$\begin{aligned} P(|X_n - c| > \varepsilon) &= 1 - P(c - \varepsilon \leq X_n \leq c + \varepsilon) \\ &= 1 - F_{X_n}(c+\varepsilon) + F_{X_n}(c-\varepsilon) - P(X_n = c - \varepsilon) \\ &\leq 1 - F_{X_n}(c+\varepsilon) + F_{X_n}(c-\varepsilon) \\ &\to 0 \quad \text{cuando } n \to \infty, \end{aligned}$$

ya que $F_{X_n}(c+\varepsilon) \to F_X(c+\varepsilon) = 1,\ F_{X_n}(c-\varepsilon) \to F_X(c-\varepsilon) = 0$, y $c-\varepsilon,\ c+\varepsilon \in \mathcal{C}(F_X)$. ■

5.5.1. Caracterización de la Convergencia en Distribución

Teorema 5.8 *Sea $\{X_n,\ n \geq 1\}$ una sucesión de v.a. y supongamos que $X_n \xrightarrow{d} X$ cuando $n \to \infty$. Si h es una función continua a valores reales o complejos definida en el intervalo acotado $[a, b]$, donde $a, b \in \mathcal{C}(F_X)$, entonces*

$$\operatorname{E} h(X_n) \to \operatorname{E} h(X) \quad \textit{cuando } n \to \infty. \tag{5.6}$$

Demostración. El caso complejo sigue del caso real. Usaremos el lema de aproximación. Sea $A \subset \mathcal{C}(F_X) \subset \mathbb{R}$ un subconjunto denso numerable. Si $h(x) = \mathbf{1}_{(c,d]}$ para $c, d \in A$, $a \leq c < d \leq b$, (5.6) se reduce a demostrar que $P(c < X_n \leq d) \to P(c < X \leq d)$, lo cual es cierto por hipótesis. Por linealidad la conclusión vale para funciones escalera cuyos 'escalones' tengan extremos en A. Sea ahora $h \in [a, b]$, y sea g una función simple aproximante, entonces

$$\begin{aligned}|\operatorname{E} h(X_n) - \operatorname{E}(h(X)| &\leq |\operatorname{E} h(X_n) - \operatorname{E} g(X_n)| + |\operatorname{E} g(X_n) - \operatorname{E} g(X)| + |\operatorname{E} g(X) - \operatorname{E} h(X)| \\ &\leq \operatorname{E}|h(X_n) - g(X_n)| + |\operatorname{E} g(X_n) - \operatorname{E} g(X)| + \operatorname{E}|g(X) - h(X)| \\ &\leq \varepsilon + |\operatorname{E} g(X_n) - \operatorname{E} g(X)| + \varepsilon.\end{aligned}$$

Como ya hemos visto que para funciones simples el término central tiende a cero cuando $n \to \infty$ obtenemos

$$\limsup_{n\to\infty} |\operatorname{E} h(X_n) - \operatorname{E} h(X)| < 2\varepsilon,$$

lo cual demuestra el teorema. ■

Teorema 5.9 *Sea $\{X_n,\ n \geq 1\}$ una sucesión de v.a. y supongamos que $X_n \xrightarrow{d} X$ cuando $n \to \infty$. Si h es una función a valores reales o complejos, continua y acotada, entonces*

$$\operatorname{E} h(X_n) \to \operatorname{E} h(X) \quad \textit{cuando } n \to \infty. \tag{5.7}$$

Demostración. De nuevo, el caso complejo sigue del caso real. Supongamos que $|h| \leq M$, entonces

$$\begin{aligned}|\operatorname{E} h(X_n) - \operatorname{E} h(X)| &\leq |\operatorname{E} h(X_n)\mathbf{1}_{\{|X_n|\leq K\}} - \operatorname{E} h(X)\mathbf{1}_{\{|X|\leq K\}}| + |\operatorname{E} h(X_n)\mathbf{1}_{\{|X_n|>K\}}| + |\operatorname{E} h(X)\mathbf{1}_{\{|X|>K\}}| \\ &\leq |\operatorname{E} h(X_n)\mathbf{1}_{\{|X_n|\leq K\}} - \operatorname{E} h(X)\mathbf{1}_{\{|X|\leq K\}}| + \operatorname{E}|h(X_n)|\mathbf{1}_{\{|X_n|>K\}} + \operatorname{E}|h(X)|\mathbf{1}_{\{|X|>K\}} \\ &\leq |\operatorname{E} h(X_n)\mathbf{1}_{\{|X_n|\leq K\}} - \operatorname{E} h(X)\mathbf{1}_{\{|X|\leq K\}}| + MP|X_n| > K) + MP(|X| > K)\end{aligned}$$

Sea $\varepsilon > 0$ y escojamos $K \in \mathcal{C}(F_X)$ suficientemente grande como para que $2MP(|X| > K) < \varepsilon$. Usando el teorema 5.8 para el primer término y la convergencia en distribución para los otros dos, obtenemos que

$$\limsup_{n\to\infty} |\operatorname{E} h(X_n) - \operatorname{E} h(X)| < 2MP(|X| > K) < \varepsilon.$$

■

Teorema 5.10 *Sea $\{X_n,\ n \geq 1\}$ una sucesión de v.a. Entonces $X_n \xrightarrow{d} X$ cuando $n \to \infty$ si y sólo si*

$$\operatorname{E} h(X_n) \to \operatorname{E} h(X) \quad \textit{cuando } n \to \infty. \tag{5.8}$$

para toda función real continua y acotada.

Demostración. Por el teorema 5.9 basta demostrar la suficiencia. Sean $a, b \in \mathcal{C}(F)$, $-\infty < a < b < \infty$. Ponemos

$$g_k(x) = \begin{cases} 0 & \text{para } x < a - \delta_k, \\ \frac{x-(a-\delta_k)}{\delta_k} & \text{para } x \in [a - \delta_k, a], \\ 1 & \text{para } x \in [a, b], \\ \frac{b+\delta_k-x}{\delta_k} & \text{para } x \in [b, b + \delta_k], \\ 0 & \text{para } x > b + \delta_k \end{cases}$$

con $\delta_k \downarrow 0$ cuando $k \to \infty$ y observamos que $\mathbf{1}_{(a,b]}(x) \leq g_k(x)$. Supongamos que (5.8) vale. Por la monotonía de la función de distribución y el teorema 5.9 obtenemos

$$F_n(a,b] = \int_a^b dF_n(x) \leq \mathrm{E}\, g_k(X_n) \to \mathrm{E}\, g_k(X) \quad \text{cuando } n \to \infty.$$

Haciendo $n \to \infty$

$$\limsup_{n\to\infty} F_n(a,b] \leq \mathrm{E}\, g_k(X) \leq F([a-\delta_k, b+\delta_k]).$$

Haciendo ahora $k \to \infty$ tenemos $\delta_k \to 0$ y $g_k(x) \downarrow \mathbf{1}_{[a,b]}$, y como a y b son puntos de continuidad concluimos que

$$\limsup_{n\to\infty} F_n(a,b] \leq F(a,b]. \tag{5.9}$$

Si, en cambio,

$$h_k(x) = \begin{cases} 0 & \text{para } x < a, \\ \frac{x-a}{\delta_k} & \text{para } x \in [a, a+\delta_k], \\ 1 & \text{para } x \in [a+\delta_k, b-\delta_k], \\ \frac{b-x}{\delta_k} & \text{para } x \in [b-\delta_k, b], \\ 0 & \text{para } x > b \end{cases}$$

los mismos argumentos nos dan

$$\liminf_{n\to\infty} F_n(a,b] \geq F([a+\delta_k, b-\delta_k]),$$

y como ahora $h_k(x) \uparrow \mathbf{1}_{(a,b)}$, tenemos finalmente

$$\liminf_{n\to\infty} F_n(a,b] \geq F(a,b]. \tag{5.10}$$

Las ecuaciones (5.9) y (5.10) demuestran que $X_n \xrightarrow{d} X$ cuando $n \to \infty$. ■

Como corolario obtenemos los siguientes resultados:

Corolario 5.4 *Sea $\{F_n,\ n \geq 0\}$ una familia de funciones de distribución. Las siguientes afirmaciones son equivalentes*

(1) $F_n \xrightarrow{d} F_0$.

(2) Para toda función real f acotada y uniformemente continua,

$$\int f\, dF_n \to \int f\, dF_0,$$

cuando $n \to \infty$.

Demostración. Basta observar que las funciones g_k y h_k que usamos en la demostración del teorema anterior son continuas con soporte compacto y por lo tanto son uniformemente continuas. ■

Teorema 5.11 *Sean X e Y v.a.*

$$X \stackrel{d}{=} Y \quad \Leftrightarrow \quad \mathrm{E}\, h(X) = \mathrm{E}\, h(Y)$$

para toda función h continua y acotada.

Teorema 5.12 *Sean X y $\{X_n,\ n \geq 1\}$ v.a. y supongamos que $X_n \xrightarrow{d} X$ cuando $n \to \infty$. Entonces*

$$\mathrm{E}\,|X| \leq \liminf_{n\to\infty} \mathrm{E}\,|X_n|.$$

Demostración. Sea $K \in \mathcal{C}(F_X)$ un número positivo. Por el teorema 5.8 tenemos

$$\liminf_{n\to\infty} \mathrm{E}\,|X_n| \geq \liminf_{n\to\infty} \mathrm{E}\,|X_n|\mathbf{1}_{\{|X_n|\leq K\}} = \mathrm{E}\,|X|\mathbf{1}_{\{|X|\leq K\}}.$$

La conclusión sigue haciendo K tender a infinito a través de una sucesión de puntos de continuidad de F_X. ■