Capítulo 1

Sucesiones de Variables Aleatorias

1.1. 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 1.1 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 1.2 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 1.1

- 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.

1.2. Convergencia Casi Segura

Definición 1.3 Sean $X_n,\ n \geq 1$ y X como en la definición 1.1. 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 1.1
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 que no es vacío.

Proposición 1.1 *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

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

de modo que F_1 es medible. ■

Definición 1.4 Sea $X_n,\ n \geq 1$ como en la definición 1.1. 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 1.2 *Una sucesión de funciones medibles (X_n) converge c.s. si y sólo si es de Cauchy c.s.*

La demostración queda como ejercicio.

1.3. Convergencia en Medida

Definición 1.5 Sean $X_n,\ n \geq 1$ y X como en la definición 1.1. 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 1.3 *Sean $X_n,\ n \geq 1$ y X como en la definición 1.1 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 existe $\varepsilon > 0$

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

■

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

Ejemplo 1.2
Sea $(\Omega, \mathcal{F}, \mu) = ([0,1], \mathcal{B}, \lambda)$ el espacio de Lebesgue y definimos la siguiente sucesión de variables aleatorias:

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

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 1.3
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.

1.3.1. 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* $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ 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.

1.3.2. 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 1.1 (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 en medida 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 1.1 *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

$$\mu(|X_{n_k} - X| > \varepsilon) \geq \delta. \tag{1.1}$$

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 (1.1). ∎

Corolario 1.2 *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 $N \in \mathcal{F}$ 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 1.3 (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. Como las integrales forman una sucesión de números reales, 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 medida 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 1.4 (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) 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 $(X_n Y_n)$ tiene una subsucesión que converge c.s. ∎

Proposición 1.5 (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. ∎

1.4. 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 1.6 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 1.6 *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 1.4
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 1.2. 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 1.3. Aún cuando la medida del espacio sea finita, es posible dar ejemplo de sucesiones que convergen c.s. pero no convergen en L^p.

1.4.1. Desigualdades

Teorema 1.2 (Desigualdad de Hölder) *Sea p, q números reales tales que $p, q > 1$ y*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{1.2}$$

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

$$\int |XY| \, d\mu \leq \Big(\int |X|^p \, d\mu \Big)^{1/p} \Big(\int |Y|^q \, d\mu \Big)^{1/q}. \tag{1.3}$$

En términos de las normas la desigualdad dice que

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

Si p y q satisfacen (1.2) 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 (1.3) 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 1.3 (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. Si $p = 1$ 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 1.4 (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 $(\mathrm{E}(X), g(\mathrm{E}(X)))$, $a\,\mathrm{E}(X) + b = g(\mathrm{E}(X))$. ■

Ejemplo 1.5 (Desigualdad de Lyapunov)
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 $r < p$, $X_n \xrightarrow{L^r} X$.

Antes de presentar el siguiente resultado recordamos el lema de Fatou, que usaremos en la próxima demostración.

Lema 1.1 (Lema de Fatou) *Si $X_n \geq 0$ para $n \geq 1$ entonces*

$$\mathrm{E}(\liminf_n X_n) \leq \liminf_n \mathrm{E}(X_n),$$

y en particular si $X_n \to X$ c.p.1,

$$\mathrm{E}(X) \leq \liminf_n \mathrm{E}(X_n).$$

Teorema 1.5 *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. Sea (X_n) una sucesión de Cauchy en L^p. Dado $\varepsilon > 0$ sea $N = N(\varepsilon) \in \mathbb{N}$ tal que si $n, m \geq N$,

$$\int |X_n - X_m|^p \, d\mu < \varepsilon^{p+1}. \tag{1.4}$$

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

$$\int |X_n - X_m|^p \, d\mu \geq \int_{A(\varepsilon,m,n)} |X_n - X_m|^p \geq \varepsilon^p \mu(A(\varepsilon, m, n)) \tag{1.5}$$

y combinando (1.4) y (1.5) 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 $\omega \notin B_k$

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

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 1.2 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*.

1.5. Convergencia en Distribución

Definición 1.7 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 1.8 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 1.3 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 1.6
Sea $X_n \sim \delta_{1/n}$, es decir, la delta de Dirac concentrada en el punto $1/n$. Si la definición 1.8 tiene sentido deberíamos tener $X_n \xrightarrow{d} \delta_0$ cuando $n \to \infty$. Tenemos

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

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 1.6 *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 1.4 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 1.7
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 1.8
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 1.7 *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 $\varepsilon > 0$. Si $c - \varepsilon,\ c + \varepsilon \in \mathcal{C}(F_X)$ 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 $F_{X_n}(c+\varepsilon) \to F_X(c+\varepsilon) = 1,\ F_{X_n}(c-\varepsilon) \to F_X(c-\varepsilon) = 0,.$ ∎

1.5.1. Caracterización de la Convergencia en Distribución

Lema 1.2 (Lema de Aproximación) *Sea f una función a valores reales tal que se cumple alguna de las siguientes condiciones:*

- $f \in \mathcal{C}_0$ *(funciones continuas que se anulan en $\pm\infty$),*
- $f \in \mathcal{C}[a,b]$ *para $a, b \in \mathbb{R}$,*

entonces, para cada $\varepsilon > 0$ existe una función simple g tal que

$$\sup_x |f(x) - g(x)| < \varepsilon.$$

Teorema 1.8 *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*

$$\operatorname{E} h(X_n) \to \operatorname{E} h(X) \quad \text{cuando } n \to \infty. \tag{1.6}$$

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 $c, d \in A$, $a \leq c < d \leq b$, (1.6) 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 $h \in \mathcal{C}[a,b]$, y sea g una función simple aproximante, entonces usando la desigualdad de Jensen,

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

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 1.9 *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*

$$\operatorname{E} h(X_n) \to \operatorname{E} h(X) \quad \text{cuando } n \to \infty. \tag{1.7}$$

Demostración. De nuevo, el caso complejo sigue del caso real. Supongamos que $|h(x)| \leq M$ para todo x y sea $K > 0$, 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 1.8 para el primer término y la convergencia en distribución para el segundo, obtenemos que

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

∎

Teorema 1.10 *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*

$$\mathrm{E}\, h(X_n) \to \mathrm{E}\, h(X) \quad \textit{cuando } n \to \infty. \tag{1.8}$$

para toda función real continua y acotada.

Demostración. Por el teorema 1.9 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 (1.8) vale. Por la monotonía de la función de distribución y el teorema 1.9 obtenemos

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

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

$$\limsup_{n\to\infty} F_n(a,b] \leq F(a,b]. \tag{1.9}$$

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

$$\liminf_{n\to\infty} F_n(a,b] \geq F(a,b]. \tag{1.10}$$

Las ecuaciones (1.9) y (1.10) demuestran que $X_n \xrightarrow{d} X$ cuando $n \to \infty$. ∎

Como corolario obtenemos los siguientes resultados:

Corolario 1.4 *Sea $\{F_n,\ n \geq 0\}$ una familia de funciones de distribución y F también una f.d. 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 1.11 *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 1.12 *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 1.8 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. ■

1.6. Integrabilidad Uniforme

Definición 1.9 Una sucesión de v.a. $(X_n,\ n \geq 1)$ es *uniformemente integrable* si

$$\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) \to 0 \quad \text{cuando} \quad a \to \infty,$$

uniformemente en n.

Notación: Abreviaremos 'uniformemente integrable' por u.i.

Una manera equivalente de expresar la integrabilidad uniforme es a través de las funciones de distribución: $\{X_n,\ n \geq 1\}$ es uniformemente integrable sii

$$\int_{|x|>a} |x| dF_{X_n}(x) \to 0 \quad \text{cuando} \quad a \to \infty,$$

uniformemente en n.

Observación 1.5 Si las variables X_n tienen media finita entonces $\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) \to 0$ cuando $a \to \infty$, para cada n, ya que las colas de integrales convergentes tienden a cero. La integrabilidad uniforme quiere decir que las colas tienden a cero de manera uniforme para todos los miembros de la sucesión.

Para cualquier sucesión de v.a.u.i.,

$$\mathrm{E}\,|X_n| = \mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|\leq a\}}) + \mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) \leq a + 1,$$

para a suficientemente grande. Por lo tanto los momentos están acotados uniformemente. Sin embargo, la integrabilidad uniforme es más fuerte, como lo muestra el siguiente teorema.

Teorema 1.13 *Las variables $X_n,\ n \geq 1$ son u.i. si y sólo si*

(a) $\sup_n \mathrm{E}\,|X_n| < \infty$,

(b) Para cualquier $\varepsilon > 0$ existe $\delta > 0$ tal que para cualquier conjunto A con $P(A) < \delta$,

$$\mathrm{E}(|X_n|\mathbf{1}_A) < \varepsilon \quad \textit{uniformemente en } n.$$

Demostración. Supongamos que $\{X_n,\ n \geq 1\}$ es u.i., sea A un conjunto tal que $P(A) < \delta$, donde δ será escogido más adelante. Ya hemos verificado que los momentos son uniformemente acotados. Veamos (b):

$$\begin{aligned}\mathrm{E}\,|X_n|\mathbf{1}_A &= \mathrm{E}(|X_n|\mathbf{1}_{A\cap\{|X_n|\leq a\}}) + \mathrm{E}(|X_n|\mathbf{1}_{A\cap\{|X_n|>a\}}) \\ &\leq aP(A) + \mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) \leq a\delta + \varepsilon/2 < \varepsilon,\end{aligned}$$

si escogemos primero a suficientemente grande como para que el segundo sumando sea menor que $\varepsilon/2$ y luego δ suficientemente pequeño como para que $a\delta < \varepsilon/2$.

Por otro lado, si se satisfacen las condiciones del teorema ponemos $A_n = \{|X_n| > a\}$ y aplicamos la desigualdad de Markov y (a) para obtener

$$P(A_n) \leq \frac{\mathrm{E}\,|X_n|}{a} \leq \frac{\sup_n \mathrm{E}\,|X_n|}{a} < \delta$$

uniformemente en n para a suficientemente grande, lo cual, por (b), muestra que

$$\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) = \mathrm{E}(|X_n|\mathbf{1}_{A_n}) < \varepsilon$$

uniformemente en n, lo que demuestra la i.u. ■

A continuación presentamos algunos criterios suficientes para i.u.

Teorema 1.14 *Sea $\{X_n,\ n \geq 1\}$ v.a. y supongamos que*

$$\sup_n \mathrm{E}(|X_n|^p) < \infty \quad \textit{para algún } p > 1.$$

entonces $\{X_n,\ n \geq 1\}$ es u.i. En particular esto es cierto si $\{|X_n|^p,\ n \geq 1\}$ es u.i. para algún $p > 1$.

Demostración. Tenemos

$$\begin{aligned}\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) &\leq a^{1-p}\,\mathrm{E}(|X_n|^p\mathbf{1}_{\{|X_n|>a\}}) \leq a^{1-p}\,\mathrm{E}(|X_n|^p) \\ &\leq a^{1-p}\sup_n \mathrm{E}(|X_n|^p) \to 0 \quad \text{cuando } a \to \infty,\end{aligned}$$

uniformemente en n. ■

La siguiente generalización se demuestra de manera similar.

Teorema 1.15 *Sea $\{X_n,\ n \geq 1\}$ v.a. y sea g una función no-negativa y creciente tal que $g(x)/x \to \infty$ cuando $x \to \infty$. Si*

$$\sup_n \mathrm{E}(g(X_n)) < \infty,$$

entonces $\{X_n,\ n \geq 1\}$ es u.i.

Teorema 1.16 *Sea $\{X_n,\ n \geq 1\}$ v.a. tales que*

$$|X_n| \leq Y \ \textit{c.s.} \quad \textit{para todo } n,$$

donde Y es una v.a. positiva integrable. Entonces $\{X_n,\ n \geq 1\}$ es u.i.

Demostración. Basta observar que

$$\mathrm{E}(|X_n|\mathbf{1}_{\{|X_n|>a\}}) \leq \mathrm{E}(|Y|\mathbf{1}_{\{Y>a\}}) \to 0 \quad (a \to \infty),$$

uniformemente en n. ∎

Corolario 1.5 *Sea $\{X_n,\ n \geq 1\}$ una sucesión de v.a. Si*

$$\mathrm{E}(\sup_n |X_n|) < \infty,$$

entonces $\{X_n,\ n \geq 1\}$ es u.i.

Ejemplo 1.9
Sea $X_i \in L^1$ para $i = 1, \ldots, n$ entonces el conjunto de v.a. $\{X_1, \ldots, X_n\}$ es uniformemente integrable ya que tenemos la siguiente acotación:

$$|X_i| \leq \sum_{i=1}^{n} |X_i| \in L^1$$

y por el teorema 1.16 el conjunto es u.i.

Ejemplo 1.10
Supongamos que $X_t \in L^1,\ X_t \in L^1$ y $|X_t| \leq |Y_t|$ para todo $t \in T$. Si el conjunto $\{Y_t\}$ es u.i. entonces $\{X_t\}$ también lo es. Esto sigue fácilmente de la definición.

1.6.1. Convergencia de Momentos

Lema 1.3

$$X_n \xrightarrow{L^1} X \quad \Leftrightarrow \quad \sup_{A \in \mathcal{F}} \Big| \int_A X_n dP - \int_A X dP \Big| \to 0.$$

Demostración. Si $X_n \xrightarrow{L^1} X$ entonces

$$\begin{aligned}\sup_A \Big| \int_A X_n dP - \int_A X dP \Big| &= \sup_A \Big| \int_A (X_n - X) dP \Big| \leq \sup_A \int_A |X_n - X| dP \\ &\leq \int |X_n - X| \, dP = \mathrm{E}(|X_n - X|) \to 0.\end{aligned}$$

Veamos ahora el recíproco, tenemos

$$\begin{aligned}\mathrm{E}\,|X_n - X| &= \int_{\{X_n > X\}} (X_n - X) \, dP + \int_{\{X_n \leq X\}} (X - X_n) \, dP \\ &= \int_{\{X_n > X\}} X_n \, dP - \int_{\{X_n > X\}} X \, dP + \int_{\{X_n \leq X\}} X \, dP - \int_{\{X_n \leq X\}} X_n \, dP \\ &\leq 2 \sup_A \Big| \int_A X_n \, dP - \int_A X \, dP \Big|.\end{aligned}$$

∎

Lema 1.4 *Si $X_n \xrightarrow{L^p} X$ para $p \geq 1$ entonces $\mathrm{E}(|X_n|^p) \to \mathrm{E}(|X|^p)$ o equivalentemente $||X_n||_p \to ||X||_p$.*

Demostración. $X = X_n + X - X_n$. Por la desigualdad de Minkowski

$$||X||_p \leq ||X_n||_p + ||X - X_n||_p.$$

De manera similar

$$||X_n||_p \leq ||X||_p + ||X - X_n||_p.$$

Combinando obtenemos

$$\Big| ||X_n||_p - ||X||_p \Big| \leq ||X - X_n||_p \to 0. \tag{1.11}$$

■

Teorema 1.17 *Supongamos que $X_n \in L^1$ para $n \geq 1$. Las siguientes afirmaciones son equivalentes*

a) $\{X_n\}$ converge en L^1.

b) $\{X_n\}$ es de Cauchy en L^1: $\mathrm{E}\,|X_n - X_m| \to 0$ cuando $n, m \to \infty$.

c) $\{X_n\}$ es u.i. y converge en probabilidad.

Demostración. (a) $\Rightarrow$ (b) por la desigualdad triangular.

(b) $\Rightarrow$ (c) Veamos primero que $\{X_n\}$ es u.i. Dado $\varepsilon > 0$ existe N_ε tal que si $m, n \geq N_\varepsilon$ entonces

$$\int |X_n - X_m|\, dP < \frac{\varepsilon}{2}.$$

Para ver que $\{X_n\}$ es u.i. usaremos la caracterización del teorema 1.13. Para cualquier $A \in \mathcal{F}$,

$$\int_A |X_n|\, dP \leq \int_A |X_n - X_{N_\varepsilon} + X_{N_\varepsilon}|\, dP \leq \int_A |X_{N_\varepsilon}|\, dP + \int_A |X_n - X_{N_\varepsilon}|\, dP.$$

Si $n \geq N_\varepsilon$,

$$\int_A |X_n|\, dP \leq \int_A |X_{N_\varepsilon}|\, dP + \frac{\varepsilon}{2}$$

de modo que

$$\sup_{n \geq N_\varepsilon} \int_A |X_n|\, dP \leq \int_A |X_{N_\varepsilon}|\, dP + \frac{\varepsilon}{2}$$

lo que implica que

$$\sup_n \int_A |X_n|\, dP \leq \sup_{m \leq N_\varepsilon} \int_A |X_m|\, dP + \frac{\varepsilon}{2}$$

Si $A = \Omega$ concluimos que

$$\sup_n \mathrm{E}\,|X_n| \leq \sup_{m \leq N_\varepsilon} \mathrm{E}\,|X_m| + \frac{\varepsilon}{2} < \infty.$$

Además, como familias finitas de variables en L^1 son u.i., $\{X_m, m \leq N_\varepsilon\}$ es u.i. y dado $\varepsilon > 0$ existe δ tal que si $P(A) < \delta$,

$$\sup_{m \leq N_\varepsilon} \int_A |X_m|\, dP < \frac{\varepsilon}{2},$$

de modo que si $P(A) < \delta$ tenemos que

$$\sup_n \int_A |X_n|\, dP \leq \varepsilon.$$

Falta ver que (X_n) converge en probabilidad. Tenemos

$$P(|X_n - X_m| > \varepsilon) \leq \frac{1}{\varepsilon}\,\mathrm{E}(|X_n - X_m|) \to 0$$

cuando $n, m \to \infty$. Por lo tanto (X_n) es de Cauchy en probabilidad y por lo tanto converge en probabilidad.

(c) $\Rightarrow$ (a) Si $X_n \to X$ en probabilidad, existe una subsucesión X_{n_k} tal que $X_{n_k} \to X$ c.s.. Por el lema de Fatou

$$\begin{aligned} \mathrm{E}(|X|) &= \mathrm{E}(\liminf |X_{n_k}|) \le \liminf \mathrm{E}(|X_{n_k}|) \\ &\le \sup_n \mathrm{E}(|X_n|) < \infty \end{aligned}$$

ya que (X_n) es u.i.. Por lo tanto $X \in L^1$.

Dado $\varepsilon > 0$ tenemos

$$\begin{aligned} \int |X_n - X|\, dP &\le \int_{\{|X_n - X| \le \varepsilon\}} |X_n - X|\, dP + \int_{\{|X_n - X| > \varepsilon\}} |X_n|\, dP + \int_{\{|X_n - X| > \varepsilon\}} |X|\, dP \\ &\le \varepsilon + T_1 + T_2 \end{aligned}$$

Como $X_n \xrightarrow{P} X$, $P(|X_n - X| > \varepsilon) \to 0$ y por lo tanto $T_2 \to 0$ cuando $n \to \infty$. Para ver que $T_1 \to 0$, como (X_n) es u.i., dado $\varepsilon > 0$ existe δ tal que si $P(A) < \delta$,

$$\sup_{n \ge 1} \int_A |X_n|\, dP < \varepsilon$$

Basta escoger n grande de modo que $P(|X_n - X| > \varepsilon) < \delta$ y se tiene que $T_1 < \varepsilon$. ■

1.7. El Lema de Scheffé

Hemos visto en secciones anterior la relación entre convergencia de momentos y distintos tipos de convergencia. La pregunta que nos hacemos ahora es cual es la relación entre convergencia débil y convergencia de densidades.

Ejemplo 1.11
Sea $X_n,\ n \ge 1$ una sucesión de v.a. con densidades

$$f_{X_n}(x) = \begin{cases} 1 - \cos(2\pi n x), & \text{para } 0 \le x \le 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Entonces

$$F_{X_n}(x) = \begin{cases} 0, & \text{para } x \le 0, \\ x - \frac{\operatorname{sen}(2\pi n x)}{2\pi n} \to x, & \text{para } 0 \le x \le 1, \\ 1, & \text{para } x \ge 1. \end{cases}$$

lo cual implica que $X_n \xrightarrow{d} \mathcal{U}(0,1)$ cuando $n \to \infty$. Sin embargo, las densidades oscilan y por lo tanto no pueden converger.

Definición 1.10 La *distancia en variación total* entre las f.d. F y G se define como

$$d(F, G) = \sup_{A \in \mathcal{B}} |F(A) - G(A)|.$$

Si las variables X e Y tienen f.d. F y G respectivamente, la definición es equivalente a

$$d(X, Y) = \sup_{A \in \mathcal{B}} |P(X \in A) - P(Y \in A)|.$$

Si X y X_n, $n \geq 1$ son v.a. tales que

$$d(X_n, X) \to 0 \quad (n \to \infty),$$

decimos que X_n *converge a* X *en variación total* cuando $n \to \infty$.

Como los conjuntos $(-\infty, x]$ son de Borel para todo x, tenemos que

$$|P(X_n \leq x) - P(X \leq x)| \leq \sup_{A \in \mathcal{B}} |P(X_n \in A) - P(X \in A)|,$$

lo cual implica el siguiente resultado.

Lema 1.5 *Sea* $\{X_n,\ n \geq 1\}$ *una sucesión de v.a. Si* $X_n \to X$ *en variación total cuando* $n \to \infty$ *entonces* $X_n \xrightarrow{d} X$ *cuando* $n \to \infty$.

Teorema 1.18 (Lema de Scheffé) *Supongamos que* X *y* X_n, $n \geq 1$ *son v.a. absolutamente continuas. Entonces*

$$\sup_{A \in \mathcal{B}} |P(X_n \in A) - P(X \in A)| \leq \int_{\mathbb{R}} |f_{X_n}(x) - f_X(x)|\, dx, \tag{1.12}$$

y si $f_{X_n}(x) \to f_X(x)$ *c.s. cuando* $n \to \infty$, *entonces* $d(X_n, X) \to 0$ *cuando* $n \to \infty$. *En particular,* $X_n \xrightarrow{d} X$ *cuando* $n \to \infty$.

Demostración. Sea $A \in \mathcal{B}$, tenemos

$$\begin{aligned} |P(X_n \in A) - P(X \in A)| &= \Big| \int_A f_{X_n}(x) - \int_A f_X(x)\, dx \Big| \\ &\leq \int_A |f_{X_n}(x) - f_X(x)|\, dx \\ &\leq \int_{\mathbb{R}} |f_{X_n}(x) - f_X(x)|\, dx, \end{aligned}$$

lo cual demuestra (1.12).

Para demostrar la convergencia observamos que

$$\int_{\mathbb{R}} (f_{X_n}(x) - f_X(x))\, dx = 1 - 1 = 0,$$

de modo que

$$\int_{\mathbb{R}} (f_{X_n}(x) - f_X(x))^+\, dx = \int_{\mathbb{R}} (f_{X_n}(x) - f_X(x))^-\, dx.$$

Además, $(f_X(x) - f_{X_n}(x))^+ \to 0$ c.s. cuando $n \to \infty$, y $0 \leq (f_X(x) - f_{X_n}(x))^+ \leq f_X(x)$, y usando el teorema de convergencia dominada

$$\int_{\mathbb{R}} |f_X(x) - f_{X_n}(x)| dx = 2 \int_{\mathbb{R}} (f_X(x) - f_{X_n}(x))^+ dx \to 0 \quad (n \to 0).$$

Esto demuestra la convergencia en variación total. ■

Observación 1.6 Es posible demostrar que

$$d(X_n, X) = \frac{1}{2} \int_{\mathbb{R}} |f_{X_n}(x) - f_X(x)|\, dx.$$

1.8. Convergencia Vaga

En nuestra definición de convergencia en distribución hemos supuesto que la función de distribución límite F es propia, es decir que $F(\mathbb{R}) = 1$ pero este no es siempre el caso. Esta es la distinción fundamental entre la convergencia en distribución y un nuevo tipo de convergencia que definimos a continuación.

Definición 1.11 Una sucesión $\{F_n,\ n \geq 1\}$ de f.d. converge vagamente a la f.d (posiblemente impropia) H si, para todo intervalo finito $I = (a, b]$ con $a, b \in \mathcal{C}(H)$,

$$F_n(I) \to H(I) \quad (n \to \infty)$$

Notación: $F \xrightarrow{v} H$ cuando $n \to \infty$.

Ejemplo 1.12
Sea $P(X_n = n) = P(X_n = -n) = 1/2$, entonces

$$F_n(x) = \begin{cases} 0, & \text{para } x < -n, \\ \frac{1}{2}, & \text{para } -n \leq x < n, \\ 1, & \text{para } x \geq n. \end{cases}$$

De modo que para todo $x \in \mathbb{R}$

$$F_n(x) \to H(x) = \frac{1}{2} \quad (n \to \infty).$$

H tiene todas las propiedades de una función de distribución salvo que $F(\mathbb{R}) = 0$.

Ejemplo 1.13
Sea $X_n \sim \mathcal{U}(-n, n),\ n \geq 1$. La función de distribución es

$$F_{X_n}(x) = \begin{cases} 0, & \text{para } x < -n, \\ \frac{x+n}{2n}, & \text{para } -n \leq x < n, \\ 1, & \text{para } x \geq n. \end{cases}$$

De nuevo, esta sucesión converge cuando $n \to \infty$ a $H(x) = 1/2$ para todo $x \in \mathbb{R}$.

Si $X_n \xrightarrow{d} X$ cuando $n \to \infty$ entonces, por definición, $F_{X_n}(x) \to F_X(x)$ cuando $n \to \infty$ para $x \in \mathcal{C}(F_X)$, de donde se obtiene de inmediato que para cualquier intervalo $I = (a, b]$ acotado tal que $a, b \in \mathcal{C}(F_X)$,

$$F_{X_n}(I) = F_{X_n}(b) - F_{X_n}(a) \to F_X(b) - F_X(a) = F_X(I)$$

cuando $n \to \infty$, de modo que convergencia en distribución siempre implica convergencia vaga.

Teorema 1.19 (Principio de Selección de Helly) *Sea $\{F_n,\ n \geq 1\}$ una sucesión de f.d. Entonces existe una subsucesión no-decreciente $\{n_k,\ k \geq 1\}$ tal que*

$$F_{n_k} \xrightarrow{v} H \quad \textit{cuando } k \to \infty,$$

para alguna función de distribución H, posiblemente impropia.

Demostración. La demostración usa el método diagonal. Sea $\{r_k,\ k \geq 1\}$ una enumeración de los racionales. Como la sucesión $(F_n(r_1))_{n \geq 1}$ está acotada, por el teorema de Bolzano-Weierstrass existe un punto de acumulación j_1 y una subsucesión $\{F_{1,k}(r_1),\ k \geq 1\}$ tal que

$$F_{1,k}(r_1) \to j_1 \quad \text{cuando } k \to \infty,$$

donde $0 \leq j_1 \leq 1$. El mismo argumento aplicado a la subsucesión $\{F_{2,k}(r_2)\}$ produce un punto de acumulación j_2 y una subsubsucesión $\{F_{2,k}(r_2), k \geq 1\}$ tal que

$$F_{2,k}(r_2) \to j_2 \quad \text{cuando } k \to \infty,$$

donde $0 \leq j_2 \leq 1$. Además, esta subsucesión converge también para r_1, ya que es una subsucesión de la primera subsucesión. Continuando este procedimiento obtenemos, cuando $k \to \infty$,

$$\begin{aligned} F_{1,k}(r_1) &\to j_1, \\ F_{2,k}(r_i) &\to j_i, \quad \text{para } i = 1, 2, \\ F_{3,k}(r_i) &\to j_i, \quad \text{para } i = 1, 2, 3 \\ &\vdots \qquad\quad \vdots \\ F_{m,k}(r_i) &\to j_i, \quad \text{para } i = 1, 2, \dots, m \\ &\vdots \qquad\quad \vdots \end{aligned}$$

Ahora consideramos la diagonal de este arreglo, $\{F_{k,k},\ k \geq 1\}$ que, salvo por un número finito de términos al comienzo, es una subsucesión de todas las subsucesiones horizontales, y por lo tanto converge en todo $\mathbb{Q}$. Tenemos, por lo tanto, un conjunto denso numerable en el cual la sucesión converge.

Definimos ahora la función H: Ponemos

$$H(r_i) = j_i, \quad \text{para } r_i \in \mathbb{Q}.$$

Como $0 \leq j_i \leq 1$ para todo i tenemos que $0 \leq H(r_i) \leq 1$ para $r_i \in \mathbb{Q}$. Además, si $r < s$, $r, s \in \mathbb{Q}$, entonces

$$0 \leq F_{k,k}(s) - F_{k,k}(r) \to H(s) - H(r)$$

cuando $k \to \infty$, lo cual muestra que H es no-decreciente en $\mathbb{Q}$.

Sea ahora $x \in \mathbb{R}$ arbitrario y escogemos $r, s \in \mathbb{Q}$ tal que $r < x < s$. Como $F_{k,k}(r) \leq F_{k,k}(x) \leq F_{k,k}(s)$, tomando límites obtenemos que

$$H(r) \leq \liminf_{k\to\infty} F_{k,k}(x) \leq \limsup_{k\to\infty} F_{k,k}(x) \leq H(s),$$

y como podemos escoger r y s tan cerca como queramos de x,

$$H(x^-) \leq \liminf_{k\to\infty} F_{k,k}(x) \leq \limsup_{k\to\infty} F_{k,k}(x) \leq H(x^+)$$

Esto prueba que siempre existe una subsucesión que converge a una función no-decreciente H que toma valores en $[0, 1]$ en sus puntos de continuidad. Para que H sea continua por la derecha, completamos su definición poniendo

$$H(x) = \lim_{r_i \downarrow x} H(r_i) \quad \text{para } x \in \mathbb{R},$$

es decir, escogemos la versión de H que es continua por la derecha. ■

Observación 1.7 Es necesario notar que el teorema de Helly no dice que $F_{k,k}(\infty) \to H(\infty)$.

Si llamamos $\mathcal{M}$ a la clase de las funciones de distribución posiblemente degeneradas, es decir, a las funciones $H : \mathbb{R} \to \mathbb{R}$ que son no-decrecientes, continuas por la derecha y satisfacen $G(-\infty) \geq 0$, $G(\infty) \leq 1$, entonces el teorema de Helly dice que $\mathcal{M}$ es *relativamente o secuencialmente compacto*. Ya hemos visto en los ejemplos 1.12 y 1.13 que la clase de las funciones de distribución no es compacta.

Corolario 1.6 *Sea F_n una sucesión en $\mathcal{M}$. Si existe $H \in \mathcal{M}$ tal que para toda subsucesión vagamente convergente (F_{n_k}) se tiene que $F_{n_k} \xrightarrow{v} H$, entonces la sucesión también converge vagamente a H: $F_n \xrightarrow{v} H$.*

Demostración. Supongamos que F_n no converge vagamente a H, entonces existe $x_0 \in \mathcal{C}(H)$ tal que $F_n(x_0) \nrightarrow H(x_0)$. Pero toda subsucesión de las F_n tiene una subsucesión convergente F_{n_k} y $F_{n_k}(x_0) \to H(x_0)$. Esto implica que $F_n(x_0) \to H(x_0)$, lo cual es una contradicción. ■

Observación 1.8 Denotamos por $\mathcal{C}_0(\mathbb{R})$ a la clase de funciones reales continuas tales que $|h(x)| \to 0$ cuando $x \to \pm\infty$. Llamamos $\mathcal{C}_K(\mathbb{R})$ a la clase de funciones continuas a soporte compacto ($f \in \mathcal{C}_K$ si existe un conjunto compacto $K(f)$ tal que f se anula fuera de $K(f)$). Finalmente, llamamos $\mathcal{C}_B(\mathbb{R})$ a las funciones continuas y acotadas. Tenemos

$$\mathcal{C}_K \subset \mathcal{C}_0 \subset \mathcal{C}_B \subset \mathcal{C}.$$

Hemos visto en el teorema 1.10 que las integrales de funciones en $\mathcal{C}_B$ caracterizan la convergencia en distribución: $F_n \xrightarrow{d} F$ si y sólo si $\int f\,dF_n \to \int f\,dF$ para toda $f \in \mathcal{C}_B(\mathbb{R})$. El siguiente teorema nos da una caracterización similar para la convergencia vaga. Primero unas definiciones y un lema de aproximación.

Una función cualquiera en un espacio arbitrario tiene *soporte* en un subconjunto S del espacio si f se anula fuera del conjunto S. Por lo tanto, si $f \in \mathcal{C}_K(\mathbb{R})$, f tiene soporte en algún conjunto compacto y por lo tanto también en un intervalo compacto. Una *función escalera* en el intervalo (finito o infinito) (a, b) es una función con soporte en este intervalo tal que para alguna partición finita $a = a_1 < \cdots < a_k = b$ se tiene que $f(x) = c_j$ para $x \in (a_j, a_{j+1})$ y $1 \le j \le k$, donde los c_j son números reales. Decimos que la función es a valores en D si $a_j \in D$, $c_j \in D$ para $1 \le j \le k$. Observamos que hemos dejado sin especificar los valores de la función f en los puntos de la partición para tener flexibilidad.

Lema 1.6 (de Aproximación) *Supongamos que $f \in \mathcal{C}_K(\mathbb{R})$ tiene soporte en el intervalo compacto $[a, b]$. Dado cualquier conjunto denso A de $\mathbb{R}$ y $\varepsilon > 0$ existe una función escalera en (a, b) definida usando una partición con valores en A tal que*

$$\sup_{s \in \mathbb{R}} |f(x) - f_\varepsilon(x)| \le \varepsilon. \tag{1.13}$$

Demostración. Este lema es un caso particular del teorema de Stone-Weierstrass. ■

Teorema 1.20 *Sea $\{H_n,\ n \ge 1\}$ una sucesión de funciones en $\mathcal{M}$. Entonces $H_n \xrightarrow{v} H$ si y sólo si para toda $f \in \mathcal{C}_K$,*

$$\int_{\mathbb{R}} f\,dH_n \to \int f\,dH. \tag{1.14}$$

Demostración. Supongamos que $H_n \xrightarrow{d} H$. (1.14) es cierta si f es la función indicadora de $(a, b]$ para $a, b \in D$ donde D es un subconjunto denso de $\mathcal{C}(H)$. Por la linealidad de las integrales también es cierto si f es una función escalera definida a partir de una partición con valores en D. Si $f \in \mathcal{C}_K$ y $\varepsilon > 0$, por el lema de aproximación existe una función escalera f_ε que satisface (1.13). Tenemos

$$|\int f\,dH_n - \int f\,dH| \le |\int (f - f_\varepsilon)\,dH_n| + |\int f_\varepsilon\,dH_n - \int f_\varepsilon\,dH| + |\int (f_\varepsilon - f)\,dG|. \tag{1.15}$$

El primer término del lado derecho está acotado por

$$\int |f - f_\varepsilon|\,dH_n \le \varepsilon \int dH_n \le \varepsilon,$$

y una acotación similar sirve para el tercer término. El segundo término tiende a cero porque f_ε es una función escalera a valores en D. Por lo tanto el lado derecho de (1.15) está acotado por 2ε cuando $n \to \infty$ y por lo tanto converge a 0.

Para demostrar el recíproco basta observar que las funciones g_k y h_k que definimos para la demostración del teorema 1.10 son funciones continuas a soporte compacto y por lo tanto el mismo argumento demuestra la convergencia

$$H_n(a,b] \to H(a,b] \quad (n \to \infty)$$

para $a, b \in \mathcal{C}(H)$. ■

1.8.1. Convergencia Vaga y Tensión.

Ahora consideraremos la relación entre convergencia vaga y convergencia en distribución.

Teorema 1.21 (Prohorov) *Sea $\{F_n,\ n \geq 1\}$ una sucesión de f.d. Para que toda subsucesión que converge vagamente converja en distribución es necesario y suficiente que*

$$\int_{|x|>a} dF_n(x) \to 0 \quad \textit{cuando } a \to \infty \tag{1.16}$$

uniformemente en n.

Demostración.

Supongamos que $F_{n_k} \xrightarrow{v} H$ para alguna $H \in \mathcal{M}$ cuando $k \to \infty$ y que (1.16) vale. Como la masa total de H es finita tenemos, para cualquier $\varepsilon > 0$, existe a_H tal que

$$\int_{|x|>a} dH(x) < \varepsilon \quad \text{para } a > a_H.$$

Además, por hipótesis tenemos que

$$\sup_n \int_{|x|>a} dF_n(x) < \varepsilon \quad \text{para } a > a_0.$$

Entonces, para $a > \text{máx}\{a_h, a_0\}$, tenemos que para todo k,

$$\begin{aligned}
1 - \int_{-\infty}^{\infty} dH(x) &= \int_{-\infty}^{\infty} dF_{n_k}(x) - \int_{-\infty}^{\infty} dH(x) \\
&\leq \int_{|x|>a} dF_{n_k}(x) + \Big| \int_{-a}^{a} dF_{n_k}(x) - \int_{-a}^{a} dH(x) \Big| + \int_{|x|>a} dH(x) \\
&\leq \varepsilon + \Big| \int_{-a}^{a} dF_{n_k}(x) - \int_{-a}^{a} dH(x) \Big| + \varepsilon
\end{aligned}$$

de modo que, por el teorema 1.20,

$$1 - \int_{-\infty}^{\infty} dH(x) < 2\varepsilon,$$

lo cual implica que la masa total de H es 1.

Veamos ahora la demostración de la necesidad. Sabemos que $F_n \xrightarrow{d} H$ cuando $n \to \infty$, de modo que H es propia, y queremos verificar que (1.16) vale. Supongamos lo contrario, entonces, dado $\varepsilon > 0$ existen sucesiones crecientes y no acotadas $\{n_k,\ k \geq 1\}$ y $\{a_k,\ k \geq 1\}$ tales que

$$\int_{|x|>a_k} dF_{n_k}(x) > 2\varepsilon.$$

Por otro lado, como antes,

$$\int_{|x|>a} dH(x) < \varepsilon \quad \text{para } a > a_H,$$

Además, como sólo estamos considerando una cantidad numerable de distribuciones y una f.d. tiene, a lo sumo, una cantidad numerable de discontinuidades, podemos suponer sin restricción que a y $(a_k)_{k\geq 1}$ son puntos de continuidad de todas las distribuciones que consideramos.

Dado cualquier $a > a_H$ existe un k_0 tal que $a_k > a$ para $k > k_0$. Sean a y $k > k_0$ dados, combinando los resultados anteriores obtenemos

$$\begin{aligned}
1 - \int_{-\infty}^{\infty} dH(x) &= \int_{-\infty}^{\infty} dF_{n_k}(x) - \int_{-\infty}^{\infty} dH(x) \\
&= \int_{|x|>a} dF_{n_k}(x) - \int_{|x|>a} dH(x) + \Big(\int_{-a}^{a} dF_{n_k}(x) - \int_{-a}^{a} dH(x) \Big) \\
&> 2\varepsilon - \varepsilon + \Big(\int_{-a}^{a} dF_{n_k}(x) - \int_{-a}^{a} dH(x) \Big)
\end{aligned}$$

y por el teorema 1.20,

$$1 - \int_{-\infty}^{\infty} dH(x) > \varepsilon,$$

de modo que H no es propia. Esta contradicción concluye la demostración del teorema. ■

Una sucesión de distribuciones que satisface (1.16) no permite que la masa escape a infinito por la uniformidad de la condición en n: Todas las distribuciones tienen masas uniformemente pequeñas fuera de intervalos suficientemente grandes. Decimos que una sucesión de distribuciones que satisfacen esta propiedad es *tensa*. Si $\{X_n,\ n \geq 1\}$ es una sucesión de v.a. cuyas distribuciones asociadas forman una sucesión tensa decimos que la sucesión $\{X_n,\ n \geq 1\}$ es *tensa*.

El teorema de Prohorov dice que una familia de f.d. es tensa si y sólo si es relativamente compacta. Combinando el teorema de Prohorov y el corolario 1.6 obtenemos el siguiente resultado.

Corolario 1.7 *Sea $\{F_n,\ n \geq 1\}$ una sucesión tensa de f.d. Si todas las subsucesiones convergentes de F_n convergen al mismo límite F, entonces $F_n \xrightarrow{d} F$.*

Observamos que para cualquier $p > 0$

$$P(|X_n| > a) \leq \frac{1}{a^p} \operatorname{E} |X_n|^p \mathbf{1}_{\{|X|>a\}} \leq \frac{1}{a^p} \sup_n \operatorname{E} |X_n|^p,$$

de donde obtenemos la primera parte del siguiente corolario.

Corolario 1.8 *(i) Sea $p > 0$. Una sucesión acotada en L^p es tensa. En particular, sucesiones uniformemente integrables son tensas.*

(ii) Sea $g \uparrow \infty$ una función no negativa y supongamos que $\sup_n \operatorname{E} g(X_n) < \infty$. Entonces $\{X_n,\ n \geq 1\}$ es tensa.

Usando el teorema 1.21 podemos demostrar el siguiente resultado.

Teorema 1.22 *Sea $\{X_n,\ n \geq 1\}$ una sucesión tensa. Entonces $X_n \xrightarrow{d} X$ cuando $n \to \infty$ si y sólo si toda subsucesión contiene una subsucesión que converge vagamente a X.*

1.8.2. Clases Separantes de Funciones

Definición 1.12 Una clase $\mathcal{E}$ de funciones continuas y acotadas en $\mathbb{R}$ es *separante* si para cualquier par $F,\ G$ de f.d. se tiene que

$$\int f\,dF = \int f\,dG, \quad \forall f \in \mathcal{E} \quad \Rightarrow \quad F = G.$$

Proposición 1.7 *Sea $\mathcal{E}$ una clase separante y $\{F_n,\ n \geq 1\}$ una sucesión tensa de f.d. Entonces existe una f.d. F tal que $F_n \xrightarrow{d} F$ si y sólo si*

$$\lim_n \int f\,dF_n \text{ existe para toda } f \in \mathcal{E}.$$

Si esto es cierto, entonces $\lim_n \int f\,dF_n = \int f\,dF$, para toda $f \in \mathcal{E}$.

Demostración. Si $F_n \xrightarrow{d} F$ entonces $\lim_n \int f dF_n = \int f\,dF$ porque las funciones en $\mathcal{E}$ son continuas y acotadas. Para ver el recíproco sea F_{n_k} una subsucesión convergente. Por la tensión $F_{n_k} \xrightarrow{d} F$ donde F es una f.d. propia. Si tomamos cualquier otra subsucesión débilmente convergente $F_{n'_k} \xrightarrow{d} G$, para $f \in \mathcal{E}$ se tiene que

$$\lim_k \int f\,dF_{n_k} = \int f\,dF, \quad \lim_k \int f\,dF_{n'_k} = \int f\,dG,$$

de modo que $\int f\,dF = \int f\,dG$ para toda $f \in \mathcal{E}$. Esto implica que $F = G$ y en consecuencia todas las subsucesiones convergentes de F_n tienen el mismo límite F. Por el corolario 1.7 tenemos $F_n \xrightarrow{d} F$. ∎

Corolario 1.9 *Sea $\mathcal{E}$ una clase separante y $\{F_n,\ n \geq 1\}$ una colección tensa de f.d. Si F es una f.d. tal que $\int f\,dF_n \to \int f\,dF$ para toda $f \in \mathcal{E}$, entonces $F_n \xrightarrow{d} F$.*

1.9. El Teorema de Skorohod

Sabemos que convergencia casi segura implica convergencia débil y que el recíproco es falso. Sin embargo, el teorema de Skorohod nos da un recíproco parcial.

Teorema 1.23 (Skorohod) *Sea $\{X_n,\ n \geq 0\}$ una sucesión de v.a. definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ tales que $X_n \xrightarrow{d} X_0$. Entonces existen v.a. $\{Y_n,\ n \geq 0\}$ definidas en el espacio de Lebesgue $([0,1], \mathcal{B}[0,1], m)$ tales que para cada $n \geq 0$ fijo, $X_n \stackrel{d}{=} Y_n$ y $Y_n \xrightarrow{c.s.} Y_0$ donde la convergencia casi segura es respecto a la medida de Lebesgue.*

Para la demostración del teorema necesitamos el siguiente resultado

Lema 1.7 *Sea F_n la f.d. de X_n, de modo que $F_n \xrightarrow{d} F_0$. Si $t \in (0,1) \cap \mathcal{C}(F_0^{\leftarrow})$, entonces $F_n^{\leftarrow}(t) \to F_0^{\leftarrow}(t)$.*

Demostración. Como $\mathcal{C}(F_0)^c$ es a lo sumo numerable, dado $\varepsilon > 0$ existe $x \in \mathcal{C}(F_0)$ tal que

$$F_0^{\leftarrow}(t) - \varepsilon < x < F_0^{\leftarrow}(t).$$

A partir de la definición de la función inversa, $x < F_0^{\leftarrow}(t)$ implica que $F_0(x) < t$. Además, $x \in \mathcal{C}(F_0)$ implica $F_n(x) \to F_0(x)$, de modo que para n grande tenemos $F_n(t) < t$. De nuevo, usando la definición de la función inversa, tenemos $x \leq F_n^{\leftarrow}(t)$. Por lo tanto,

$$F_0^{\leftarrow}(t) - \varepsilon < x \leq F_n^{\leftarrow}(t).$$

para todo n grande, y como $\varepsilon > 0$ es arbitrario, obtenemos que

$$F_0^{\leftarrow}(t) \leq \liminf_{n\to\infty} F_n^{\leftarrow}(t). \tag{1.17}$$

Para cualquier $t' > t$ podemos hallar $y \in \mathcal{C}(F_0)$ tal que

$$F_0^{\leftarrow}(t') < y < F_0^{\leftarrow}(t') + \varepsilon.$$

y esto implica $F_0(y) \geq t' > t$. Como $y \in \mathcal{C}(F_0)$, $F_n(y) \to F_0(y)$ y para n grande, $F_n(y) \geq t$, de modo que $y \geq F_n^{\leftarrow}(t)$ y en consecuencia

$$F_n^{\leftarrow}(t) \leq y < F_0^{\leftarrow}(t') + \varepsilon.$$

para todo n grande. Como $\varepsilon > 0$ es arbitrario obtenemos

$$\limsup_{n\to\infty} F_n^{\leftarrow}(t) \leq F_0^{\leftarrow}(t').$$

Hacemos ahora $t' \downarrow t$ y usamos la continuidad de $F_0^{\leftarrow}$ en t para obtener

$$\limsup_{n\to\infty} F_n^{\leftarrow}(t) \leq F_0^{\leftarrow}(t). \tag{1.18}$$

Combinando (1.17) y (1.18) obtenemos el resultado. ■

Demostración del teorema de Skorohod. En el espacio $[0,1]$ definimos la variable aleatoria $U(t) = t$ de modo que U tiene distribución uniforme, ya que para $0 \leq x \leq 1$

$$m(U \leq x) = m(t \in [0,1] : U(t) \leq x) = m([0,x]) = x.$$

Para $n \geq 0$ definimos Y_n en $[0,1]$ por $Y_n = F_n^{\leftarrow}(U)$. Entonces para $y \in \mathbb{R}$,

$$m(Y_n \leq y) = m(t \in [0,1] : F_n^{\leftarrow}(t) \leq y) = m(t \in [0,1] : t \leq F_n(y)) = F_n(y),$$

y concluimos que $Y_n \stackrel{d}{=} X_n$ para todo $n \geq 0$. Ahora escribimos

$$m(t \in [0,1] : Y_n(t) \nrightarrow Y_0(t)) = m(t \in [0,1] : F_n^{\leftarrow}(t) \nrightarrow F_0^{\leftarrow}(t)),$$

y usando el lema 1.7, esto esta acotado por

$$m(t \in [0,1] : F_0^{\leftarrow} \text{ no es continua en } t) = 0$$

porque el conjunto de discontinuidades es numerable. ■

1.9.1. El Método Delta

Supongamos que tenemos una familia de modelos $\{(\Omega, \mathcal{F}, P_\theta),\ \theta \in \Theta\}$, y queremos estimar el parámetro θ con base en una muestra de tamaño n usando el estadístico $T_n = T_n(X_1, \dots, X_n)$. El estimador T_n es consistente si

$$T_n \xrightarrow{P_\theta} \theta,$$

para cada θ. El estimador es consistente y asintóticamente normal (CAN) si para todo $\theta \in \Theta$

$$\lim_{n\to\infty} P_\theta(\sigma_n(T_n - \theta) \leq x) = \mathcal{N}(x; 0, 1),$$

para alguna sucesión $\sigma_n \to \infty$.

Supongamos que tenemos un estimador CAN de θ pero queremos estimar una función suave $g(\theta)$. Por ejemplo, en una familia de densidades exponenciales, θ puede representar la media pero estamos interesados en la varianza θ^2. Usando el método delta vemos que $g(T_n)$ también es CAN para $g(\theta)$.

Ilustramos el método usando el TCL que demostraremos más adelante. Sea $\{X_j,\ j \geq 1\}$ una sucesión i.i.d. con $\mathrm{E}(X_n) = \mu$ y $\mathrm{Var}(X_n) = \sigma^2$. El TCL dice que

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

Equivalentemente, podemos expresar este resultado en términos de $\bar{X}_n = \sum_1^n X_i/n$ como

$$\sqrt{n}\Big(\frac{\bar{X}_n - \mu}{\sigma}\Big) \xrightarrow{d} \mathcal{N}(0,1),$$

de modo que $\bar{X}$ es un estimador CAN de μ. El método delta dice que si $g(x)$ tiene derivada no nula $g'(\mu)$ en μ entonces

$$\sqrt{n}\Big(\frac{g(\bar{X}_n) - g(\mu)}{\sigma g'(\mu)}\Big) \xrightarrow{d} \mathcal{N}(0,1). \tag{1.19}$$

Vemos que $g(\bar{X})$ es CAN para $g(\mu)$.

Observación 1.9 La demostración no depende de que la variable límite tenga distribución $\mathcal{N}(0,1)$.

Demostración de (1.19). Por el teorema de Skorohod existen variables Z'_n y N' en el espacio de Lebesgue tales que

$$Z'_n \stackrel{d}{=} \sqrt{n}\Big(\frac{\bar{X} - \mu}{\sigma}\Big), \qquad N' \sim \mathcal{N}(0,1)$$

y

$$Z'_n \to N' \quad c.s.\ (m).$$

Definimos

$$\bar{X}'_n = \mu + \sigma Z'_n/\sqrt{n},$$

de modo que $\bar{X}_n \stackrel{d}{=} \bar{X}'_n$. Usando la definición de derivada

$$\begin{aligned}\sqrt{n}\Big(\frac{g(\bar{X}_n) - g(\mu)}{\sigma g'(\mu)}\Big) &\stackrel{d}{=} \sqrt{n}\Big(\frac{g(\mu + \sigma Z'_n/\sqrt{n}) - g(\mu)}{\sigma g'(\mu)}\Big)\\ &= \frac{g(\mu + \sigma Z'_n/\sqrt{n}) - g(\mu)}{\sigma Z'_n/\sqrt{n}} \frac{Z'_n}{g'(\mu)}\\ &\xrightarrow{c.s.} g'(\mu)\frac{N'}{g'(\mu)} = N' \sim \mathcal{N}(0,1),\end{aligned}$$

ya que $\sigma Z'_n/\sqrt{n} \to 0$ c.s. ■

1.10. Convergencia en Distribución II

Teorema 1.24 (Slutsky) *Sean $\{X, X_n, Y_n, \xi_n, n \geq 1\}$ variables aleatorias. Si $X_n \xrightarrow{d} X$ y $X_n - Y_n \xrightarrow{P} 0$ entonces $Y_n \xrightarrow{d} X$. Equivalentemente, si $X_n \xrightarrow{d} X$ y $\xi_n \xrightarrow{P} 0$ entonces $X_n + \xi_n \xrightarrow{d} X$.*

Demostración. Basta probar la segunda afirmación. Sea f una función real, acotada y uniformemente continua. Definimos el módulo de continuidad por

$$\omega_\delta(f) = \sup_{|x-y|\leq\delta} |f(x) - f(y)|.$$

Como f es uniformemente continua tenemos que

$$\omega_\delta(f) \to 0, \qquad \text{cuando } \delta \to 0. \tag{1.20}$$

Por el corolario 1.4 basta ver que $\mathrm{E}(f(X_n + \xi_n)) \to \mathrm{E}(f(X))$. Para esto observemos que

$$\begin{aligned} |\,\mathrm{E}(f(X_n+\xi_n)) - \mathrm{E}(f(X))| &\le |\,\mathrm{E}(f(X_n+\xi_n)) - \mathrm{E}(f(X_n))| + |\,\mathrm{E}(f(X_n)) - \mathrm{E}(f(X))| \\ &\le \mathrm{E}\left(|f(X_n+\xi_n) - f(X_n)|\mathbf{1}_{\{|\xi_n|\le\delta\}}\right) + 2\sup_x |f(x)| P(|\xi_n| > \delta) + o(1) \\ &\le \omega_\delta(f) + \mathrm{Const}\, P(|\xi_n| > \delta) + o(1). \end{aligned}$$

Haciendo $n \to \infty$ y luego $\delta \to 0$ y usando (1.20) obtenemos el resultado. ■

A continuación presentamos una generalización del teorema de Slutsky. La demostración de este resultado puede verse en el libro de Resnick.

Teorema 1.25 *Sean $\{X_{un}, X_u, Y_n, X, n \ge 1\}$ variables aleatorias tales que para cada n, Y_n, X_{un}, $u \ge 1$ están definidas en un mismo espacio de probabilidad. Supongamos que para cada u, cuando $n \to \infty$, $X_{un} \xrightarrow{d} X_u$ y cuando $u \to \infty$, $X_u \xrightarrow{d} X$. Supongamos además que para todo $\varepsilon > 0$,*

$$\lim_{u\to\infty} \operatorname*{limsup}_{n\to\infty} P(|X_{un} - Y_n| > \varepsilon) = 0.$$

Entonces tenemos $Y_n \xrightarrow{d} X$ cuando $n \to \infty$.

1.11. Teorema de Convergencia a Familias

Muchos resultados de convergencia de variables aleatorias son del siguiente tipo: Para una sucesión de v.a. $\xi_n, n \ge 1$ y constantes $a_n > 0$ y $b_n \in \mathbb{R}$, se demuestra que

$$\frac{\xi_n - b_n}{a_n} \xrightarrow{d} Y,$$

donde Y es una v. a. no-degenerada. Usando esto tenemos

$$P(\frac{\xi_n - b_n}{a_n} \le x) \approx P(Y \le x) = G(x),$$

o poniendo $y = a_n x + b_n$,

$$P(\xi_n \le y) \approx G(\frac{y - b_n}{a_n}).$$

Esto permite aproximar la distribución de ξ_n por una familia de distribuciones con parámetros de ubicación y escala.

La pregunta ahora es ¿Hasta qué punto son únicas estas constantes de normalización a_n y b_n? La respuesta la da el teorema de convergencia a familias de distribuciones: Las constantes están determinadas salvo por equivalencias asintóticas y la distribución límite está determinada salvo por parámetros de ubicación y escala.

Definición 1.13 Dos distribuciones F y G son *del mismo tipo* o *pertenecen a la misma familia* si para algunas constantes $a > 0$, $b \in \mathbb{R}$,

$$G(x) = F(ax + b), \qquad x \in \mathbb{R}.$$

En términos de variables aleatorias, si $X \sim F$ y $Y \sim G$ entonces

$$Y \stackrel{d}{=} \frac{X-b}{a}.$$

Por ejemplo, podemos considerar la familia gaussiana. Si $X_{\mu,\sigma}$ tiene distribución $N(\mu,\sigma^2)$, entonces $X_{\mu,\sigma} \stackrel{d}{=} \sigma X_{0,1} + \mu$.

Teorema 1.26 (Convergencia a familias, Gnedenko & Khinchin) *Sean $G(x)$ y $H(x)$ dos funciones de distribución propias, ninguna de las cuales está concentrada en un punto. Supongamos que para $n \geq 1$, X_n son v.a. con funciones de distribución F_n y U y V son v.a. con f.d. G y H, respectivamente. Sean, además, constantes $a_n > 0$, $\alpha_n > 0$, $b_n \in \mathbb{R}, \beta_n \in \mathbb{R}$.*
a) Si

$$F_n(a_n x + b_n) \to G(x), \quad F_n(\alpha_n x + \beta_n) \to H(x) \tag{1.21}$$

o equivalentemente

$$\frac{X_n - b_n}{a_n} \xrightarrow{d} U, \quad \frac{X_n - \beta_n}{\alpha_n} \xrightarrow{d} V,$$

entonces existen constantes $A > 0$ y $B \in \mathbb{R}$ tales que, cuando $n \to \infty$,

$$\frac{\alpha_n}{a_n} \to A > 0, \quad \frac{\beta_n - b_n}{a_n} \to B, \tag{1.22}$$

y

$$H(x) = G(Ax+B), \quad V \stackrel{d}{=} \frac{U-B}{A}. \tag{1.23}$$

b) Recíprocamente, si (1.22) vale, entonces cualquiera de las relaciones en (1.21) implica la otra y (1.23) vale.

Demostración
Veamos primero la demostración de (b). Supongamos que

$$G_n(x) := F_n(a_n x + b_n) \to G(x)$$

y

$$\frac{\alpha_n}{a_n} \to A > 0, \quad \frac{\beta_n - b_n}{a_n} \to B,$$

entonces

$$F_n(\alpha_n x + \beta_n) = G_n(\frac{\alpha_n}{a_n} x + \frac{\beta_n - b_n}{a_n}).$$

Escogemos $x \in \mathcal{C}(G(A \cdot + B))$. Supongamos que $x > 0$, un argumento similar sirve si $x \leq 0$. Dado $\varepsilon > 0$ para n grande tenemos

$$(A-\varepsilon)x + B - \varepsilon \leq \frac{\alpha_n}{a_n} x + \frac{\beta_n - b_n}{a_n} \leq (A+\varepsilon)x + B + \varepsilon$$

y en consecuencia

$$\limsup_{n\to\infty} F_n(\alpha_n x + \beta_n) \leq \limsup_{n\to\infty} G_n((A+\varepsilon)x + B + \varepsilon).$$

Por lo tanto, para cualquier $z \in \mathcal{C}(G)$ con $z > (A+\varepsilon)x + B + \varepsilon$ tenemos

$$\limsup_{n\to\infty} F_n(\alpha_n x + \beta_n) \leq \limsup_{n\to\infty} G_n(z) = G(z).$$

En consecuencia,

$$\limsup_{n\to\infty} F_n(\alpha_n x + \beta_n) \leq \inf\{G(z) : z > (A+\varepsilon)x + B + \varepsilon, z \in \mathcal{C}(G)\}.$$

y como $\varepsilon > 0$ es arbitrario,

$$\limsup_{n\to\infty} F_n(\alpha_n x + \beta_n) \leq \inf\{G(z) : z > Ax + B\} = G(Ax + B)$$

por la continuidad por la derecha de G. De manera similar,

$$\begin{aligned}\liminf_{n\to\infty} F_n(\alpha_n x + \beta_n) &\geq \liminf_{n\to\infty} G_n((A-\varepsilon)x + B - \varepsilon)\\ &\geq \liminf_{n\to\infty} G_n(z) = G(z)\end{aligned}$$

para cualquier $z < (A-\varepsilon)x + B - \varepsilon$, $z \in \mathcal{C}(G)$. Como estas desigualdades valen para todo $\varepsilon > 0$,

$$\liminf_{n\to\infty} F_n(\alpha_n x + \beta_n) \geq \sup\{G(z) : z < Ax + B,\ z \in \mathcal{C}(G)\} = G(Ax + B)$$

porque $Ax + B \in \mathcal{C}(G)$.

Veamos ahora la demostración de la parte (a). Supongamos que

$$G_n(x) := F_n(a_n x + b_n) \to G(x), \qquad H_n(x) := F_n(\alpha_n x + \beta_n) \to H(x).$$

Usando el lema 1.7 tenemos que $G_n^{\leftarrow}(y) \to G^{\leftarrow}(y)$ para $y \in \mathcal{C}(G^{\leftarrow})$ y $H_n^{\leftarrow}(y) \to H^{\leftarrow}(y)$ para $y \in \mathcal{C}(H^{\leftarrow})$, y es fácil verificar que

$$G_n^{\leftarrow}(y) = \frac{F_n^{\leftarrow}(y) - b_n}{a_n}, \qquad H_n^{\leftarrow}(y) = \frac{F_n^{\leftarrow}(y) - \beta_n}{\alpha_n}.$$

Por lo tanto

$$\begin{aligned}\frac{F_n^{\leftarrow}(y) - b_n}{a_n} &\to G^{\leftarrow}(y), \quad y \in \mathcal{C}(G^{\leftarrow}),\\ \frac{F_n^{\leftarrow}(y) - \beta_n}{\alpha_n} &\to H^{\leftarrow}(y), \quad y \in \mathcal{C}(H^{\leftarrow}).\end{aligned}$$

Como las f.d. $G(x)$ y $H(x)$ no están concentradas en un punto, podemos hallar $y_1 < y_2$ con $y_i \in \mathcal{C}(G^{\leftarrow}) \cap \mathcal{C}(H^{\leftarrow})$ para $i = 1, 2$, tales que

$$-\infty < G^{\leftarrow}(y_1) < G^{\leftarrow}(y_2) < \infty, \qquad -\infty < H^{\leftarrow}(y_1) < H^{\leftarrow}(y_2) < \infty.$$

Por lo tanto, para $i = 1, 2$ tenemos

$$\frac{F_n^{\leftarrow}(y_i) - b_n}{a_n} \to G^{\leftarrow}(y_i), \quad \frac{F_n^{\leftarrow}(y_i) - \beta_n}{\alpha_n} \to H^{\leftarrow}(y_i). \tag{1.24}$$

En las expresiones anteriores restamos las ecuaciones con $i = 1$ de las expresiones con $i = 2$ para obtener

$$\frac{F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)}{a_n} \to G^{\leftarrow}(y_2) - G^{\leftarrow}(y_1), \tag{1.25}$$

$$\frac{F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)}{\alpha_n} \to H^{\leftarrow}(y_2) - H^{\leftarrow}(y_1). \tag{1.26}$$

Ahora dividimos (1.25) entre (1.26) y obtenemos

$$\frac{\alpha_n}{a_n} \to \frac{G^{\leftarrow}(y_2) - G^{\leftarrow}(y_1)}{H^{\leftarrow}(y_2) - H^{\leftarrow}(y_1)} := A > 0.$$

Además, de (1.24) vemos que

$$\frac{F_n^{\leftarrow}(y_1) - b_n}{a_n} \to G^{\leftarrow}(y_1),$$

$$\frac{F_n^{\leftarrow}(y_1)-\beta_n}{a_n}=\frac{F_n^{\leftarrow}(y_1)-\beta_n}{\alpha_n}\frac{\alpha_n}{a_n}\to H^{\leftarrow}(y_1)A,$$

y restando obtenemos

$$\frac{\beta_n-b_n}{a_n}\to H^{\leftarrow}(y_1)A-G^{\leftarrow}(y_1):=B,$$

como queríamos ver, de modo que (1.22) vale. Por la parte (b) obtenemos (1.23).

■

Observación 1.10 Una consecuencia de la demostración es que, a partir de (1.24), (1.25) y (1.26), una posible selección de las constantes de normalización es

$$a_n=F_n^{\leftarrow}(y_2)-F_n^{\leftarrow}(y_1), \qquad b_n=F_n^{\leftarrow}(y_1).$$

El siguiente ejemplo muestra la importancia de la hipótesis de que las distribuciones límite no estén concentradas en un punto.

Ejemplo 1.14
Sea

$$G(x)=\begin{cases}0, & \text{si } t<c,\\ 1, & \text{si } t\geq c.\end{cases}$$

Entonces,

$$G^{\leftarrow}(t)=\inf\{y:G(y)\geq t\}=\begin{cases}-\infty, & \text{si } t=0\\ c, & \text{si } 0<t\leq 1,\\ \infty, & \text{si } t>1.\end{cases}$$

El siguiente corolario es consecuencia inmediata del teorema de convergencia a familias.

Corolario 1.10 *Sea F_n una sucesión de f. d. y $a_n>0$ y b_n sucesiones de constantes tales que*

$$F_n(a_nx+b_n)\to G(x) \tag{1.27}$$

en todo punto de continuidad de G, que es un a f.d. propia y no está concentrada en un punto. Sean $c_n>0$ y d_n sucesiones de constantes tales que

$$\frac{a_n}{c_n}\to 1, \qquad \frac{d_n-b_n}{a_n}\to 0.$$

Entonces (1.27) vale con c_n y d_n en lugar de a_n y b_n.