# Índice general

# Capítulo 1

# LEYES $0-1$. CONVERGENCIA DE SERIES DE VV.AA.

## 1.1. Ley $0-1$ de Kolmogorov

**Definición 1.1.1.** *Sean $\{X_1, X_2, \cdots\}$, una sucesión de variables aleatorias (vv.aa.), y sea $\mathcal{F}_n$ la $\sigma$-álgebra generada por las variables $\{X_{n+k} : k = 0, 1, \cdots\}$, es decir, $\mathcal{F}_n := \sigma\{X_n, X_{n+1}, \cdots\}$. La $\sigma$-álgebra $\mathcal{F}_\infty := \cap_{n=1}^{\infty}\mathcal{F}_n$ se llama la $\sigma$-*álgebra cola *de la sucesión de vv.aa. $\{X_n\}$. Un evento $A \in \mathcal{F}_\infty$ se llama* evento cola*, y si $X$ es una v.a. $\mathcal{F}_\infty$-medible, entonces diremos que $X$ es una función cola con respecto a $\{X_n\}$.*

**Observación 1.1.2.** *Intuitivamente, un evento (o función) cola es aquel cuya ocurrencia o no ocurrencia no varía con ningún cambio finito de las $X_i$'s, es decir, no depende de ningún número finito de sus "coordenadas"$X_i$'s.*

**Ejemplo 1.1.0.1.**

*(a) Sea $\{B_n\}$ una sucesión de borelianos de $\mathbb{R}$. Entonces*

$$A := \limsup_{n\to\infty}[X_n \in B_n]$$

*es un evento cola. En efecto, note que*

$$A = \cap_{m=k}^{\infty}(\cup_{n=m}^{\infty}[X_n \in B_n]) \in \mathcal{F}_k \quad \forall k \in \mathbb{N}.$$

*Entonces* $A \in \cap_{k=1}^{\infty}\mathcal{F}_k = \mathcal{F}_\infty$, *es decir* $A$ *es evento cola. De aquí se sigue que* $B := \liminf_{n\to\infty}[X_n \in B_n]$ *es también un evento cola.*

*(b) Las vv.aa.* $\limsup_{n\to\infty} X_n$ *y* $\liminf_{n\to\infty} X_n$ *son funciones cola. En efecto,*

$$\limsup_{n\to\infty} X_n = \inf_{m\in\mathbb{N}} \sup_{n\geq m} X_n = \inf_{m\geq k} \sup_{n\geq m} X_n \quad \forall k \in \mathbb{N},$$

*de donde* $\limsup_{n\to\infty} X_n$ *es* $\mathcal{F}_k$*-medible para todo* $k \in \mathbb{N}$. *Entonces* $\limsup_{n\to\infty} X_n$ *es* $\cap_{k=1}^{\infty}\mathcal{F}_k = \mathcal{F}_\infty$*-medible.*

*(c) Sea* $A := \{\omega \in \Omega : \text{ existe } \lim_{n\to\infty} X_n(\omega) \in \overline{\mathbb{R}}\}$.

**Lema 1.1.3.** (Teorema de aproximación.) *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad,* $\mathcal{F}_0$ *una álgebra de subconjuntos de* $\Omega$ *tal que* $\mathcal{F} = \sigma(\mathcal{F}_0)$ *y* $\mu$ *una medida* $\sigma$*-finita sobre* $\mathcal{F}_0$*, es decir, existe una sucesión* $\{C_n\}$ *de subconjuntos de* $\Omega$ *elementos de* $\mathcal{F}_0$ *satisfaciendo:*

*(a)* $\Omega = \cup_{n=1}^{\infty} C_n$*;*

*(b)* $\mu(C_n) < \infty$ *para todo* $n \in \mathbb{N}$.

*Entonces* $\forall\varepsilon > 0$, *y* $\forall A \in \mathcal{F}$, *con* $\mu(A) < \infty$, $\exists B \in \mathcal{F}_0$ *tal que* $\mu(A\Delta B) < \varepsilon$.

**Dem.**

**Parte I**. En esta parte supondremos que $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de medida finito. En esta parte demostraremos que: $\forall\varepsilon > 0$, y $\forall A \in \mathcal{F}$, $\exists B \in \mathcal{F}_0$ tal que $\mu(A\Delta B) < \varepsilon$, donde $A\Delta B$ es la *diferencia simétrica* entre $A$ y $B$, definido por $A\Delta B := A - B \cup B - A$.

Definimos

$$\mathcal{H} := \{A \in \mathcal{F} : \forall\varepsilon > 0 \exists B \in \mathcal{F}_0 \text{ tal que } \mu(A\Delta B) < \varepsilon\}.$$

Afirmamos que $\mathcal{H}$ es una $\sigma$-álgebra sobre $\Omega$. Claramente $\mathcal{F}_0 \subset \Omega \subset \mathcal{F}$, de donde $\Omega \in \mathcal{F}_0$ implica necesariamente que $\Omega \in \mathcal{H}$, además, como $A\Delta B = \mathcal{C}A\Delta\mathcal{C}B$, se sigue que $\mathcal{H}$ es cerrado bajo complementación. Resta probar que $\mathcal{H}$ es cerrado bajo uniones numerables.

Sea $\{A_n\}$ una sucesión de conjuntos elementos de la familia $\mathcal{H}$. Definamos $A := \cup_{n=1}^{\infty} A_n$. Dado $\varepsilon > 0$ arbitrario, como $\mu$ es medida finita, y $\mu(A) = \lim_{N\to\infty} \mu(\cup_{n=1}^{N} A_n)$, existe $N_0 \in \mathbb{N}$ tal $\mu(A - (\cup_{n=1}^{N_0} A_n)) < \varepsilon/2$. De aquí para cada $n = 1, \cdots, N_0$ existe $B_n \in \mathcal{F}_0$ tal que $\mu(A_n\Delta B_n) < \varepsilon/(2N_0)$. Definamos $B := \cup_{n=1}^{N_0} B_n \in \mathcal{F}_0$. Entonces

$$\begin{aligned}
\mu(A\Delta B) &\leq \mu(A\Delta(\cup_{n=1}^{N_0}A_n)) + \mu((\cup_{n=1}^{N_0}A_n)\Delta(\cup_{n=1}^{N_0}B_n)) \\
&\leq \varepsilon/2 + \sum_{n=1}^{N_0}\mu(A_n\Delta B_n) \\
&\leq \varepsilon/2 + \sum_{n=1}^{N_0}\varepsilon/(2N_0) = \varepsilon.
\end{aligned}$$

Luego $A = \cup_{n=1}^{\infty}A_n \in \mathcal{H}$. Así $\mathcal{H}$ es una $\sigma$-álgebra que contiene a la álgebra $\mathcal{F}_0$ y está contenida en la $\sigma$-álgebra $\mathcal{F}$, de donde $\sigma(\mathcal{F}_0) = \mathcal{F} \subset \mathcal{H} \subset \mathcal{F}$, o sea $\mathcal{H} = \mathcal{F}$, probándose la primera parte.

**Parte II**. Supongamos ahora que $\mu$ es $\sigma$-finita sobre $\mathcal{F}_0$, es decir, existe una sucesión $\{C_n\}$ de subconjuntos de $\Omega$ elementos de la álgebra $\mathcal{F}_0$ satisfaciendo los incisos (a) y (b) del enunciado de éste lema. Sin pérdida de generalidad podemos suponer que $\{C_n\}$ es sucesión creciente. Luego, dado $\varepsilon > 0$, $A \in \mathcal{F}$ con $\mu(A) < \infty$, de donde $\mu(A) = \lim_{n\to\infty}\mu(A \cap C_n)$, y así existe $N_0 \in \mathbb{N}$ tal que

$$\mu(A) - \mu(A \cap C_{N_0}) < \varepsilon/2. \tag{1.1.1}$$

Definimos la medida finita $\nu : \mathcal{F} \to [0,\infty)$, por $\nu(D) = \mu(D \cap C_{N_0})$ para cada $D \in \mathcal{F}$. Por la parte I, existe $B \in \mathcal{F}_0$ tal que $\nu(A\Delta B) = \mu((A\Delta B)\cap C_{N_0}) < \varepsilon/2$ que junto a (1.1.1), definiendo $B_0 := B \cap C_{N_0} \in \mathcal{F}_0$, se tiene

$$\begin{aligned}
\mu(A\Delta B_0) &\leq \mu(A\Delta(A\cap C_{N_0})) + \mu((A\cap C_{N_0})\Delta(B\cap C_{N_0})) \\
&\leq \varepsilon/2 + \mu((A\Delta B)\cap C_{N_0}) \\
&\leq \varepsilon/2 + \varepsilon/2 = \varepsilon.
\end{aligned}$$

Esto prueba el lema. ■

**Observación 1.1.4. Aplicación**. Sea $\{X_n\}$ una sucesión de vv.aa. y sea $\mathcal{F}_0 := \cup_{n=1}^{\infty}\sigma(X_1,\cdots,X_n)$. Claramente $\mathcal{F}_0$ es una álgebra de subconjuntos de $\Omega$. Definimos $\mathcal{F}_1 := \sigma(X_1, X_2, \cdots)$. Entonces $\mathcal{F}_1 = \sigma(\mathcal{F}_0)$. Por el Lema 1.1.3 para cada $\varepsilon > 0$ y para cada $A \in \mathcal{F}_1$ existe $n \in \mathbb{N}$, y además $B \in \sigma(X_1,\cdots,X_n)$ tal que $\mathbb{P}(A\Delta B) < \varepsilon$.

**Teorema 1.1.5. Ley $0-1$ de Kolmogorov** Sean $X_1, X_2, \cdots$, vv.aa. independientes y $\mathcal{F}_\infty$ la correspondiente $\sigma$-álgebra cola. Entonces cualquier evento cola $A \in \mathcal{F}_\infty$ tiene probabilidad 0 ó 1. Además, cualquier función cola es constante.

**Dem.** Sea $A \in \mathcal{F}_\infty$ evento cola. A grandes rasgos, la idea de la demostración consiste en probar que $A$ es independiente de sí mismo, en cuyo caso, $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, de donde $\mathbb{P}(A) = 0$, ó bien $\mathbb{P}(A) = 1$.

Usando la notación de la observación 1.1.4, tenemos que si $A \in \mathcal{F}_\infty$, entonces $A \in \mathcal{F}_1$ (pues $\mathcal{F}_\infty \subset \mathcal{F}_1$). Por el lema 1.1.3, existen $\{B_n\}$ conjuntos tales que $B_n \in \sigma(X_1, \cdots, X_{N_n})$, con $\{N_n\}$ sucesión creciente de números naturales divergentes al $+\infty$, tal que $\mathbb{P}(A\Delta B_n) \to 0$ cuando $n \to \infty$.

Como $\mathbb{P}(A\Delta B_n) = \mathbb{P}(A - A \cap B_n) + \mathbb{P}(B_n - A \cap B_n)$, se sigue de aquí que

$$\mathbb{P}(A \cap B_n) \to \mathbb{P}(A), \qquad \text{y} \quad \mathbb{P}(B_n) \to \mathbb{P}(A) \text{ cuando } n \to \infty. \tag{1.1.2}$$

Por otro lado, como $A \in \mathcal{F}_{N_n+1} = \sigma(X_{N_n+1}, X_{N_n+2}, \cdots)$, y en consecuencia, $A$ y $B_n$ son independientes. De los límites en (1.1.2), $\mathbb{P}(A \cap B_n) = \mathbb{P}(A)\mathbb{P}(B_n) \to \mathbb{P}(A)^2 = \mathbb{P}(A)$, de donde $\mathbb{P}(A) = 0$ ó $\mathbb{P}(A) = 1$.

Finalmente, sea $g : \Omega \to \mathbb{R}$ una función cola. Entonces para toda $c \in \mathbb{R}$, $A_c := [g < c] \in \mathcal{F}_\infty$. Por la primera parte $\mathbb{P}(A_c) = 0$ ó $\mathbb{P}(A_c) = 1$. Definamos ahora

$$k := \sup\{c \in \mathbb{R} : \mathbb{P}(g < c) = 0\}.$$

Como $[g < k] = \cup_{n=1}^{\infty}[g < c_n]$, donde $c_n \uparrow k$, se sigue inmediatamente que $\mathbb{P}(g < k) = 0$. De la definición de $k$ se tiene que $\mathbb{P}(g < k + 1/n) = 1$ para cada $n$. Poniendo $C_n := [k \leq g < k + 1/n] = [g < k + 1/n] - [g < k]$, se tiene que $\mathbb{P}(C_n) = 1$, con $C_n$ es sucesión decreciente tal que $C_n \downarrow [g = k]$, por lo tanto

$$\mathbb{P}(g = k) = \lim_{n\to\infty} \mathbb{P}(C_n) = 1.$$

■

**Observación 1.1.6.** *La ley* $0-1$ *de Kolmogorov tiene consecuencias muy interesantes. Por ejemplo se sigue que si* $X_1, X_2, \cdots$ *son vv.aa. independientes, entonces*

$$\mathbb{P}(\sum_{n=1}^{\infty} X_n \textit{ converge }) = 0 \textit{ ó } 1;$$

*y*

$$\mathbb{P}(\lim_{n\to\infty} X_n \textit{ existe }) = 0 \textit{ ó } 1.$$

## 1.2. Ley $0-1$ de Hewitt-Savage

**Definición 1.2.1.** *Sean $X_1, X_2, \cdots$ vv.aa., $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \cdots)$, $n \in \mathbb{N}$. Claramente, un evento $A \in \mathcal{F}_1$ es de la forma*

$$[(X_1, X_2, \cdots) \in A'] \quad \text{con } A' \in (\mathcal{B}(\mathbb{R}))^\infty =: \mathcal{B}^\infty.$$

*Decimos que $A$ es un* evento simétrico *si para cualquier permutación $T : \{1, 2, \cdots\} \to \{T(1), T(2), \cdots\}$ que permuta sólo un número finito de coordenadas, se satisface que $A = A(T)$, donde $A(T) := [(X_{T(1)}, X_{T(2)}, \cdots) \in A']$. En otras palabras, $A \in \mathcal{F}_1$ es un evento simétrico si y sólo si la ocurrencia o no ocurrencia de $A$ no se altera por una permutación de un número finito de las $X_i$'s.*

**Observación 1.2.2.** *Dado $n \in \mathbb{N}$, sea $\varphi_n : \Omega \to \mathbb{R}^\infty$ la aplicación $\varphi_n(\omega) = (X_n(\omega), X_{n+1}(\omega), \cdots) \in \mathbb{R}^\infty$ es una función medible de $(\Omega, \mathcal{F})$ en $(\mathbb{R}^\infty, \mathcal{B}^\infty)$, donde $\mathcal{B}^\infty$ es la $\sigma$-álgebra de Borel del espacio topológico producro $\mathbb{R}^\infty$. Así, $\mathcal{F}_n = \sigma(\varphi_n) = \varphi_n^{-1}(\mathcal{B}^\infty)$.*

*Consideremos una permutación $T : \{1, 2, \cdots\} \to \{T(1), T(2), \cdots\}$ que permuta sólo un número finito de coordenadas, y sea el homeomorfismo $\varphi_T : \mathbb{R}^\infty \to \mathbb{R}^\infty$, con $\varphi_T(x_1, x_2, \cdots) = (x_{T(1)}, x_{T(2)}, \cdots)$. Entonces $A \in \mathcal{F}_1$ es evento simétrico si y sólo si existe $A' \in \mathcal{B}^\infty$ tal que*

$$A = \varphi_1^{-1}(A') = (\varphi_T \circ \varphi_1)^{-1}(A') \quad \text{para toda permutación } T. \tag{1.2.1}$$

*Denotemos por Sym el conjunto de todos los eventos $A \in \mathcal{F}_1$ simétricos. Por (1.2.1) se tiene que Sym es una subsigma-álgebra de $\mathcal{F}_1$. De aquí, por la proposición 1.2.3 abajo*

$$\mathcal{F}_\infty \subset Sym \subset \mathcal{F}_1.$$

**Proposición 1.2.3.** *Cualquier evento cola es simétrico. El recíproco es falso.*

**Dem.** Sea $A$ un evento cola y sea $T$ una permutación que permuta sólo las primeras $n$ coordenadas. Puesto que $A \in \mathcal{F}_{n+1}$ podemos escribir $A$ en la forma $[(X_{n+1}, X_{n+2}, \cdots) \in A_{n+1}]$, por lo tanto

$$A = [(X_1, X_2, \cdots) \in \mathbb{R}^n \times A_{n+1}] = [(X_{T(1)}, X_{T(2)}, \cdots) \in \mathbb{R}^n \times A_{n+1}],$$

pues $T(k) = k$ para todo $k > n$.

El recíproco es falso. Por ejemplo $A = \{X_n = 0 \text{ para todo } n\} \in \mathcal{F}_1$ es un evento simétrico pero no es un evento cola porque $A \notin \mathcal{F}_n$ para todo $n > 1$. (Recuérdese que, intuitivamente, un evento cola es un evento cuya ocurrencia o no ocurrencia no se afecta cambiando un número finito de las $X_n$'s). ∎

**Teorema 1.2.4. (Ley $0-1$ de Hewitt-Savage)**. Sean $X_1, X_2, \cdots$ vv.aa.i.i.d. Si $A \in \mathcal{F}_1$ es un evento simétrico, entonces $\mathbb{P}(A) = 0$ ó bien $\mathbb{P}(A) = 1$.

**Dem.** Por el teorema de aproximación (lema 1.1.3), existen $A_n \in \sigma(X_1, \cdots, X_n)$ tales que $\mathbb{P}(A\Delta A_n) \to 0$ cuando $n \to \infty$. Similarmente a como se hizo en (1.1.2),

$$\mathbb{P}(A \cap A_n) \to \mathbb{P}(A), \qquad \text{y} \quad \mathbb{P}(A_n) \to \mathbb{P}(A) \text{ cuando } n \to \infty. \tag{1.2.2}$$

También, $A_n$ se puede escribir como

$$A_n = [(X_1, \cdots, X_n) \in B_n], \qquad B_n \in \mathcal{B}^n \quad \forall n,$$

donde $\mathcal{B}^n = \mathcal{B}(\mathbb{R}^n)$. Por hipótesis, $X(T) = (X_{T(1)}, X_{T(2)}, \cdots)$ tiene la misma distribución que $X = (X_1, X_2, \cdots)$, y , en consecuencia,

$$\mathbb{P}(A) = \mathbb{P}(X \in A') = \mathbb{P}(X(T) \in A').$$

Puesto que la segunda igualdad se cumple para todo $A' \in \mathcal{B}^\infty$, tenemos que

$$\mathbb{P}([X(T) \in A']\Delta[X(T) \in B_n\}) = \mathbb{P}([X \in A']\Delta[X \in B_n]). \tag{1.2.3}$$

Sea $T$ la permutación que manda $(1, 2, \cdots, 2n)$ en $(2n, 2n-1, \cdots, 1)$ y $T(k) = k$ para todo $k > 2n$, y sea

$$A(T)_n := [X(T) \in B_n] = [(X_{2n}, X_{2n-1}, \cdots, X_{n+1}) \in B_n].$$

Los eventos $A(T)_n$ y $A_n$ tienen la misma distribución. Por (1.2.3), y por simetría, con $A'$ tal que $A = [X \in A'] = [X(T) \in A']$,

$$\mathbb{P}(A\Delta A(T)_n) = \mathbb{P}(A\Delta A_n) \to 0, \tag{1.2.4}$$

de donde semejante a lo expuesto en (1.2.2)

$$\mathbb{P}(A_n) \to \mathbb{P}(A), \qquad \text{y} \quad \mathbb{P}(A(T)_n) \to \mathbb{P}(A) \text{ cuando } n \to \infty. \tag{1.2.5}$$

Ahora bien, como $A_n\Delta A(T)_n \subset (A\Delta A(T)_n) \cup (A\Delta A_n)$, entonces $\mathbb{P}(A_n\Delta A(T)_n) \to 0$ por (1.2.4), y puesto que ambas $\mathbb{P}(A_n)$ y $\mathbb{P}(A(T)_n)$ convergen a $\mathbb{P}(A)$, se sigue que

$$\begin{aligned}
\lim \mathbb{P}(A_n \Delta A(T)_n) &= \lim[\mathbb{P}(A_n) + \mathbb{P}(A(T)_n) - 2\mathbb{P}(A_n \cap A(T)_n)] \\
&= 2\mathbb{P}(A) - 2\lim \mathbb{P}(A_n \cap A(T)_n).
\end{aligned}$$

Luego $\mathbb{P}(A_n \cap A(T)_n) \to \mathbb{P}(A)$, pero $A_n$ y $A(T)_n$ son independientes, de modo que, usando (1.2.5),

$$\mathbb{P}(A_n \cap A(T)_n) = \mathbb{P}(A_n)\mathbb{P}(A(T)_n) \to \mathbb{P}(A)^2,$$

por lo que $\mathbb{P}(A) = \mathbb{P}(A)^2$, i.e., $\mathbb{P}(A) = 0$ ó $\mathbb{P}(A) = 1$. ■

## 1.3. Convergencia de series de variables aleatorias

En las leyes de los grandes números se estudia la convergencia de ciertos tipos de sucesiones de vv.aa. Ahora deseamos ahondar un poco más en el estudio de la convergencia de vv.aa. independientes. El resultado fuerte de este capítulo es el célebre teorema de *las tres series* debido a Kolmogorov, el cual proporciona condiciones necesarias y suficientes de convergencia. Antes proporcionamos el siguiente lema, al que conviene comparar con la desigualdad de Kolmogorov.

**Lema 1.3.1.** *Sean $X_1, \cdots, X_n$ vv.aa. independientes con media cero y tales que $|X_j| \leq C$ c.p.1., para todo $j = 1, \cdots, n$. Entonces para cualquier $\varepsilon > 0$,*

$$\mathbb{P}(\max_{1\leq j\leq n} |S_j| > \varepsilon) \geq 1 - \frac{(\varepsilon + C)^2}{\mathrm{Var}(S_n)}. \tag{1.3.1}$$

*Por cierto, note que la condición $|X_j| \leq C$ implica que* $\mathrm{Var}(X_j) < \infty$.

**Dem.**
La demostración es igual a la dada en la desigualdad de Kolmogorov.
Definimos los eventos

$$A := [\max_{1\leq j\leq n} |S_j| > \varepsilon],$$

y

$$A_k := [|S_j| \leq \varepsilon, \quad \text{para} \quad 0 \leq j \leq k-1, \quad \text{y} \quad |S_k| > \varepsilon], \quad \forall k = 1, \cdots, n,$$

donde previamente hemos definido $S_0 = 0$. Los eventos $A_k$ son ajenos a pares y claramente $A = \cup_{k=1}^n A_k$. Entonces vale

$$\begin{aligned}\mathbb{E}(S_n^2 I_A) &= \sum_{k=1}^n \mathbb{E}(S_n^2 I_{A_k}) \\ &= \sum_{k=1}^n \mathbb{E}\{(S_k + (S_n - S_k))^2 I_{A_k}\} \\ &= \sum_{k=1}^n \mathbb{E}(S_k^2 I_{A_k}) + \sum_{k=1}^n \mathbb{E}\{(S_n - S_k)^2 I_{A_k}\} + 2\sum_{k=1}^n \mathbb{E}\{S_k I_{A_k}(S_n - S_k)\},\end{aligned}$$

en donde por las hipótesis del lema, el último término es cero. Ahora bien, sobre $A_k$, se tiene que $|S_{k-1}| \leq \varepsilon$ y, por lo tanto, $|S_k| \leq |S_{k-1}| + |X_k| \leq \varepsilon + C$ c.p.1. en $A_k$. Luego, puesto que $I_{A_k}$ y $S_n - S_k$ son independientes,

$$\begin{aligned}\mathbb{E}(S_n^2 I_A) &\leq (\varepsilon + C)^2 \sum_{k=1}^n \mathbb{P}(A_k) + \sum_{k=1}^n \mathbb{P}(A_k) \sum_{j=k+1}^n \mathrm{Var}(X_j) \\ &\leq ((\varepsilon + C)^2 + \mathrm{Var}(S_n))\mathbb{P}(A), \qquad (1.3.2)\end{aligned}$$

pues $\sum_{j=k+1}^n \mathrm{Var}(X_j) \leq \sum_{j=1}^n \mathrm{Var}(X_j) = \mathrm{Var}(S_n)$.

Por otro lado,

$$\begin{aligned}\mathbb{E}(S_n^2 I_A) &= \mathbb{E}(S_n)^2 - \mathbb{E}(S_n^2 I_{A^c} \\ &\geq \mathrm{Var}(S_n) - \varepsilon^2 \mathbb{P}(A^c) \\ &= \mathrm{Var}(S_n) - \varepsilon^2 + \varepsilon^2 \mathbb{P}(A). \qquad (1.3.3)\end{aligned}$$

Finamelnte, comparando (1.3.2) y (1.3.3), obtenemos que

$$\mathbb{P}(A)((\varepsilon + C)^2 + \mathrm{Var}(S_n) - \varepsilon^2) \geq \mathrm{Var}(S_n) - \varepsilon^2$$

de donde

$$\mathbb{P}(A) \geq 1 - \frac{(\varepsilon + C)^2}{(\varepsilon + C)^2 + \mathrm{Var}(S_n) - \varepsilon^2} \geq 1 - \frac{(\varepsilon + C)^2}{(\mathrm{Var}(S_n)}$$

lo cual implica la desigualdad deseada. ∎

**Observación 1.3.2.** *En un teorema del curso pasado vimos que si $X_1, X_2, \cdots$, son vv.aa. independientes en $L^2$ tales que $\sum_{n=1}^\infty \mathrm{Var}(X_n)$ converge, entonces la serie $\sum_{n=1}^\infty (X_n - \mathbb{E}X_n)$ converge c.p.1. Este resultado se puede completar como sigue:*

**Teorema 1.3.3.** *Sean $X_1, X_2, \cdots$, vv.aa. independientes con* <u>*media cero*</u> *y uniformemente acotadas, es decir, tales que existe una constante $C$ con $|X_i| \leq C$ c.p.1. para todo $i = 1, 2, \cdots$. Entonces la serie $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)$ converge si y sólo si $\sum_{n=1}^{\infty} X_n$ converge c.p.1.*

**Dem.** La necesidad se sigue del teorema mencionado en la observación precedente 1.3.2.

Ahora supongamos que $\sum_{n=1}^{\infty} X_n$ converge c.p.1., y sea $S_n = X_1 + X_2 + \cdots + X_n$, $n = 1, 2, \cdots$. Entonces la sucesión $\{S_n\}$ es de Cauchy c.p.1., i.e.

$$\max_{i \geq 0} |S_{N+i} - S_N| \to 0 \quad \text{c.p.1. cuando } N \to \infty;$$

en particular, se tiene la convergencia en probabilidad (recordemos que convergencia c.p.1. implica convergencia en probabilidad): para cualquier $\varepsilon > 0$,

$$\lim_{N \to \infty} \mathbb{P}(\max_{i \geq 0} |S_{N+i} - S_N| > \varepsilon) = 0.$$

Fijemos $N$ tal que ésta probabilidad es menor o igual a $1/2$. Por el lema 1.3.1, para cada $n \in \mathbb{N}$,

$$\mathbb{P}(\max_{i \geq 0} |S_{N+i} - S_N| > \varepsilon) \geq \mathbb{P}(\max_{0 \leq i \leq n} |S_{N+i} - S_N| > \varepsilon) \geq 1 - \frac{(\varepsilon + C)^2}{\sum_{i=N+1}^{N+n} \mathrm{Var}(X_i)}.$$

Si suponemos que la serie de las varianzas no converge, entonces el lado derecho de la última desigualdad tiende a 1 cuando $n \to \infty$, lo cual constituye una contradicción. ■

**Ejemplo 1.3.4. Continuación del problema de los signos aleatorios** En el curso precedente, vimos que si $\sum_{n=1}^{\infty} a_n^2 < \infty$, entonces la serie con signos aleatorios $\sum_{n=1}^{\infty} a_n Y_n$ converge c.p.1., en donde $Y_1, Y_2, \cdots$, son variables i.i.d. con

$$\mathbb{P}(Y_n = 1) = \mathbb{P}(Y_n = -1) = \frac{1}{2}.$$

Recíprocamente, si $\sum_{n=1}^{\infty} a_n Y_n$ converge c.p.1., entonces $a_n \to 0$, y por lo tanto, la sucesión $\{a_n\}$ es acotada. Luego, por el teorema 1.3.3, se sigue inmediatamente que $\sum_{n=1}^{\infty} \mathrm{Var}(a_n Y_n) = \sum_{n=1}^{\infty} a_n^2$ converge.

En el siguiente teorema eliminaremos la hipótesis del teorema 1.3.3, de que las variables tienen media cero.

**Teorema 1.3.5.** *Sean* $X_1, X_2, \cdots$ *, vv.aa. independientes uniformemente acotadas, digamos* $|X_i| \leq C$ *c.p.1., para* $i = 1, 2, \cdots$*. Entonces la serie* $\sum_{k=1}^{\infty} X_k$ *converge c.p.1. si y sólo si las series* $\sum_{k=1}^{\infty} \mathrm{Var}(X_k)$ *y* $\sum_{k=1}^{\infty} \mathbb{E}X_k$ *convergen (en* $\mathbb{R}$*).*

**Dem.** (**Suficiencia**.) Puesto que $X_k - \mathbb{E}X_k$ son variables independientes con media cero y uniformemente acotadas, se sigue ( de la observación 1.3.2 ó de ) teorema 1.3.3 que la serie $\sum_{k=1}^{\infty}(X_k - \mathbb{E}X_k)$ converge c.p.1., pues $\sum_{k=1}^{\infty} \mathrm{Var}(X_k)$ converge. Luego, si $\sum_{k=1}^{\infty} \mathbb{E}X_k$ converge, concluimos que $\sum_{k=1}^{\infty} X_k$ converge c.p.1.

(**Necesidad**.) Supongamos ahora que $\sum_{k=1}^{\infty} X_k$ converge c.p.1., y consideramos una nueva sucesión $\{Y_k\}$ de vv.aa. (un teorema debido a Kolmogorov garantiza su existencia), tales que $X_1, Y_1, X_2, Y_2, \cdots$ , son independientes y para cada $k$, $Y_k$ tiene la misma distribución que $X_k$. Entonces las variables $X_k - Y_k$ tienen media cero y variancias

$$\mathrm{Var}(X_k - Y_k) = \mathrm{Var}(X_k) + \mathrm{Var}(Y_k) = 2\mathrm{Var}(X_k).$$

Puesto que $\sum_{k=1}^{\infty} X_k$ converge c.p.1., la serie $\sum_{k=1}^{\infty} Y_k$ también converge c.p.1. y por lo tanto $\sum_{k=1}^{\infty}(X_k - Y_k)$ converge c.p.1. Luego, del teorema 1.3.3 se sigue que

$$\sum_{k=1}^{\infty} \mathrm{Var}(X_k) = \frac{1}{2}\sum_{k=1}^{\infty} \mathrm{Var}(X_k - Y_k) < \infty.$$

De nuevo por el teorema 1.3.3 aplicado a la serie de vv.aa. con media cero de término general $X_k - \mathbb{E}X_k$, se tiene necesariamente que $\sum_{k=1}^{\infty}(X_k - \mathbb{E}X_k)$ converge c.p.1., y en consecuencia $\sum_{k=1}^{\infty} \mathbb{E}X_k$ converge. ∎

## 1.4. Teorema de las tres series de Kolmogorov

Recordemos los lemas de Borel-Cantelli.

**Lema 1.4.1.** *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad,* $\{A_n\}$ *sucesión de eventos.*

*(i)* <u>*Primer lema de Borel-Cantelli*</u>*. Si* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$*, entonces*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

(ii) *Segundo lema de Borel-Cantelli. Si $\{A_n\}$ son eventos independientes y $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, entonces*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 1.$$

(ii') *El segundo lema de Borel-Cantelli equivale: Si $\{A_n\}$ son eventos independientes y $\mathbb{P}(\limsup_{n\to\infty} A_n) < 1$, entonces $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.*

Proporcionamos el siguiente criterio general de convergencia de series de vv.aa. independientes.

**Teorema 1.4.2.** (**Teorema de las tres series de Kolmogorov**). Sean $X_1, X_2, \cdots$, vv.aa. independientes. Si $M > 0$ definimos las variables truncadas

$$\begin{aligned} \overline{X}_j &= X_j \quad \text{si } |X_j| \le M \\ &= 0 \quad \text{si } |X_j| > M. \end{aligned}$$

(a) Si $\sum_{j=1}^{\infty} X_j$ converge c.p.1., entonces para cualquier $M > 0$, las tres series

$$\sum_{j=1}^{\infty} \mathbb{P}(X_j \ne \overline{X}_j), \quad \sum_{j=1}^{\infty} \mathrm{Var}(\overline{X}_j), \quad \text{y} \quad \sum_{j=1}^{\infty} \mathbb{E}\overline{X}_j$$

convergen (en $\mathbb{R}$).

(b) Si para algún $M > 0$ , estas tres series convergen, entonces $\sum_{j=1}^{\infty} X_j$ converge c.p.1.

**Dem.** (a) Si $\sum_{j=1}^{\infty} X_j$ converge c.p.1., entonces $X_j \to 0$ c.p.1. cuando $j \to \infty$. Por lo tanto, para cualquier $M > 0$, vale

$$[X_j \to 0] \subset \liminf_{j\to\infty}[|X_j| \le M] = \cup_{N=1} \cap_{j=N}^{\infty} [|X_j| \le M],$$

lo que implica que $\mathbb{P}(\liminf_{j\to\infty}[|X_j| \le M]) = 1$, equivalentemente, como $[|X_j| > M] = [X_j \ne \overline{X}_j]$,

$$\mathbb{P}(\limsup_{j\to\infty}[|X_j| > M]) = \mathbb{P}(\limsup_{j\to\infty}[X_j \ne \overline{X}_j]) = 0. \tag{1.4.1}$$

Luego, por el segundo lema de Borel-Cantelli, $\sum_{j=1}^{\infty} \mathbb{P}(X_j \ne \overline{X}_j) < \infty$. La convergencia de las otras dos series se sigue del teorema 1.3.5.

(b) Si las series $\sum_{j=1}^{\infty} \mathrm{Var}(\overline{X}_j)$ y $\sum_{j=1}^{\infty} \mathbb{E}\overline{X}_j$ convergen para algún $M > 0$, entonces una vez más por el teorema 1.3.5 se tiene que $\sum_{j=1}^{\infty} \overline{X}_j$ converge c.p.1. Ahora, por hipótesis, $\sum_{j=1}^{\infty} \mathbb{P}(X_j \neq \overline{X}_j) < \infty$. Por el primer lema de Borel-Cantelli, se tiene que vale la relación (1.4.1), o equivalentemente $\mathbb{P}(\liminf_{j\to\infty}[|X_j| \leq M]) = 1$. Así, dado $\omega \in \Omega' := \liminf_{j\to\infty}[|X_j| \leq M] \cap [\sum_{j=1}^{\infty} \overline{X}_j \quad \text{converge}]$ (note que $\mathbb{P}(\Omega') = 1$) existe $N = N(\omega) \in \mathbb{N}$ tal que

$$|X_j(\omega)| \leq M \quad \forall j \geq N,$$

o sea

$$\overline{X}_j(\omega) = X_j(\omega) \quad \forall j \geq N,$$

de donde

$$\sum_{j=1}^{\infty} X_j(\omega) = \sum_{j=1}^{N-1} X_j(\omega) + \sum_{j=N}^{\infty} X_j(\omega) = \sum_{j=1}^{N-1} X_j(\omega) + \sum_{j=N}^{\infty} \overline{X}_j(\omega) \in \mathbb{R}.$$

Por lo tanto, $\sum_{j=1}^{\infty} X_j$ converge c.p.1. ■

## 1.5. Ejercicios

1. Si $X$ es una v.a. no-negativa, pruebe que

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \leq \mathbb{E}X \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

   De aquí pruebe que $\mathbb{E}X < \infty$ si y sólo si $\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) < \infty$.

2. Sean $X_1, X_2, \cdots$ , v.a. i.i.d. de Bernoulli, con

$$\begin{aligned} \mathbb{P}(X_1 = k) &= p && \text{si } k = 1 \\ &= 1 - p && \text{si } k = -1. \quad 0 \leq p \leq 1, p \neq 1/2. \end{aligned}$$

   Demuestre que $\mathbb{P}(S_n \neq 0 \text{ para casi todo } n) = 1$, en donde $S_n = X_1 + X_2 + \cdots + X_n$.

3. Sean $X_1, X_2, \cdots$ v.a. i.i.d. con media $\mu$ finita. Demuestre que para cualquier función $f : \mathbb{R} \to \mathbb{R}$, $f \in C(\mathbb{R})$ se tiene

$$f(S_n/n) \to f(\mu) \qquad \text{c.p.1. y en } L^1.$$

4. Sean $X_1, X_2, \cdots$ v.a. i.i.d. con media cero y varianza $\sigma^2$ finita. Demuestre que

$$\frac{S_n}{\sqrt{n}\log(n)} \to 0 \qquad \text{c.p.1.}$$

5. Sea $X_1, X_2, \cdots$ una sucesión de v.a. Pruebe que es de Cauchy c.p.1. si y sólo si

$$\lim_{m \to \infty} \mathbb{P}(\cup_{j,k=m}^{\infty} [|X_j - X_k| \geq \delta]) = 0 \quad \forall \delta > 0.$$

# Capítulo 2

# CONVERGENCIA EN DISTRIBUCIÓN

En el curso precedente y en el capítulo precedente consideramos varios tipos de convergencia de sucesiones de vv.aa. Ahora estudiaremos el concepto de *convergencia en distribución*, el cual es uno de las convergencias más importantes en la teoría de probabilidad. Muchos de los resultados de mayor interés en probabilidad y estadística matemática, así como en sus aplicaciones se refieren a este tipo de convergencia. Este capítulo esta basado en la bibliografía básica sobre el tema (ver, por ejemplo, [1, 3, 11, 12]).

## 2.1. Convergencia débil de medidas de probabilidad

En esta sección supondremos que nuestro espacio medible $(\Omega, \mathcal{F})$ consiste en un espacio métrico $\Omega$, con métrica $\rho$, y $\mathcal{F} = \mathcal{B}(\Omega)$ es la $\sigma$-álgebra de los conjuntos de Borel de $\Omega$, es decir, $\mathcal{F}$ es la $\sigma$-álgebra generada por los conjuntos abiertos (o equivalentemente, los conjuntos cerrados) de $\Omega$. Denotaremos por $C_b(\Omega)$ el espacio vectorial normado de todas las funciones continuas acotadas $f : \Omega \to \mathbb{R}$, con la norma $\|f\| := \sup_{x \in \Omega} |f(x)|$.

**Definición 2.1.1.** *Sean* $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \cdots$ *medidas de probabilidad sobre* $(\Omega, \mathcal{F})$*. Decimos que* $\{\mathbb{P}_n\}$ converge débilmente *a* $\mathbb{P}$ *( en cuyo caso escribimos*

$\mathbb{P}_n \xrightarrow{d} \mathbb{P}$) *si se satisface que*

$$\int_\Omega f \; d\mathbb{P}_n \to \int_\Omega f \; d\mathbb{P} \qquad \text{cuando } n \to \infty \tag{2.1.1}$$

*para toda función* $f \in C_b(\Omega)$.

En el teorema siguiente caracterizamos la convergencia débil de medidas de probabilidad en términos de convergencia sobre conjuntos, para el cual introducimos el concepto de conjunto $\mathbb{P}$-continuo:

**Definición 2.1.2.** *Decimos que* $A \in \mathcal{F}$ *es un conjunto* $\mathbb{P}$-continuo *si* $\mathbb{P}(\mathrm{Fr}A) = 0$, *donde* $\mathrm{Fr}A$ *es la frontera del conjunto* $A$.

El siguiente teorema proporciona condiciones equivalentes fundamentales para la convergencia débil de medidas de probabilidad.

**Teorema 2.1.3.** (**Teorema de Portmanteau.**) Las siguientes proposiciones son equivalentes:

(a) $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$;

(b) $\limsup_{n\to\infty} \mathbb{P}_n(F) \leq \mathbb{P}(F)$ para todo conjunto cerrado $F$ de $\Omega$;

(c) $\liminf_{n\to\infty} \mathbb{P}_n(G) \geq \mathbb{P}(G)$ para todo conjunto abierto $G$ de $\Omega$;

(d) $\lim_{n\to\infty} \mathbb{P}_n(A) = \mathbb{P}(A)$ para todo conjuto Borel $\mathbb{P}$-continuo $A$ de $\Omega$.

**Dem.** (a) $\Longrightarrow$ (b): Sea $F$ un conjunto cerrado y denotamos por $F^\delta$, con $\delta > 0$, a la $\overline{\delta\text{-vecindad}}$ de $F$, a saber, el conjunto abierto que contiene a $F$,

$$F^\delta := \{x \in \Omega : \rho(x, F) < \delta\}.$$

Sea $\epsilon > 0$ arbitrario. Puesto que $F^\delta \downarrow F$ cuando $\delta \downarrow 0$, existe $\delta > 0$ tal que

$$\mathbb{P}(F^\delta) < \mathbb{P}(F) + \epsilon.$$

Consideremos ahora la función $\phi : \mathbb{R} \to \mathbb{R}$ definido como

$$\begin{aligned} \phi(t) &= 1 && \text{si} \quad t \leq 0, \\ &= 1 - t && \text{si} \quad 0 \leq t \leq 1, \\ &= 0 && \text{si} \quad t \geq 1, \end{aligned}$$

y definamos $f : \Omega \to \mathbb{R}$ por $f(x) := \phi(\rho(x, F)/\delta)$. Entonces $f$ es continua y acotada, i.e., $f \in C_b(\Omega)$. Además,

(i) $f(x) = 1 \iff \rho(x, F) = 0 \iff x \in F$;

(ii) $f(x) = 0 \iff \rho(x, F) \geq \delta \iff x \notin F^\delta$;

También vale la implicación:

(iii) $x \in F^\delta - F \Longrightarrow f(x) = 1 - \rho(x, F)/\delta$.

En todo caso, de (i),(ii) y (iii), se tiene que $0 \leq f(x) \leq 1$ para todo $x \in \Omega$. Además, de (i), si $I_F$ es la función indicadora del cerrado $F$, vemos que $I_F(x) \leq f(x)$ para todo $x \in \Omega$. Luego

$$\mathbb{P}_n(F) \leq \int_\Omega f \, d\mathbb{P}_n \quad \forall n \in \mathbb{N}.$$

Por otro lado, claramente de (ii), $f(x) \leq I_{F^\delta}(x)$ para cada $x \in \Omega$, lo que implica

$$\int_\Omega f \, d\mathbb{P} = \int_{F^\delta} d\mathbb{P} \leq \mathbb{P}(F^\delta) < \mathbb{P}(F) + \epsilon.$$

Finalmente, puesto que $f \in C_b(\Omega)$ y $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$, se sigue que

$$\limsup_{n\to\infty} \mathbb{P}_n(F) \leq \lim_{n\to\infty} \int_\Omega f \, d\mathbb{P}_n = \int_\Omega f \, d\mathbb{P} < \mathbb{P}(F) + \epsilon.$$

Puesto que $\epsilon > 0$ es arbitrario, la última desigualdad prueba (b).

<u>(b) $\iff$ (c)</u>: Si (b) se cumple, entonces para cualquier abierto $G$,

$$\begin{aligned} \mathbb{P}(G) &= 1 - \mathbb{P}(G^c) \leq 1 - \limsup_{n\to\infty} \mathbb{P}_n(G^c) \\ &= \liminf_{n\to\infty} \mathbb{P}_n(G), \end{aligned}$$

lo cual prueba (c). Un argumento similar prueba que (c) implica (b).

<u>(c) $\Longrightarrow$ (d)</u>: Para cualquier $A \in \mathcal{F}$,

$$\begin{aligned} \mathbb{P}(\overline{A}) &\geq \limsup_{n\to\infty} \mathbb{P}_n(\overline{A}) \\ &\geq \limsup_{n\to\infty} \mathbb{P}_n(A) \\ &\geq \liminf_{n\to\infty} \mathbb{P}_n(A) \\ &\geq \liminf_{n\to\infty} \mathbb{P}_n(A^\circ) \\ &\geq \mathbb{P}(A^\circ), \end{aligned}$$

en donde $\overline{A}$ y $A^\circ$ denotan la adherencia y el interior de $A$, respectivamente. Por último, si $A$ es un conjunto $\mathbb{P}$-continuo , entonces

$$0 = \mathbb{P}(\mathrm{Fr}A) = \mathbb{P}(\overline{A} - A^\circ) = \mathbb{P}(\overline{A}) - \mathbb{P}(A^\circ),$$

y por lo tanto, de las desigualdades precedentes se sigue (d).

<u>(d) $\Longrightarrow$ (a)</u>: Supongamos ahora que (d) se cumple; deseamos probar (2.1.1) para cualquier $f \in C_b(\Omega)$.

Sea $f \in C_b(\Omega)$, y sean $\alpha, \beta$ números reales tales que $\alpha < f(x) < \beta$ para todo $x \in \Omega$. El conjunto de todos los $\gamma$ tales que $\mathbb{P}(x : \ f(x) = \gamma) > 0$ es a lo más numerable. Dado $\epsilon > 0$, tómense números $\alpha_i$ tales que $\alpha = \alpha_0 < \alpha_1 < \cdots < \alpha_k = \beta$, con

$$\alpha_i - \alpha_{i-1} < \epsilon \qquad \text{y} \qquad \mathbb{P}(x : \ f(x) = \alpha_i) = 0.$$

Si $z$ es un punto frontera del conjunto del conjunto

$$A_i := \{x \in \Omega : \alpha_{i-1} < f(x) \leq \alpha_i\},$$

entonces $f(z) = \alpha_i$ ó $f(z) = \alpha_{i-1}$. En efecto, $f^{-1}((\alpha_{i-1}, \alpha_i)) \subset A_i \subset f^{-1}([\alpha_{i-1}, \alpha_i])$, de donde $\mathrm{Fr}A_i = \overline{A_i} - A_i^\circ \subset f^{-1}([\alpha_{i-1}, \alpha_i] - (\alpha_{i-1}, \alpha_i)) = f^{-1}(\{\alpha_{i-1}\}) \cup f^{-1}(\{\alpha_i\})$, y por lo tanto, $\mathbb{P}(\mathrm{Fr}A_i) \leq \mathbb{P}(x : \ f(x) = \alpha_{i-1}) + \mathbb{P}(x : f(x) = \alpha_i) = 0$, por lo tanto, $\mathbb{P}(\mathrm{Fr}A_i) = 0$, de donde $A_i$ es $\mathbb{P}$-continuo para todo $i = 1, \cdots, k$. Además, los $A_i$ son eventos ajenos a pares tales que $\Omega = \cup_{i=1}^k A_i$, y se tiene que

$$\sum_{i=1}^k \alpha_{i-1} \mathbb{P}_n(A_i) \leq \int_\Omega f \ d\mathbb{P}_n \leq \sum_{i=1}^k \alpha_i \mathbb{P}_n(A_i)$$

así como

$$\sum_{i=1}^k \alpha_{i-1} \mathbb{P}(A_i) \leq \int_\Omega f \ d\mathbb{P} \leq \sum_{i=1}^k \alpha_i \mathbb{P}(A_i).$$

Combinando estas desigualdades y usando el hecho de que $\mathbb{P}_n(A_i) \to \mathbb{P}(A_i)$ cuando $n \to \infty$, vemos que

$$-\sum_{i=1}^k (\alpha_i - \alpha_{i-1}) \mathbb{P}(A_i) \leq \limsup_{n\to\infty} \int_\Omega f \ d\mathbb{P}_n - \int_\Omega f \ d\mathbb{P} \leq \sum_{i=1}^k (\alpha_i - \alpha_{i-1}) \mathbb{P}(A_i),$$

así como

$$-\sum_{i=1}^{k}(\alpha_i-\alpha_{i-1})\mathbb{P}(A_i)\le \liminf_{n\to\infty}\int_\Omega f\ d\mathbb{P}_n-\int_\Omega f\ d\mathbb{P}\le\sum_{i=1}^{k}(\alpha_i-\alpha_{i-1})\mathbb{P}(A_i),$$

de manera que

$$-\epsilon\le\limsup_{n\to\infty}\int_\Omega f\ d\mathbb{P}_n-\int_\Omega f\ d\mathbb{P}\le\epsilon;$$

y

$$-\epsilon\le\liminf_{n\to\infty}\int_\Omega f\ d\mathbb{P}_n-\int_\Omega f\ d\mathbb{P}\le\epsilon;$$

de aquí se sigue inmediatamente (a). ∎

**Ejemplo 2.1.4.** *Sean $x, x_1, x_2, \cdots$, puntos en el espacio métrico $\Omega$, y $\delta_x, \delta_{x_1}, \delta_{x_2}, \cdots$ las correspondientes medidas de Dirac: para $B \in \mathcal{B}(\Omega)$,*

$$\begin{aligned}\delta_x(B) &:= 1 \quad si \quad x\in B,\\ &= 0 \quad si \quad x\notin B.\end{aligned}$$

*Entonces $\delta_{x_n}\xrightarrow{d}\delta_x$ si $x_n\to x$ en $(\Omega,\rho)$, porque*

$$\int_\Omega f\ d\delta_{x_n}=f(x_n)\to\int_\Omega f\ d\delta_x=f(x),$$

*para todo $f\in C_b(\Omega)$. Por el teorema de Portmanteau 2.1.3, $\delta_{x_n}(A)\to\delta_x(A)$ para todo $A\in\mathcal{B}(\Omega)$ boreliano $\mathbb{P}$-continuo. Sin embargo, si $A\in\mathcal{B}(\Omega)$ es arbitrario, entonces $\delta_{x_n}\xrightarrow{d}\delta_x$ no implica necesariamente que $\delta_{x_n}(A)\to\delta_x(A)$.*

*En efecto, considere $\Omega=\mathbb{R}$, $x_n\downarrow x$, $A=(-\infty,x]$, entonces $\delta_{x_n}(A)=0\nrightarrow\delta_x(A)=1$. Esto también ilustra que $\delta_{x_n}\xrightarrow{d}\delta_x$ no implica que $\int_\Omega f\ d\delta_{x_n}=f(x_n)\to\int_\Omega f\ d\delta_x=f(x)$ si $f$ no es continua aunque $f$ sea medible y acotada, por ejemplo tome $f=I_A$ con $A=(-\infty,x]$.*

## 2.2. Unicidad del límite en la convergencia débil

Consideremos el siguiente resultado, el cual es una consecuencia del teorema de las clases de Dynkin (ver, por ejemplo, [4, teorema 1.6.2]).

**Teorema 2.2.1.** *Sean $(\Omega, \mathcal{F})$ un espacio medible, $\mathcal{C}$ una subcolección de $\Omega$ tal que satisface las siguientes condiciones*

*(a) Si $A, B \in \mathcal{C}$ entonces $A \cap B \in \mathcal{C}$;*

*(b) $\mathcal{F} = \sigma(\mathcal{C})$.*

*Si $\mathbb{P}, \mathbb{Q}$ son dos medidas de probabilidad tal que $\mathbb{P}(C) = \mathbb{Q}(C)$ para todo $C \in \mathcal{C}$. Entonces $\mathbb{P}(F) = \mathbb{Q}(F)$ para todo $F \in \mathcal{F}$.*

**Corolario 2.2.1.** *Supongamos que $(\Omega, \rho)$ es un espacio métrico, $\mathcal{F} := \mathcal{B}(\Omega)$ la $\sigma$-álgebra de Borel de $\Omega$, $\mathbb{P}$, $\mathbb{Q}$ dos medidas de probabilidad sobre $(\Omega, \mathcal{F})$. Consideremos además $\mathcal{C}$ la colección de todos los abiertos de $\Omega$ ó la colección de todos los cerrados de $\Omega$. Entonces:*

*(a) Si $\mathbb{P}(F) = \mathbb{Q}(F)$ para todo cerrado $F$ de $\Omega$, entonces $\mathbb{P} = \mathbb{Q}$ sobre $\mathcal{F} = \mathcal{B}(\Omega)$.*

*(b) Si $\mathbb{P}(G) = \mathbb{Q}(G)$ para todo abierto $G$ de $\Omega$, entonces $\mathbb{P} = \mathbb{Q}$ sobre $\mathcal{F} = \mathcal{B}(\Omega)$.*

**Observación 2.2.2.** *El corolario precedente afirma que toda medida de probabilidad $\mathbb{P}$ sobre un espacio métrico $(\Omega, \rho)$ está determinada por sus valores $\mathbb{P}(F)$ sobre conjuntos cerrados $F$ ( o de $\mathbb{P}(G)$ sobre los abiertos $G$). En el siguiente teorema se asegura que $\mathbb{P}$ también queda completamente determinada por sus valores integrales $\int_\Omega f \; d\mathbb{P}$ para todo $f \in C_b(\Omega)$.*

**Teorema 2.2.3.** *Sea $(\Omega, \rho)$ un espacio métrico, $\mathbb{P}, \mathbb{Q}$ dos probabilidades sobre $(\Omega, \mathcal{F})$, con $\mathcal{F} = \mathcal{B}(\Omega)$ la $\sigma$-álgebra de Borel de $\Omega$. Supongamos que vale*

$$\int_\Omega f \; d\mathbb{P} = \int_\Omega f \; d\mathbb{Q} \qquad \forall f \in C_b(\Omega),$$

*entonces $\mathbb{P} = \mathbb{Q}$ sobre $\mathcal{F}$.*

**Dem.** Por el corolario 2.2.1, basta probar que $\mathbb{P}(F) = \mathbb{Q}(F)$ para cualquier cerrado $F$ de $\Omega$. Para ello consideramos la función $\phi : \mathbb{R} \to \mathbb{R}$ definida en la demostración del teorema 2.1.3:

$$\begin{aligned} \phi(t) &= 1 \quad &\text{si} \quad & t \leq 0, \\ &= 1 - t \quad &\text{si} \quad & 0 \leq t \leq 1, \\ &= 0 \quad &\text{si} \quad & t \geq 1, \end{aligned}$$

y para cada $n \in \mathbb{N}$ definamos $f_n : \Omega \to \mathbb{R}$

$$f_n(x) := \phi(n\rho(x, F)) \quad \forall x \in \Omega.$$

Notemos que $\{f_n\}$ es una sucesión decreciente de funciones continuas y acotadas en $C_b(\Omega)$ tal que

$$\lim_{n\to\infty} f_n(x) = I_F(x), \qquad \text{y} \quad 0 \leq f_n(x) \leq 1 \qquad \forall x \in \Omega.$$

Por el teorema de la convergencia dominada de Lebesgue se sigue de aquí que

$$\mathbb{P}(F) = \lim_{n\to\infty} \int_\Omega f_n \, d\mathbb{P} = \lim_{n\to\infty} \int_\Omega f_n \, d\mathbb{Q} = \mathbb{Q}(F).$$

■

**Corolario 2.2.2.** *Sean $(\Omega, \rho)$ espacio métrico, $\mathbb{P}, \mathbb{Q}, \mathbb{P}_1, \mathbb{P}_2, \cdots$ medidas de probabilidad sobre $(\Omega, \mathcal{B}(\Omega))$. Si $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$, y $\mathbb{P}_n \xrightarrow{d} \mathbb{Q}$, entoces $\mathbb{P} = \mathbb{Q}$. En otras palabras, si $\{\mathbb{P}_n\}$ converge débilmente, el límite es único.*

## 2.3. Convergencia débil de funciones de distribución de probabilidad

Las definiciones y resultados anteriores son válidos para medidas finitas (no necesariamente medidas de probabilidad) arbitrarias sobre un espacio Borel medible, es decir, un espacio métrico provisto de la $\sigma$-álgebra de Borel.

Sin embargo, el caso más usual que tienen amplia aplicación a la teoría de probabilidad, es suponer que $\{\mathbb{P}_n\}$, $\mathbb{P}$ son medidas de probabilidad sobre la recta real $\mathbb{R}$ provista de la $\sigma$-álgebra de Borel. Veremos en esta sección la importancia de asociar a cada $\mathbb{P}_n$ y $\mathbb{P}$ sus funciones de distribución de probabilidad (f.d.p.),

$$F_n(x) = \mathbb{P}((-\infty, x]), \quad F(x) = \mathbb{P}((-\infty, x]), \qquad \forall x \in \mathbb{R}.$$

En lo que sigue, supondremos que las medidas de probilidad están definidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

**Notación**. Si $F : \mathbb{R} \to \mathbb{R}$ es un f.d.p. la cual es continua en un punto $x \in \mathbb{R}$, escribimos $x \in \mathrm{C}(F)$, donde $\mathrm{C}(F)$ es el subconjunto de $\mathbb{R}$ formado por todos los puntos de continuidad de la f.d.p. $F$.

**Definición 2.3.1.** *Sean $F, F_1, F_2, \cdots$ f.d.p.: $\mathbb{R} \to \mathbb{R}$. Diremos que $\{F_n\}$* converge débilmente *a $F$ ( lo que denotamos en símbolos como $F_n \xrightarrow{d} F$ ) si vale*

$$F_n(x) \to F(x) \qquad \forall x \in C(F).$$

**Teorema 2.3.2.** *Sean $\mathbb{P}, \mathbb{P}_n$, $n = 1, 2, \cdots$, probabilidades sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, y sean $F, F_n$ las correspondientes f.d.p. Entonces $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$ si y sólo si $F_n \xrightarrow{d} F$.*

**Dem. Necesidad**. Sea $x \in \mathrm{C}(F)$ un punto de continuidad de la f.d.p. $F$. Entonces $\mathbb{P}(\mathrm{Fr}(-\infty, x]) = \mathbb{P}(\{x\}) = F(x+) - F(x-) = 0$, es decir, el intervalo $(-\infty, x]$ es $\mathbb{P}$-continuo. Puesto que $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$, por el teorema 2.1.3(d), se tiene que

$$F_n(x) = \mathbb{P}_n((-\infty, x]) \to \mathbb{P}((-\infty, x]) = F(x)$$

cuando $n \to \infty$, es decir, $F_n \xrightarrow{d} F$.

**Suficiencia**. Supongamos ahora que $F_n \xrightarrow{d} F$. Probaremos que

$$\int_{\mathbb{R}} f \; d\mathbb{P}_n \to \int_{\mathbb{R}} f \; d\mathbb{P} \qquad \forall f \in C_b(\mathbb{R}). \tag{2.3.1}$$

Teniendo en mente que toda función monótona posee un conjunto a lo sumo numerable de puntos de discontinuidad, dado $\epsilon > 0$ podemos encontrar puntos $a, b \in \mathrm{C}(F)$ tales que

$$a < b, \qquad F(a) < \epsilon, \quad \text{y} \quad F(b) > 1 - \epsilon.$$

Así también, tomando una función arbitraria $f \in C_b(\mathbb{R})$, escojamos puntos $x_i \in \mathrm{C}(F)$ tales que $a = x_0 < x_1 < \cdots < x_k = b$, satisfaciendo

$$|f(x) - f(x_i)| < \epsilon \quad \forall x \in [x_{i-1}, x_i].$$

Definamos las sumas

$$S := \sum_{i=1}^{k} f(x_i)[F(x_i) - F(x_{i-1})], \qquad S_n := \sum_{i=1}^{k} f(x_i)[F_n(x_i) - F_n(x_{i-1})].$$

Por la hipótesis, $S_n \to S$ cuando $n \to \infty$. Por lo tanto, observando que

$$S = \sum_{i=1}^{k} f(x_i)\mathbb{P}((x_{i-1}, x_i]) = \sum_{i=1}^{k} \int_{(x_{i-1}, x_i]} f(x_i) \; d\mathbb{P},$$

y

$$\int_{\mathbb{R}} f \, d\mathbb{P} = \int_{(a,b]^c} f \, d\mathbb{P} + \int_{(a,b]} f \, d\mathbb{P},$$

y si $M > 0$ es tal que $|f(x)| \leq M$ para todo $x \in \mathbb{R}$, entonces

$$\begin{array}{rcl} \left|\int_{\mathbb{R}} f \, d\mathbb{P} - S\right| & = & \left|\int_{(a,b]^c} f \, d\mathbb{P} + \sum_{i=1}^{k} \int_{(x_{i-1},x_i]} (f(x) - f(x_i)) \, d\mathbb{P}\right| \\ & \leq & MF(a) + M(1 - F(b)) + \epsilon\mathbb{P}(a,b] \\ & \leq & (2M+1)\epsilon. \end{array}$$

Análogamente,

$$\begin{array}{rcl} \left|\int_{\mathbb{R}} f \, d\mathbb{P}_n - S_n\right| & = & \left|\int_{(a,b]^c} f \, d\mathbb{P}_n + \sum_{i=1}^{k} \int_{(x_{i-1},x_i]} (f(x) - f(x_i)) \, d\mathbb{P}_n\right| \\ & \leq & MF_n(a) + M(1 - F_n(b)) + \epsilon \\ & \to & MF(a) + M(1 - F(b)) + \epsilon \\ & < & 2(M+1)\epsilon \end{array}$$

para $n$ suficientemente grande. Por lo tanto

$$\left|\int_{\mathbb{R}} f \, d\mathbb{P}_n - \int_{\mathbb{R}} f \, d\mathbb{P}\right| \leq \left|\int_{\mathbb{R}} f \, d\mathbb{P}_n - S_n\right| + |S_n - S| + \left|\int_{\mathbb{R}} f \, d\mathbb{P} - S\right|$$

Tomando $\limsup$ en ambos miembros de la última desigualdad, se tiene que:

$$\begin{array}{rcl} \limsup_{n\to\infty} \left|\int_{\mathbb{R}} f \, d\mathbb{P}_n - \int_{\mathbb{R}} f \, d\mathbb{P}\right| & \leq & \limsup_{n\to\infty} \left|\int_{\mathbb{R}} f \, d\mathbb{P}_n - S_n\right| \\ & + & \limsup_{n\to\infty} |S_n - S| + \left|\int_{\mathbb{R}} f \, d\mathbb{P} - S\right| \\ & \leq & 2(M+1)\epsilon + 0 + (2M+1)\epsilon. \end{array}$$

Siendo $\epsilon > 0$ arbitrario se tiene que $\limsup_{n\to\infty} \left|\int_{\mathbb{R}} f \, d\mathbb{P}_n - \int_{\mathbb{R}} f \, d\mathbb{P}\right| = 0$, es decir, se cumple el límite (2.3.1). ■

## 2.4. Convergencia en distribución de variables aleatorias

**Definición 2.4.1.** *Sean $X, X_1, X_2, \cdots$ vv.aa., y $F, F_1, F_2, \cdots$ sus correspondientes f.d.p. Decimos que $\{X_n\}$* converge en distribución *a $X$ (y escribimos $X_n \xrightarrow{d} X$) si y sólo si $F_n \xrightarrow{d} F$.*

De la definición precedente y del teorema 2.3.2, consideremos vv.aa. $X, X_1, X_2, \cdots$ definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{Q})$. Entonces vemos que $X_n \xrightarrow{d} X$ si y sólo si $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$, en donde $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \cdots$ son las correspondientes probabilidades de distribución inducidas por las vv.aa. $X, X_1, X_2, \cdots$, i.e., $\mathbb{P} = \mathbb{Q} \circ X^{-1}$, $\mathbb{P}_n = \mathbb{Q} \circ X_n^{-1}$.

Por otra parte, puesto que

$$\mathbb{E}f(X_n) = \int_\Omega f(x)\, d\mathbb{P}_n(x), \qquad \text{y} \qquad \mathbb{E}f(X) = \int_\Omega f(x)\, d\mathbb{P}(x), \qquad \forall f \in C_b(\mathbb{R}),$$

se sigue de la definición 2.1.1 y del teorema de Portmanteau 2.1.3, que:

**Teorema 2.4.2.** *Sean $X, X_1, X_2, \cdots$ vv.aa. definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{Q})$. Entonces las siguientes proposiciones son equivalentes:*

*(a)* $X_n \xrightarrow{d} X$*;*

*(b)* $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ *para toda* $f \in C_b(\mathbb{R})$*;*

*(c)* $\limsup_{n\to\infty} \mathbb{Q}(X_n \in F) \leq \mathbb{Q}(X \in F)$ *para todo cerrado* $F$ *de* $\mathbb{R}$*;*

*(d)* $\liminf_{n\to\infty} \mathbb{Q}(X_n \in G) \geq \mathbb{Q}(X \in G)$ *para todo abierto* $G$ *de* $\mathbb{R}$*;*

*(e)* $\lim_{n\to\infty} \mathbb{Q}(X_n \in A) = \mathbb{Q}(X \in A)$ *para todo* $A \in \mathcal{B}(\mathbb{R})$ *tal que* $\mathbb{Q}(X \in \mathrm{Fr}A) = 0$.

Como consecuencia del teorema 2.4.2(a), (b), tenemos el siguiente resultado conocido como el *teorema de la función continua* (ver, por ejemplo, [3, pág. 26]).

**Corolario 2.4.3.** *Sean $(\Omega, \mathcal{F}, \mathbb{Q})$ un espacio de probabilidad, $X_n \xrightarrow{d} X$, y $g : \mathbb{R} \to \mathbb{R}$ una función medible tal que $\mathbb{Q}(X \in \mathrm{D}_g) = 0$, en donde $\mathrm{D}_g$ es el conjuto de discontinuidades de $g$, entonces $g(X_n) \xrightarrow{d} g(X)$. En particular, si $g$ es continua en $\mathbb{R}$, entonces $g(X_n) \xrightarrow{d} g(X)$.*

**Observación 2.4.4.**

*(1) La convergencia en distribución está completamente determinada por las f.d.p., o equivalentemente, por las medidas de probabilidad inducidas por las vv.aa. De aquí se sigue que, en general, para convergencia en distribución no es necesario que las vv.aa. estén definidas sobre el mismo espacio de probabilidad.*

*(2) Supongamos que $X, X_1, X_2, \cdots$ están definidos en el mismo espacio de probabilidad. En el ejercicio 3 de la lista de éste capítulo se pide demostrar que convergencia en probabilidad implica convergencia en distribución. El recíproco, en general, es falso como se prueba en el ejercicio 4; pero una excepción es el caso en el que el límite $X = a$ es una constante (ver ejercicio 5). Por otra parte, puesto que convergencia c.p.1. implica convergencia en probabilidad, se tiene que*

$$X_n \to X \quad \text{c.p.1 implica que} \quad X_n \xrightarrow{d} X.$$

*En general, el recíproco es falso. Sin embargo, una "especie de recíproco"muy útil en las aplicaciones es el siguiente resultado conocido como* teorema de representación de Skorohod*:*

**Teorema 2.4.5.** *Si $X_n \xrightarrow{d} X$, entonces existe un espacio de probabilidad $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ y vv.aa. $X^*, X_n^*$, $n \geq 1$, sobre $\Omega^*$ tales que $X_n^* \sim X_n$, $X^* \sim X$, y $X_n^*(x) \to X^*(x)$ para todo $x \in \Omega^*$.*

**Dem.** Definamos $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ como el "espacio de Lebesgue"sobre el intervalo $(0,1)$, es decir, $\Omega^* := (0,1)$, $\mathcal{F}^* := \mathcal{B}(0,1)$ y $\mathbb{P}^* = \lambda$, la medida de Lebesgue sobre $\Omega^*$.

Si $F$ es la f.d.p. de $X$, y $F_n$ es la f.d.p. de $X_n$ para todo $n \geq 1$, definimos

$$F^{-1}(x) := \text{ínf}\{t \in \mathbb{R} : F(t) \geq x\}, \quad \forall x \in (0,1),$$

y similarmente para $F_n$. Por último, definimos $X^* := F^{-1}$, $X_n^* := F_n^{-1}$ sobre $(0,1)$, y es fácil verificar que estas vv.aa. satisfacen la conclusión del teorema.

Denotemos por $F_n^*$ la f.d.p. de $X_n^*$, y $F^*$ la f.d.p. de $X^*$. Entonces

$$\begin{aligned} F_n^*(z) &= \mathbb{P}^*(X_n^* \leq z) \\ &= \lambda(\{y \in \Omega^* : X_n^*(y) \leq z\}) \\ &= \lambda(\{y \in (0,1) : F_n^{-1}(y) \leq z\}) \\ &= \lambda(\{y \in (0,1) : y \leq F_n(z)\}) \\ &= \lambda((0, F_n(z)]) \\ &= F_n(z), \end{aligned}$$

para todo $z \in \mathbb{R}$. Esto prueba que $X_n^* \sim X_n$, y similarmente se obtiene que $X^* \sim X$.

Claramente se tiene, dado $x \in \Omega^*$:

(i) $X^*(x) \leq t \Longleftrightarrow F(t) \geq x$;

(ii) $X_n^*(x) \leq t \Longleftrightarrow F_n(t) \geq x$ para cada $n \in \mathbb{N}$.

(iii) $X_n^*, X^* : (0,1) \to \mathbb{R}$ son crecientes en $\Omega^* = (0,1)$.

**(a)** Probemos que $\liminf_{n\to\infty} X_n^*(x) \geq X^*(x)$ para cada $x \in \Omega^* = (0,1)$: Sea $x \in (0,1)$ y $\epsilon > 0$. Escojamos un punto de continuidad $t \in \mathrm{C}(F)$ de $F$ tal que $X^*(x) - \epsilon < t < X^*(x)$. De inciso (i) se tiene que $F(t) < x$, y como $F_n(t) \to F(t)$, se sigue que existe $n_0 \in \mathbb{N}$ tal que

$$F_n(t) < x \qquad \forall n \geq n_0.$$

Luego, por (ii), $X_n^*(x) > t$ para todo $n \geq n_0$, de donde

$$\liminf_{n\to\infty} X_n^*(x) \geq t > X^*(x) - \epsilon$$

para cada $\epsilon > 0$. Esto prueba que $\liminf_{n\to\infty} X_n^*(x) \geq X^*(x)$.

**(b)** Si $X^*$ es continua en el punto $x$, entonces probabremos que vale $\limsup_{n\to\infty} X_n^*(x) \leq X^*(x)$:

Sean $x, x' \in (0,1)$ con $x < x'$, y sea $\epsilon > 0$ arbitrario. Escojamos un punto de continuidad de $F$, digamos $t$, tal que $X^*(x') < t < X^*(x') + \epsilon$. Ahora, como $X^*(x') < t$, luego $F(X^*(x')) \leq F(t)$. Luego de (i), $F(X^*(x')) \geq x'$. Así, $x < x' \leq F(X^*(x')) \leq F(t)$. Como $F_n(t) \to F(t)$, luego existe $n_0 \in \mathbb{N}$ tal que

$$F_n(t) \geq x \qquad \forall n \geq n_0.$$

Por (ii), $X_n^*(x) \leq t$ para todo $n \geq n_0$, de donde se tiene combinando las otras desigualdades

$$X_n^*(x) \leq t < X^*(x') + \epsilon \qquad \forall n \geq n_0,$$

lo que implica que $\limsup_{n\to\infty} X_n^*(x) \leq X^*(x') + \epsilon$ para todo $\epsilon > 0$. Consecuentemente,

$$\limsup_{n\to\infty} X_n^*(x) \leq X^*(x').$$

Como $X^*$ es continua en $x$, luego $X^*(x') \to X^*(x)$ cuando $x' \to x$. Luego

$$\limsup_{n\to\infty} X_n^*(x) \leq X^*(x) \quad \text{para todo punto de continuidad } x \text{ de } X^*.$$

**(c)** Como $X^* = F^{-1}$ es aplicación creciente $(0,1) \to \mathbb{R}$, por lo que $X^*$ posee un número a lo sumo numerable de puntos de discontinuidad $D^*$. Dicha colección de puntos al ser a lo sumo numerable, posee medida de Lebesgue

0, i.e., $\lambda(D^*) = \mathbb{P}^*(D^*) = 0$. Redefiniendo $X_n^*, X^*$, al multiplicarlas por $1_{D^*}^c$, tenemos que $X_n^*(x) = X^*(x) = 0$ para todo $x \in D^*$. Luego de (a), (b) y esta redefinición, tenemos que

$$X_n^*(x) \to X^*(x) \qquad \forall x \in \Omega^*,$$

con $X_n^* \sim X_n$ y $X^* \sim X$. ■

**Observación 2.4.6.** *En el teorema precedente, las variables $X, X_1, X_2, \cdots$ pueden no estar definidas en el mismo espacio de probabilidad. Algunas aplicaciones interesantes del teorema de representación de Skorohod se pueden encontrar en [3, 13, 14]. Para una generalización a espacios polacos, veáse por ejemplo [7].*

*Para aplicaciones de la convergencia débil, se puede ver, por ejemplo, los libros [2, 3, 11, 12], o el artículo de Iglehart [10]. Para aplicaciones de la convergencia débil en el estudio de la convergencia de procesos de Markov, ver por ejemplo, [7]. Para aplicaciones en modelos en matemáticas financieras, así como extensiones del concepto de convergencia débil, se puede ver el texto [8].*

## 2.5. Ejercicios

1. Demuestre que $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$ sí y sólo si cualquier subsucesión $\{\mathbb{P}_{n'}\}$ de $\{\mathbb{P}_n\}$ contiene una subsucesión $\{\mathbb{P}_{n''}\}$ tal que $\mathbb{P}_{n''} \xrightarrow{d} \mathbb{P}$.

2. Demuestre que si $\{X_n\}$ y $\{Y_n\}$ son dos sucesiones de vv.aa. tales que $X_n \xrightarrow{d} X$ y $|X_n - Y_n| \to 0$ en probabilidad, entonces $Y_n \xrightarrow{d} X$.

3. Sea $A \in \mathcal{B}(\mathbb{R})$ un boreliano de $\mathbb{R}$ tal que $\mathbb{P}(X \in \mathrm{Fr}A) = 0$. Demuestre que:

   (a) Si $X_n \to X$ en probabilidad, entonces
   $$\mathbb{P}([X_n \in A]\Delta[X \in A]) \to 0 \qquad \text{cuando } n \to \infty.$$
   De aquí concluya que

   (b) Convergencia en probabilidad implica convergencia en distribución.

4. Demuestre que convergencia en distribución no implica necesariamente convergencia en probabilidad. Sin embargo vea el siguiente ejercicio:

5. Sea $a$ una v.a. constante. Demuestre que $X_n \xrightarrow{d} a$ sí y sólo si $X_n \to a$ en probabilidad.

6. Pruebe que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} 1$, entonces $X_n Y_n \xrightarrow{d} X$.

7. Sea $\theta$ una v.a. distribuida uniformemente sobre $[0, 2\pi]$, y sea $X_k := \mathrm{sen}\,(k\theta)$, $k = 1, 2, \cdots$. Demuestre que $\frac{1}{n} S_n \to 0$ c.p.1., en donde $S_n = X_1 + \cdots + X_n$.

8. Sea $(\Omega, \rho)$ un espacio métrico, $\mathcal{F} = \mathcal{B}(\Omega)$ la $\sigma$-álgebra de Borel sobre $\Omega$, $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \cdots$ medidas de probabilidad borelianas sobre $\Omega$. Demuestre que
   $$\sup_{B \in \mathcal{B}(\Omega)} |\mathbb{P}_n(B) - \mathbb{P}(B)| \to 0 \quad \text{implica que } \mathbb{P}_n \xrightarrow{d} \mathbb{P}.$$
   Dé un ejemplo mostrando que el recíproco es falso.

   Nota: Si $\mathbb{P}$ y $\mathbb{Q}$ son dos medidas de probabilidad sobre un espacio medible arbitrario $(\Omega, \mathcal{F})$, la función $V(P, Q) := 2 \sup_{B \in \mathcal{F}} |\mathbb{P}(B) - \mathbb{Q}(B)|$ se llama la *distancia en variación* o *distancia de Kolmogorov* entre las medidas de probabilidad $\mathbb{P}$ y $\mathbb{Q}$.

# Capítulo 3

# FUNCIONES CARACTERÍSTICAS

Alrededor del año 1900, el matemático ruso A. M. Lyapunov (1857-1918) introdujo en la teoría de probabilidad el uso de las transformadas de Fourier. Lyapunov usó estas transformadas, conocidas en probabilidad con el nombre de *funciones características*, para demostrar el teorema del límite central, y actualmente constituyen una de las herramientas más usadas para estudiar, entre otras cosas, convergencia en distribución. Por otro lado, en lugar de usar transformadas de Fourier se podrían introducir transformadas de Laplace (bilaterales) para obtener resultados análogos. Al final de este capítulo introducimos brevemente la transformada de Laplace de una v.a. $X$, más conocida como la *función generadora de momentos de* $X$. (ver, por ejemplo, [1, 6, 11]).

## 3.1. Funciones características de medidas de probabilidad y de vv.aa.

**Definición 3.1.1.** *Sea $\mu$ una medida de probabilidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La transformada de Fourier de $\mu$ se llama la* función característica *asociada a $\mu$. Es decir, la función característica (que abreviaremos como f.c.) de $\mu$ es la función $\phi : \mathbb{R} \to \mathbb{C}$ con valores complejos definida por*

$$\phi(t) = \int_{\mathbb{R}} \exp(itx)\ \mu(dx), \qquad \forall t \in \mathbb{R}.$$

*Si* $\mu = \mathbb{P}_X$ *es la medida de distribución de probabilidad de la v.a.* $X$*, entonces podemos escribir*

$$\begin{aligned}\phi(t) &= \mathbb{E}\exp(itX) = \int_\Omega \exp(itX)\ d\mathbb{P} \\ &= \int_\mathbb{R} \exp(itx)\ \mathbb{P}_X(dx),\end{aligned}$$

*y decimos que* $\phi$ *es la* función característica *(f.c.) de* $X$*.*

*Equivalentemente, si* $F$ *es la f.d.p. asociada a* $\mu$ *(ó a* $X$*), decimos que* $\phi$ *es la* función característica *de* $F$*, y también podemos escribir*

$$\phi(t) = \int_\mathbb{R} \exp(itx)\ dF(x).$$

**Observación 3.1.2.** *Los casos más usuales en la práctica se presentan cuando* $X$ *es una v.a.* *<u>discreta</u>, en cuyo caso su f.c. es de la forma*

$$\phi(t) = \sum_x \exp(itx)\mathbb{P}(X = x),$$

*y cuando* $X$ *es* *<u>absolutamente continua</u>, en cuyo caso la f.c. resulta*

$$\phi(t) = \int_{-\infty}^{\infty} \exp(itx)\ f_X(x)\ dx,$$

*en donde* $f_X$ *es la función de densidad de* $X$*.*

**Ejemplo 3.1.3.** *(a) Supóngase que* $X$ *tiene una distribución de Poisson con parámetro* $\lambda > 0$*,*

$$\mathbb{P}(X = n) = \exp(-\lambda)\frac{\lambda^n}{n!}, \qquad n = 0, 1, 2, \cdots.$$

*Entonces su f.c. está dada por*

$$\phi(t) = \exp[\lambda(\mathrm{e}^{it} - 1)], \qquad \forall t \in \mathbb{R}.$$

*En efecto*

$$\begin{aligned}\phi(t) &= \sum_{n=0}^\infty \mathrm{e}^{itn}\mathbb{P}(X = n) \\ &= \mathrm{e}^{-\lambda} \sum_{n=0}^\infty \frac{(\lambda \mathrm{e}^{it})^n}{n!} \\ &= \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda \mathrm{e}^{it}} = \exp[\lambda(\mathrm{e}^{it} - 1)].\end{aligned}$$

(b) *Supóngase que* $Z$ *tiene una distribución* $N(0,1)$*, normal con media* $0$ *y varianza* $1$*; la densidad de* $Z$ *es*

$$f_Z(x) = \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}, \qquad \forall x \in \mathbb{R}.$$

*Entonces su f.c. es*

$$\phi_Z(t) = \mathrm{e}^{-t^2/2}, \qquad \forall t \in \mathbb{R}.$$

*En efecto,*

$$\begin{aligned} \phi_Z(t) &= \int_{-\infty}^{\infty} \mathrm{e}^{itx} f_Z(x)\ dx \\ &= \tfrac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-(x-it)^2/2}\ dx, \end{aligned}$$

*y haciendo el cambio de variable* $y = x - it$ *se obtiene el resultado deseado.*

(c) *Análogamente, considerando la v.a.* $Y = \sigma Z + \mu$*, con* $Z$ *distribuida* $N(0,1)$*, es fácil probar que* $Y$ *tiene distribución normal* $N(\mu, \sigma^2)$*, entonces se puede demostrar que la f.c. de* $Y$ *es*

$$\phi_Y(t) = \exp\left(i\mu t - \sigma^2 t^2/2\right) \qquad \forall t \in \mathbb{R}.$$

Al final de este capítulo se da una tabla de funciones características de algunas de las distribuciones usuales.

## 3.2. Propiedades de las funciones características

**Teorema 3.2.1.** *Sea* $\phi : \mathbb{R} \to \mathbb{C}$ *la f.c. de una v.a.* $X$*. Entonces*

(a) $\phi(0) = 1$*;*

(b) $|\phi(t)| \le \phi(0) = 1$ *para todo* $t \in \mathbb{R}$*;*

(c) $\phi$ *es uniformemente continua sobre* $\mathbb{R}$*;*

(d) $\overline{\phi(t)} = \phi(-t)$ *(donde* $\overline{z}$ *es el conjugado de* $z \in \mathbb{C}$*);*

*(e) Si $Y = a + bX$, con $a, b$ constantes, entonces la f.c. de $Y$ es*

$$\phi_Y(t) = \mathrm{e}^{iat}\phi_X(bt) \qquad \textit{para todo } t \in \mathbb{R}.$$

**Dem.** (a) $\phi(0) = \int_\Omega \mathrm{e}^0 \, d\mathbb{P} = \mathbb{P}(\Omega) = 1$.

(b) Puesto que $|\mathrm{e}^{ix}| = 1$ para cada $x \in \mathbb{R}$,

$$|\phi(t)| \leq \int_{\mathbb{R}} |\mathrm{e}^{itx}| \, d\mathbb{P}_X(x) = 1,$$

en donde $\mathbb{P}_X$ es la medida de probabilidad inducida por $X$.

(c) Para cualquier $t, s \in \mathbb{R}$

$$\begin{aligned} |\phi(t+s) - \phi(t)| &= \left| \int_{\mathbb{R}} \mathrm{e}^{itx}(\mathrm{e}^{isx} - 1) \, d\mathbb{P}_X(x) \right| \\ &\leq \int_{\mathbb{R}} |\mathrm{e}^{isx} - 1| \, d\mathbb{P}_X(x), \end{aligned}$$

y esta integral no depende de $t$. Ahora, puesto que $2 \geq |\mathrm{e}^{isx} - 1| \to 0$ cuando $s \to 0$, la proposición (c) se sigue del teorema de convergencia dominada de Lebesgue.

Las partes (d) y (e) son fáciles y se dejan como ejercicio. ■

**Teorema 3.2.2.** *Sean $X_1, \cdots, X_n$ vv.aa. independientes con funciones características (abreviado ff.cc.) $\phi_1, \cdots, \phi_n$, respectivamente, y sea $\phi_S$ la f.c. de la suma $S = X_1 + \cdots + X_n$. Entonces*

$$\phi_S(t) = \phi_1(t) \cdots \phi_n(t), \qquad \forall t \in \mathbb{R}.$$

*En particular, si $X_1, \cdots, X_n$ son vv.aa. i.i.d., y su f.c. común es $\phi$, entonces*

$$\phi_S(t) = \phi(t)^n, \qquad \forall t \in \mathbb{R}.$$

**Dem.**

$$\begin{aligned} \phi_S(t) &= \mathbb{E}(\mathrm{e}^{itS}) \\ &= \mathbb{E} \textstyle\prod_{k=1}^n \exp(itX_k) \\ &= \textstyle\prod_{k=1}^n \mathbb{E}(\exp(itX_k)) \\ &= \textstyle\prod_{k=1}^n \phi_k(t). \end{aligned}$$

■

## 3.3. Independencia de vv.aa. y convolución de f.d.p.

**Observación 3.3.1.** *El teorema 3.2.2 precedente nos permite calcular la f.c. de la suma* $S = X_1 + \cdots + X_n$ *, en términos de las ff.cc. individuales de las v.a.* $X_j$*, siempre y cuando estas vv.aa. sean independientes. Conviene mencionar que el teorema 3.2.2 se puede escribir en términos de la convolución de funciones de distribución. Para ello, recordemos que si* $F$ *y* $G$ *son f.d.p., su* convolución *se define como*

$$(F * G)(x) = \int_{-\infty}^{\infty} F(x-y)\ dG(y),$$

*y esta operación es conmutativa y asociativa.*

**Proposición 3.3.2.** *Si* $X$*,* $Y$ *son vv.aa. independientes con distribuciones* $F_X$ *y* $F_Y$ *respectivamente, entonces la distribución de* $X+Y$ *es la convolución de* $F_X$ *y* $F_Y$*, es decir,*

$$\mathbb{P}(X+Y \leq x) = (F_X * F_Y)(x) \qquad \forall x \in \mathbb{R}.$$

*Por la asociatividad de la convolución, el resultado se puede extender a cualquier número finito de vv.aa. independientes: Si* $X_1, \cdots, X_n$ *son vv.aa. independientes con distribuciones* $F_1, \cdots, F_n$*, respectivamente. Entonces la distribución de* $S = X_1 + \cdots + X_n$ *es la convolución*

$$F = F_1 * \cdots * F_n.$$

*(Observamos que combinando esta proposición con el teorema 3.2.2, se tiene que la f.c. o* transformada de Fourier *de una convolución de funciones de distribución de probabilidades* $F_1, \cdots, F_n$ *es el producto de las ff.cc. de* $F_1, \cdots, F_n$*.)*

**Dem.** Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función definida por

$$\begin{aligned} f(r,s) &= 1 \text{ si } r+s \leq x \\ &= 0 \text{ en caso contrario.} \end{aligned}$$

Nótese que la función $f$ es la función indicadora del conjunto cerrado $\{(r,s) \in \mathbb{R}^2 : r+s \leq x\}$, y por lo tanto, es Borel-medible. Por independencia, la distribución conjunta $F_{X,Y}$ de $X$ y $Y$ es el producto de $F_X$ y $F_Y$, vale decir, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. Puesto que $f(X,Y) = I_{[X+Y\leq x]}$, vemos que, por Tonelli

$$\begin{array}{rcl} \mathbb{P}(X+Y\leq x) & = & \mathbb{E}f(X,Y) \\ & = & \int_{\mathbb{R}^2} f(r,s)\ dF_{X,Y}(r,s) \\ & = & \int_{\mathbb{R}^2} I_{\{(r,s)\in\mathbb{R}^2: r+s\leq x\}}(r,s)\ dF_{X,Y}(r,s) \\ & = & \int_{\mathbb{R}^2} I_{(-\infty,\infty)}(r) I_{(-\infty,x-r)}(s)\ dF_{X,Y}(r,s) \\ & = & \int_{-\infty}^{\infty}\left(\int_{-\infty}^{x-r}\ dF_Y(s)\right)\ dF_X(r) \\ & = & \int_{-\infty}^{\infty} F_Y(x-r)\ dF_X(r). \\ & = & (F_X * F_Y)(x). \end{array}$$

■

**Corolario 3.3.3.** *Sean $X$, $Y$ vv.aa. independientes como en la proposición 3.3.2, y supóngase que $X$ es absolutamente continua con densidad $f_X$, es decir,*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\ dt, \quad \forall x\in\mathbb{R}.$$

*Entonces $X+Y$ es absolutamente continua y su función de densidad de probabilidad es (c.t.p.)*

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(t-y)\ dF_Y(y).$$

**Dem.** Por el teorema de Tonelli y la proposición 3.3.2, tenemos

$$\begin{array}{rcl} F_{X+Y}(x) & = & \mathbb{P}(X+Y\leq x) \\ & = & (F_X * F_Y)(x) \\ & = & \int_{-\infty}^{\infty} F_X(x-y)\ dF_Y(y) \\ & = & \int_{-\infty}^{\infty}\left(\int_{-\infty}^{x-y} f_X(t)\ dt\right)\ dF_Y(y) \\ & = & \int_{-\infty}^{\infty}\left(\int_{-\infty}^{x} f_X(t-y)\ dt\right)\ dF_Y(y) \\ & = & \int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f_X(t-y)\ dF_Y(y)\right)\ dt. \end{array}$$

■

Como una aplicación del corolario 3.3.3, consideremos dos vv.aa. independientes $X$ y $Y^\sigma$, en donde $X$ tiene distribución $F$ y $Y^\sigma$ está distribuida normalmente con media cero y varianza $\sigma^2$ ($\sigma > 0$), es decir, $Y^\sigma$ tiene distribución

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2\sigma^2}\, dt, \quad \forall x \in \mathbb{R}.$$

Por otro lado, si $F_\sigma$ denota la f.d.p. de $X + Y^\sigma$, entonces de la proposición 3.3.2 y el corolario 3.3.3 se sigue $F_\sigma(x) = (F * G_\sigma)(x)$ tiene densidad

$$f_\sigma(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-u)^2/2\sigma^2}\, dF(u). \tag{3.3.1}$$

Usaremos este resultado en la demostración del siguiente teorema:

## 3.4. Fórmula de inversión. Fórmula integral de Fourier

**Teorema 3.4.1** (Fórmula de inversión). *Sea $X$ una v.a. con distribución $F$ y f.c. $\phi$. Si $a$ y $b$ son puntos de continuidad de $F$, con $a < b$, entonces*

$$F(b) - F(a) = \lim_{\sigma \to 0^+} \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) \mathrm{e}^{-\sigma^2 t^2/2} \left( \int_a^b \mathrm{e}^{-itx}\, dx \right) dt.$$

**Dem.** Sea $Y^\sigma \sim N(0, \sigma^2)$, con $\sigma > 0$, una v.a. independiente de $X$. En el párrafo anterior observamos que la distribución $F_\sigma$ se $X + Y^\sigma$ es $F_\sigma(x) = (F * G_\sigma)(x)$ y tiene densidad $f_\sigma(x)$ dada por (3.3.1).

Ahora bien, como primer paso de la demostración, probaremos que

$$f_\sigma(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-itx} \phi_\sigma(t)\, dt, \tag{3.4.1}$$

en donde

$$\phi_\sigma(t) = \phi(t)\mathrm{e}^{-\sigma^2 t^2/2} = \mathrm{e}^{-\sigma^2 t^2/2} \int_{-\infty}^{\infty} \mathrm{e}^{itu}\, dF(u) \tag{3.4.2}$$

es la f.c. de $X + Y^\sigma$ (véase el ejemplo 3.1.3-(c) y el teorema 3.2.2). Para probar (3.4.1), denotemos por $g_\sigma(x)$ la densidad de $G_\sigma(x)$:

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-x^2/2\sigma^2\right).$$

Entonces el lado derecho de (3.4.1) resulta, tomando en cuenta (3.4.2),

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\mathrm{e}^{-\sigma^2t^2/2}\int_{-\infty}^{\infty} \mathrm{e}^{itu}\, dF(u)\, dt =$$

(por el teorema de Fubini y denotando por $\phi_{1/\sigma}$ como la f.c. de una v.a. distribuida $N(0,1/\sigma^2)$, así como de la relación (3.3.1) arriba,)

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{it(u-x)}\mathrm{e}^{-\sigma^2t^2/2}\, dt\, dF(u)\\
&= \int_{-\infty}^{\infty} \tfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{it(u-x)}g_{1/\sigma}(t)\, dt\, dF(u)\\
&= \int_{-\infty}^{\infty} \tfrac{1}{\sigma\sqrt{2\pi}}\phi_{1/\sigma}(u-x)\, dF(u)\\
&= \int_{-\infty}^{\infty} \tfrac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-u)^2/2\sigma^2}\, dF(u)\\
&= f_\sigma(x),
\end{aligned}$$

lo cual prueba (3.4.1).

Ahora, integrando (3.4.1) de $a$ hacia $b$, tenemos:

$$F_\sigma(b) - F_\sigma(a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_\sigma(t)\left(\int_a^b \mathrm{e}^{-itx}\, dx\right)\, dt.$$

Por lo tanto, para completar la demostración, basta probar que

$$\lim_{\sigma\to 0^+} F_\sigma(x) = F(x) \qquad \text{si } x \text{ es punto de continuidad de } F. \tag{3.4.3}$$

Para ésto, notemos primero que, por la desigualdad de Chebyshev, dado $\epsilon > 0$, podemos escoger $\sigma > 0$ tal que $\mathbb{P}(|Y^\sigma| \geq \epsilon) \leq \epsilon$. Para tal $\sigma$, vemos entonces que

$$\begin{aligned}
F_\sigma(x) &= \mathbb{P}(X + Y^\sigma \leq x)\\
&= \mathbb{P}(X \leq x - Y^\sigma, |Y^\sigma| < \epsilon) + \mathbb{P}(X \leq x - Y^\sigma, |Y^\sigma| \geq \epsilon)\\
&\leq \mathbb{P}(X \leq x + \epsilon) + \epsilon = F(x+\epsilon) + \epsilon,
\end{aligned}$$

y también

$$\begin{aligned}
F_\sigma(x) &\geq \mathbb{P}(X \leq x - \epsilon, |Y^\sigma| < \epsilon)\\
&\geq \mathbb{P}(X \leq x - \epsilon) - \epsilon\\
&= F(x-\epsilon) - \epsilon.
\end{aligned}$$

Finalmente, combinando las últimas dos desigualdades con el hecho de que

$$F(x-\epsilon) \leq F(x) \leq F(x+\epsilon)$$

y $x \in C(F)$, donde $C(F)$ es el conjunto de puntos de continuidad de $F$, de donde concluímos que (3.4.3) se cumple. ■

**Observación 3.4.2.** *(1) Si por ejemplo, $b \notin C(F)$, entonces el límite en el teorema 3.4.1 aún existe, pero su valor ya no es $F(b)$ sino*

$$(F(b+) + F(b-))/2.$$

*La misma modificación se aplica a $F(a)$ si $a \notin C(F)$.*

*(2) Como una consecuencia de la fórmula de inversión en el teorema 3.4.1 tenemos que una f.d.p. está determinada en forma única por su f.c. Es decir, si $F$ y $G$ son dos f.d.p. con la misma f.c. $\phi$, se sigue del teorema 3.4.1 que $F(x) = G(x)$ para todo $x \in C(F) \cap C(G)$, y por la continuidad por la derecha de $F$ y $G$, y por el hecho de que el complemento de $C(F) \cap C(G)$ es a lo sumo numerable, obtenemos inmediatamente que $F(x) = G(x)$ para todo $x \in \mathbb{R}$.*

Si en lugar de considerar funciones de distribución consideramos las correspondientes medidas de probabilidad, concluimos:

**Teorema 3.4.3** (Teorema de unicidad)**.** *Si $\mathbb{P}$ y $\mathbb{Q}$ son medidas de probabilidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y su función característica es la misma, es decir,*

$$\int_{\mathbb{R}} \mathrm{e}^{itx}\ \mathbb{P}(dx) = \int_{\mathbb{R}} \mathrm{e}^{itx}\ \mathbb{Q}(dx) \qquad \text{para todo } t \in \mathbb{R},$$

*entonces $\mathbb{P} = \mathbb{Q}$.*

Ahora, en general, la f.c. de una v.a. no es integrable. Por ejemplo, si $X$ tiene una distribución exponencial de parámetro $\alpha = 1$, i.e., si su función de densidad de probabilidad $f_X : \mathbb{R} \to \mathbb{R}$ es

$$f_X(x) = \begin{cases} \mathrm{e}^{-x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0, \end{cases}$$

entonces su función característica es

$$\phi(t) = \frac{1}{1 - it},$$

la cual no es integrable. Pero cuando la f.c. es integrable, tenemos como una consecuencia del teorema 3.4.1 que:

**Corolario 3.4.4.** *Si $\phi(t)$ es la f.c. de una v.a. $X$ y $\phi \in L^1$, entonces $X$ es absolutamente continua y su densidad está dada por*

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\phi(t)\ dt. \tag{3.4.4}$$

*Así pues, en este caso tenemos el "par de transformadas de Fourier"dado por (3.4.4) y*

$$\phi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} f_X(x)\ dx. \tag{3.4.5}$$

*A la expresión para $f_X$ en (3.4.4) se le llama la* transformada de Fourier inversa *de $\phi$. Además, combinando (3.4.4) y (3.4.5) obtenemos la* fórmula integral de Fourier

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-it(x-y)} f_X(y)\ dy\ dt. \tag{3.4.6}$$

**Dem.** Puesto que $\phi \in L^1$, podemos usar el teorema de la convergencia dominada de Lebesgue para intercambiar el límite con la integral en el teorema 3.4.1. El resultado para $a, b \in C(F)$, $a < b$, es

$$\begin{aligned} F(b) - F(a) &= \tfrac{1}{2\pi}\textstyle\int_{-\infty}^{\infty} \phi(t) \int_a^b \mathrm{e}^{-itx}\ dx\ dt \\ &= \textstyle\int_a^b \left(\tfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\phi(t)\ dt\right)\ dx \\ &= \textstyle\int_a^b f_X(x)\ dx, \end{aligned}$$

en donde $F$ es la f.d.p. de $X$. Puesto que $F$ está determinada por sus puntos de continuidad, se sigue de la expresión anterior que la función $f_X$ en (3.4.4) es la densidad de $F$

## 3.5. Función generadora de momentos

**Definición 3.5.1.** *La* función generadora de momentos (f.g.m.) *de una v.a. $X$ se define como*

$$M_X(t) = \mathbb{E}(\mathrm{e}^{tX}) = \int_{-\infty}^{\infty} \mathrm{e}^{tx}\ dF_X(x) \tag{3.5.1}$$

*para valores $t \in \mathbb{R}$ para los cuales la integral existe. La función $M_X$ también se conoce como la* transformada (bilateral) de Laplace *de la v.a. $X$.*

A diferencia de la función característica, la cual está definida para todo $t \in \mathbb{R}$, la función generadora de momentos en general sólo existe en subconjuntos de $\mathbb{R}$. Por ejemplo, si $X$ está exponencialmente distribuida con parámetro $\lambda > 0$, su f.g.m. es

$$M_X(t) = \lambda \int_0^\infty \mathrm{e}^{-(\lambda - t)x} \, dx = \frac{\lambda}{\lambda - t}, \qquad \forall t < \lambda.$$

En general, $M_X(t)$ siempre está definida para $t = 0$, con $M_X(0) = 1$, pero puede no estar definida para todo $t \neq 0$. Así pues, decimos que $X$ tiene una f.g.m. si existe $t_0 > 0$ talque $M_X(t)$ existe para todo $|t| \leq t_0$. En este último caso, se sigue de la expansión de Taylor alrededor de $t = 0$ para la función exponencial, que

$$M_X(t) = \sum_{k=0}^\infty \frac{t^k}{k!} \mathbb{E}(X^k), \qquad |t| \leq t_0.$$

Por lo tanto, todas las derivadas de $M_X$ existen en $t = 0$ y satisfacen

$$M_X^{(k)}(0) = \mathbb{E}(X^k), \qquad k = 0, 1, 2, \cdots . \tag{3.5.2}$$

Es precisamente esta relación (3.5.2) la que da a $M_X$ el nombre de *función generadora de momentos*.

La f.g.m. tiene propiedades análogas a la función característica; en particular, satisface las versiones correspondientes del teorema 3.2.2 y del teorema de unicidad 3.4.3.

En la siguiente tabla proporcionamos las funciones características de algunas distribuciones importantes, tanto discretas como absolutamente continuas.

| Distribución | Función característica |
|---|---|
| Poisson con parámetro $\lambda > 0$:<br>$\mathbb{P}(X = k) = \mathrm{e}^{-\lambda}\lambda^k/k!,\ k = 0, 1, \cdots$ | $\phi(t) = \mathrm{e}^{\lambda(\mathrm{e}^{it}-1)}$ |
| Binomial con parámetros $N$ y $0 < p < 1$<br>$\mathbb{P}(X = n) = \frac{N!}{(N-n)!n!}p^n(1-p)^{N-n},\ n = 0, 1, \cdots, N$ | $\phi(t) = (1 - p + p\mathrm{e}^{it})^N$ |
| Geométrica con parámetro $0 < p < 1$<br>$\mathbb{P}(X = n) = p(1-p)^n,\ n = 0, 1, \cdots$ | $\phi(t) = p(1 - q\mathrm{e}^{it})^{-1},\ q = 1 - p$ |
| Normal con paráms. $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, $N(\mu, \sigma^2)$:<br>$f(x) = (2\pi\sigma^2)^{-1/2}\exp\left[-(x-\mu)^2/2\sigma^2\right],\ x \in \mathbb{R}$ | $\phi(t) = \exp\,(i\mu t - \frac{1}{2}\sigma^2 t^2$ |
| Gamma con parámetros $p, b > 0$<br>densidad $\Gamma(x; p, b)$:<br>$f(x) = \begin{cases} \frac{b^p}{\Gamma(p)}x^{p-1}\mathrm{e}^{-bx} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ | $\phi(t) = (1 - it/b)^{-p}$ |
| Exponencial con parámetro $b > 0$<br>densidad $\Gamma(x; 1, b)$:<br>$f(x) = \begin{cases} b\mathrm{e}^{-bx} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ | $\phi(t) = b/(b - it)$ |
| Uniforme sobre $(a, b)$ con $a < b$<br>densidad:<br>$f(x) = \begin{cases} (b-a)^{-1} & \text{si } x \in (a, b) \\ 0 & \text{si } x \notin (a, b) \end{cases}$ | $\phi(t) = \frac{\mathrm{e}^{ibt} - \mathrm{e}^{iat}}{it(b-a)}$ |

## 3.6. Ejercicios

1. Demuestre que la función

$$f(x) = \frac{c}{\pi}(c^2 + x^2)^{-1}, \quad c > 0, \quad -\infty < x < \infty,$$

   es la densidad de una distribución de probabilidad con función característica $\phi(t) = \mathrm{e}^{-c|t|}$. La función $f$ se llama *densidad de Cauchy con parámetro c*.

   Si $X_1, \cdots, X_n$ son vv.aa. independientes con distribución de Cauchy con la misma $c$, calcule la distribución de $S_n/n$, con $S_n = X_1 + \cdots + X_n$.

2. Use funciones características para verificar que una combinación lineal de vv.aa. independientes normalmente distribuidas también es normal (es decir, si $X_k$ tiene distribución $N(\mu_k, \sigma_k^2)$, $k = 1, \cdots, n$, entonces $a_1X_1 + \cdots + a_nX_n$ tiene distribución $N(a_1\mu_1 + \cdots + a_n\mu_n, a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2)$.

3. Supóngase que $\int_{\mathbb{R}} |x|^n \, dF(x) < \infty$ para algún entero positivo $n > 0$, y sea $\phi$ la función característica de $F$. Demuestre que la $n$-ésima derivada $\phi^{(n)}(t)$ de $\phi$ existe, es continua y satisface:

$$\phi^{(n)}(t) = \int_{\mathbb{R}} (ix)^n \mathrm{e}^{itx} \, dF(x).$$

   En particular, nótese que $\phi^{(n)}(0) = i^n \mathbb{E}(X^n)$.

4. Demuestre que si $\phi$ es una función característica, entonces en *no-negativa definida*, es decir, para cualquier $n \in \mathbb{N}$, $t_1, \cdots, t_n \in \mathbb{R}$, $z_1, \cdots, z_n \in \mathbb{C}$, vale

$$\sum_{k=1}^{n} \sum_{j=1}^{n} \phi(t_k - t_j) z_k \overline{z_j} \geq 0.$$

5. Si $X = (X_1, \cdots, X_n)$ es un vector aleatorio $n$-dimensional, su función característica conjunta se define como

$$\phi(u) = \mathbb{E}(\mathrm{e}^{iu \cdot X}) \quad \text{para todo } u = (u_1, \cdots, u_n) \in \mathbb{R}^n,$$

en donde

$$u \cdot X = \sum_{j=1}^{n} u_j X_j,$$

ó equivalentemente

$$\phi(u) = \int_{\mathbb{R}^n} \mathrm{e}^{iu\cdot x}\, dF(x),$$

con $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, y " $\cdot$ " es el producto interno euclidiano canónico en $\mathbb{R}^n$, donde $F : \mathbb{R}^n \to \mathbb{R}$ es la f.d.p. de $X$. Demuestre que:

(a) $\phi(0) = 1$, con $0 = (0, \cdots, 0) \in \mathbb{R}^n$;

(b) $|\phi(u)| \leq 1$ para todo $u \in \mathbb{R}^n$;

(c) $\phi$ es uniformemente continua sobre $\mathbb{R}^n$;

(d) Si $Y = (Y_1, \cdots, Y_n)$ es un vector aleatorio definido por $Y_k = a_k + b_k X_k$, con $a_k, b_k \in \mathbb{R}$ para $k = 1, \cdots, n$. Pruebe que la función característica $\psi$ de $Y$ es

$$\psi(u) = \mathrm{e}^{ia\cdot u}\phi(b_1 u_1, \cdots, b_n u_n), \qquad a = (a_1, \cdots, a_n).$$

6. Sea $X \sim \Gamma(p, b)$ una v.a. con distribución gamma con parámetros $p$ y $b$. Use el ejercicio 3 para probar que $\mathbb{E}X = p/b$, $\mathbb{E}X^2 = p(p+1)/b^2$, y por lo tanto, $\mathrm{var}(X) = p/b^2$.

7. Sean $X_1, \cdots, X_n$ v.a. independientes con $X_k \sim \Gamma(p_k, b)$, $k = 1, \cdots, n$. Demuestre que $X_1 + \cdots + X_n$ tiene distribución $\Gamma(p_1 + \cdots + p_n, b)$. En particular, si $X_k \sim \Gamma(p, b)$, $k = 1, \cdots, n$, son vv.aa. i.i.d., entonces

$$X_1 + \cdots + X_n \sim \Gamma(np, b).$$

8. Pruebe que si $X \sim N(0, 1)$, entonces $X^2 \sim \Gamma(1/2, 1/2)$.

9. Sean $X_1, \cdots, X_n$ v.a. i.i.d., con $X_k \sim N(0, 1)$. Pruebe que la v.a. *ji-cuadrada con $n$ grados de libertad*:

$$\chi^2_{(n)} := X_1^2 + \cdots + X_n^2$$

tiene distribución $\Gamma(n/2, 1/2)$. Demuestre de aquí que $\mathbb{E}(\chi^2_{(n)}) = n$ y $\mathrm{var}(\chi^2_{(n)}) = 2n$.

10. Obtenga resultados similares al problema precedente si $X_k \sim N(0, \sigma^2)$.

# Capítulo 4

# COMPACIDAD DE MEDIDAS Y TEOREMA DE LEVY

En este capítulo estudiaremos la relación entre convergencia débil de medidas y funciones características. Por razones que explicaremos más adelante, ahora en lugar de medidas de probabilidad, conviene considerar medidas finitas, es decir, medidas $\mu$ con $\mu(\Omega) < \infty$, pero no necesariamente $\mu(\Omega) = 1$. Así mismo, consideraremos funciones de distribución generales: funciones $F$ no-decrecientes y continuas por la derecha. Si $F$ es una función de distribución de probabilidad, es decir, si además $F$ satisface que $F(+\infty) = 1$ y $F(-\infty) = 0$, lo mencionaremos explícitamente abreviando f.d.p.

Sea $F$ una función de distribución acotada, lo cual significa que existe $M > 0$ tal que $F(+\infty) - F(-\infty) \leq M$. Entonces del curso de Probabilidad I, la función $\mu$ definida sobre intervalos acotados $(a, b]$ de $\mathbb{R}$ como $\mu(a, b] = F(a, b]$, en donde

$$F(a, b] = F(b) - F(a), \tag{4.0.1}$$

se puede extender en forma única a una medida finita $\mu$ sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ver, por ejemplo, [4]). Recíprocamente, si $\mu$ es una medida finita sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, entonces la función $F$ definida por

$$F(x) = \mu(-\infty, x], \quad \forall x \in \mathbb{R},$$

es una función de distribución acotada. Además, identificando las funciones de distribución que sólo difieren por una constante aditiva, la correspondencia entre medidas finitas y funciones de distribución acotadas es uno-a-uno.

## 4.1. Teorema de Helly

**Definición 4.1.1.** *(a) Sea $F$ una función de distribución. Definimos el* conjunto de puntos de continuidad de $F$ *como*

$$C(F) := \{x \in \mathbb{R} : F \text{ es continua en } x\} \cup \{-\infty, +\infty\},$$

*es decir, convendremos en que $-\infty$ y $+\infty$ son puntos de continuidad de $F$.*

*(b) Si $F$, $F_n$, $n = 1, 2, \cdots$, son funciones de distribución, decimos que* $F_n$ converge débilmente *a $F$, lo que escribimos $F_n \xrightarrow{d} F$, si*

$$F_n(a, b] \to F(a, b] \quad \text{para todo } a, b \in C(F).$$

*Aquí usamos la notación (4.0.1).*

**Ejemplo 4.1.2.** *Sea $\{X_n\}$ una sucesión de vv.aa.constantes definidas como $X_n = n$, $n = 1, 2, \cdots$. La correspondiente f.d.p. de $X_n$ es*

$$F_n(x) = \mathbb{P}(X_n \leq x) = \begin{cases} 0 & \text{si} \quad x < n \\ \\ 1 & \text{si} \quad x \geq n \end{cases}$$

*para $n = 1, 2, \cdots$, y $F_n(x) \to F(x)$ para todo $x \in \mathbb{R}$, en donde $F = 0$ es la función de distribución idéntica a cero. Así pues, tenemos convergencia puntual de funciones de distribución de probabilidad a una función de distribución que no es de probabilidad. Obsérvese que $F_n$ no converge débilmente a $F$ porque $F_n(+\infty) - F_n(-\infty) = 1 \nrightarrow F(+\infty) - F(-\infty) = 0$. En el siguiente teorema veremos cómo generalizar este resultado.*

**Teorema 4.1.3** (Teorema de Helly)**.** *Sea $\{F_n\}$ una sucesión acotada de funciones de distribución. Supóngase que $F_n(-\infty) = 0$ y $F_n(+\infty) \leq M$ para toda $n$. Entonces existe una función de distribución $F$ y una subsucesión $\{F_{n_k}\}$ tal que*

$$F_{n_k}(x) \to F(x) \qquad \text{para todo } x \in \mathbb{R} \cap C(F).$$

**Dem.** Sea $D = \{x_1, x_2, \cdots\}$ un conjunto numerable denso en $\mathbb{R}$. Puesto que $\{F_n(x_1)\}$ es una sucesión acotada, existe ua subsucesión $\{F_{1j}\}$ de $\{F_n\}$ tal que $\{F_{1j}(x_1)\}$ converge a un límite finito $y_1$ ($\leq M$) cuando $j \to \infty$. Así mismo, puesto que $\{F_{1j}(x_2)\}$ es acotada, existe una subsucesión $\{F_{2j}\}$ de

$\{F_{1j}\}$ tal que $\{F_{2j}(x_2)\}$ converge a un límite finito $y_2$. En general, procediendo inductivamente,existe una subsucesión $\{F_{mj}\}$ de $\{F_{m-1j}\}$ tal que $F_{mj}(x_m) \to y_m$ cuando $j \to \infty$, $(m = 1, 2, \cdots)$.

Ahora definamos $F_D : D \to \mathbb{R}$ como $F_D(x_m) = y_m$ para todo $m = 1, 2, \cdots$. Definamos $F_{n_k} := F_{kk}$, $k = 1, 2, \cdots$ la *sucesión diagonal*. Entonces

$$F_{n_k}(x) \to F_D(x), \qquad \text{para todo } x \in D.$$

También $F_D$ es no-decreciente, pues si $x, y \in D$, con $x < y$, luego $F_{n_k}(x) \leq F_{n_k}(y)$. Tomando el límite cuando $k \to \infty$, se tiene que $F_D(x) \leq F_D(y)$.

Definamos $F : \mathbb{R} \to \mathbb{R}$ como

$$F(x) := \text{ínf}\{F_D(y) : y \in D,\ y > x\}.$$

Probaremos que $F$ es una función de distribución. En primer lugar, vemos que $F$ es no-decreciente, porque si $r < s$, entonces existe $y \in D$ tal que $r < y < s$ y por lo que $F(r) \leq F_D(y) \leq F(s)$.

Para probar que $F$ es continua por la derecha, tomemos una sucesión decreciente $\{z_n\}$ que converge a $x$, de manera que $\{F(z_n)\}$ tiende a un límite $b \geq F(x)$. Si $F(x) < b$, sea $y_0 \in D$ tal que $y_0 > x$, y $b \geq F_D(y_0)$. Para $n$ suficientemente grande tenemos que $x < z_n < y_0$, de manera que $F(z_n) \leq F_D(y_0) < b$, y por lo tanto, $\text{lím}_{n\to\infty} F(z_n) \leq F_D(y_0) < b$. Pero esto es una contradicción, así concluimos que $F(z_n) \to F(x)$.

Falta demostrar la última afirmación. Sea $x \in \mathbb{R} \cap C(F)$. Si $x < y$ con $y \in D$, entonces

$$\limsup_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(y) = F_D(y).$$

Por lo tanto, tomando el ínfimo sobre los $y \in D$ con $y > x$, obtenemos

$$\limsup_{k\to\infty} F_{n_k}(x) \leq F(x).$$

Por otra parte, si $x' < y < x$, con $y \in D$, entonces

$$\begin{aligned} F(x') &\leq F_D(y) = \limsup_{k\to\infty} F_{n_k}(y) \\ &= \liminf_{k\to\infty} F_{n_k}(y) \leq \liminf_{k\to\infty} F_{n_k}(x). \end{aligned}$$

Si hacemos tender $x'$ a $x$, resulta

$$F(x-) \leq \liminf_{k\to\infty} F_{n_k}(x).$$

Pero $x \in \mathbb{R} \cap C(F)$, luego $F(x-) = F(x)$. Por tanto, $F_{n_k}(x) \to F(x)$. $\blacksquare$

**Observación 4.1.4.** *Por el teorema de Helly precedente, supusimos que $F_n$ es una función de distribución acotada con $F_n(-\infty) = 0$. Hicimos esto para simplificar la demostración, pero es claro que es una suposición inofensiva. Así, si suponemos que $F_n(+\infty) - F_n(-\infty) \leq M$ y no necesariamente $F_n(-\infty) = 0$, podríamos entonces considerar una nueva sucesión de funciones $G_n(x) := F_n(x) - F_n(-\infty)$ y aplicar el teorema de Helly a esta nueva sucesión.*

*Por otro lado, en el ejemplo 4.1.2 vimos que una sucesión de f.d.p.'s puede converger a una función de distribución que no es de probabilidad. La situación especial que debemos observar en dicho ejemplo es que las masas de probabilidad se "escapan al infinito". La medida de probabilidad $\mathbb{P}_n$ inducida por $X_n = n$ sobre $\mathbb{R}$ es la medida de Dirac concentrada en $\{n\}$, es decir, para cualquier conjunto de Borel $B$,*

$$\mathbb{P}_n(B) = \delta_{\{n\}}(B) = \begin{cases} = & 1 \quad si\ n \in B \\ = & 0 \quad si\ n \notin B. \end{cases}$$

*Para impedir el "escape"de probabilidad al infinito debemos imponer una condición de compacidad sobre el conjunto de probabilidades.*

## 4.2. Teorema de Prohorov

**Definición 4.2.1.** *Sea $(\Omega, \mathcal{F})$ un espacio medible, en donde $\Omega$ es un espacio métrico y $\mathcal{F}$ es la $\sigma$-álgebra de conjuntos Borel de $\Omega$, y sea $\mathcal{M} = \{\mu_i : i \in T\}$ una familia de medidas finitas sobre $(\Omega, \mathcal{F})$. Decimos que*

*(a) La famila $\mathcal{M}$ es* compacta *("tight" en inglés) si dado cualquier $\epsilon > 0$, existe un subconjunto compacto de $\Omega$ tal que $\mu_i(K^c) < \epsilon$ para todo $i \in T$. En particular, si $\Omega = \mathbb{R}$, entonces $\mathcal{M}$ es* compacta *si dado cualquier $\epsilon > 0$ existe un intervalo $I$ tal que $\mu_i(I^c) < \epsilon$ para todo $i \in T$.*

*(b) La famila $\mathcal{M}$ es* relativamente compacta *si cualquier sucesión en $\mathcal{M}$ posee una subsucesión que converge débilmente a una medida finita sobre $(\Omega, \mathcal{F})$.*

*Para medidas finitas $\mu, \mu_n$, $n = 1, 2, \cdots$, la definición de convergencia débil es como la dada en la definición 2.1.1: $\mu_n$* converge débilmente *a $\mu$ ($\mu_n \xrightarrow{d} \mu$) si*

$$\int_\Omega f_n\ d\mu_n \to \int_\Omega f\ d\mu \qquad \text{para toda función } f \in C_b(\Omega).$$

*(c) Si $\{F_i : i \in T\}$ es una familia de funciones de distribución,* compacidad relativa *de $\{F_i\}$ significa que la correspondiente familia de medidas asociadas es compacta (Resp. relativamente compacta).*

**Teorema 4.2.2** (Teorema de Prohorov)**.** *Sea $\mathcal{M} := \{F_i : i \in T\}$ una familia acotada, digamos $F_i(+\infty) - F_i(-\infty) \leq M$ para todo $i \in T$, de funciones de distribución en $\mathbb{R}$. Entonces $\mathcal{M}$ es compacta si y sólo si es relativamente compacta.*

**Nota importante.** El teorema de Prohorov es válido en cualquier espacio métrico separable y completo (espacio Polaco); véase, por ejemplo, [3]. En dicha referencia el teorema se escribe en términos de medidas (es decir, no en términos de funciones de distribución).

**Dem. (a)** Supóngase que $\mathcal{M}$ es compacta y sea $F_1, F_2, \cdots$ una sucesión en $\mathcal{M}$. Por el teorema de Helly 4.1.3, existe una subsucesión $F_{n_k}$ y una función de distribución $F$ tal que

$$F_{n_k}(a,b] \to F(a,b] \qquad \forall a,b \in \mathbb{R} \cap C(F). \tag{4.2.1}$$

Para probar que $F_{n_k} \xrightarrow{d} F$ debemos verificar que (4.2.1) también se cumple en $+\infty$ y en $-\infty$, es decir,

$$F_{n_k}(+\infty) - F_{n_k}(-\infty) \to F(+\infty) - F(-\infty) \qquad \text{cuando } k \to \infty.$$

Dado $\epsilon > 0$, sean $a, b \in \mathbb{R} \cap C(F)$ puntos tales que para todo $n \in \mathbb{N}$

$$F_n(\mathbb{R} - (a,b]) < \epsilon, \qquad F(\mathbb{R} - (a,b]) < \epsilon.$$

Si $x \in \mathbb{R} \cap C(F)$, escribimos

$$|(F_{n_k}(+\infty) - F_{n_k}(x)) - (F(+\infty) - F(x))|$$

$$\begin{aligned} \leq \;& |F_{n_k}(+\infty) - F_{n_k}(b)| + |F(+\infty) - F(b)| \\ + \;& |(F_{n_k}(b) - F_{n_k}(x)) - (F(b) - F(x))|. \end{aligned}$$

Cada uno de los dos primeros términos del lado derecho son menores que $\epsilon$, y el último término tiende a cero cuando $k \to \infty$. Por lo tanto,

$$F_{n_k}(+\infty) - F_{n_k}(x) \to F(+\infty) - F(x).$$

Similarmente se obtiene que

$$F_{n_k}(x) - F_{n_k}(-\infty) \to F(x) - F(-\infty),$$

y combinando estos resultados con (4.2.1), concluimos que $F_{n_k} \xrightarrow{d} F$.

**(b)** Para probar la implicación recíproca, razonemos por contradicción. Supongamos ahora que $\mathcal{M}$ es relativamente compacta, pero que no es compacta. Entonces existe $\epsilon > 0$ tal que para todo entero $n$ y alguna función de distribución denotada por $F_n \in \mathcal{M}$,

$$F_n(\mathbb{R} - (-n, n)) \geq \epsilon.$$

Puesto que $\mathcal{M}$ es relativamente compacta, existe una subsucesión $\{F_{n_k}\}$ de $\{F_n\}$, y una función de distribución $F$ tal que $F_{n_k} \xrightarrow{d} F$. Luego, puesto que $\mathbb{R} - (-n, n)$ es cerrado, se sigue del teorema de Portmanteau 2.1.3 que

$$\limsup_{k\to\infty} F_{n_k}(\mathbb{R} - (-n, n)) \leq F(\mathbb{R} - (-n, n))$$

es decir, $F(\mathbb{R} - (-n, n)) \geq \epsilon$ para toda $n$. Luego, cuando $n \to \infty$, se obtiene que $0 \geq \epsilon$, lo cual constituye una contradicción. ■

Algunas consecuencias del teorema de Prohorov son:

**Corolario 4.2.3.** *Sea $\{F_n\}$ una sucesión acotada de funciones de distribución sobre $\mathbb{R}$ y supóngase que $\{F_n\}$ es compacta.*

*(a) Si toda subsucesión débilmente convergente de $\{F_n\}$ converge a una función de distribución $F$, entonces $F_n \xrightarrow{d} F$.*

*(b) $\{F_n\}$ converge débilmente si y sólo si las funciones características*

$$\phi_n(t) = \int_{\mathbb{R}} \mathrm{e}^{itx}\, dF_n(x)$$

*convergen a un límite finito cuando $n \to \infty$ para todo $t \in \mathbb{R}$.*

**Dem.** **(a)** Si $\{F_n\}$ no converge débilmente a $F$, existe una función $f \in C_b(\mathbb{R})$ tal que

$$\int_{\mathbb{R}} f(x)\, dF_n(x) \nrightarrow \int_{\mathbb{R}} f(x)\, dF(x) \qquad \text{cuando } n \to \infty.$$

Es decir, existe $\epsilon > 0$ y un conjunto infinito $T$ de números enteros tal que

$$\left| \int_{\mathbb{R}} f(x)\ dF_n(x) - \int_{\mathbb{R}} f(x)\ dF(x) \right| \geq \epsilon \qquad \text{para todo } n \in T.$$

Por el teorema de Prohorov 4.2.2, existe una subsucesión $\{F_{n_k}\}$ de $\{F_n : n \in T\}$ tal que $\{F_{n_k}\}$ converge débilmente a una función de distribución $G$. Por hipótesis, $G = F$ y se tiene que

$$\int_{\mathbb{R}} f(x)\ dF_n(x) \to \int_{\mathbb{R}} f(x)\ dF(x) \qquad \text{cuando } n \to \infty$$

lo cual es una contradicción. Por lo tanto $F_n \xrightarrow{d} F$.

**(b)** Si $\{F_n\}$ converge débilmente, la convergencia de las funciones características es inmediata, pues las funciones $\cos tx$ y $\operatorname{sen} tx$ pertenecen a $C_b(\mathbb{R})$, y puesto que $\mathrm{e}^{itx} = \cos tx + \mathrm{i} \operatorname{sen} tx$.

Ahora supóngase que para toda $t \in \mathbb{R}$, $\phi_n(t)$ tiene un límite finito. Por el teorema de Prohorov, existe una subsucesión $\{F_{n_k}\}$ y una función de distribución $F$ tal que $F_{n_k} \xrightarrow{d} F$. Si la sucesión original no converge débilmente a $F$, se sigue de la parte (a) que existe una subsucesión $\{F_{m_k}\}$ y una función de distribución $G \neq F$ tal que $F_{m_k} \xrightarrow{d} G$. Pero, por hipótesis, $\int_{\mathbb{R}} \mathrm{e}^{itx}\ dF_{n_k}(x)$ e $\int_{\mathbb{R}} \mathrm{e}^{itx}\ dF_{m_k}(x)$ tienen el mismo límite cuando $k \to \infty$. Por lo tanto

$$\int_{\mathbb{R}} \mathrm{e}^{itx}\ dF(x) = \int_{\mathbb{R}} \mathrm{e}^{itx}\ dG(x) \qquad \forall t \in \mathbb{R},$$

lo cual implica que $F = G$, y esto es una contradicción. Luego $F_n \xrightarrow{d} F$. ■

## 4.3. Teorema de continuidad de Lévy

**Lema 4.3.1** (Desigualdad de truncamiento.). *Sea $F$ una función de distribución acotada sobre $R$, con función característica $\phi$. Entonces existe una constante $k > 0$ (independiente de $F$) tal que para todo $u > 0$, se tiene*

$$\begin{aligned} F(\mathbb{R} - (-1/u, 1/u)) &= \int_{|x| \geq 1/u} dF(x) \\ &\leq \tfrac{k}{u} \int_0^u (\phi(0) - \mathrm{Re}\phi(v))\ dv, \end{aligned}$$

*en donde* $\mathrm{Re}(z)$ *es la parte real del número complejo* $z$.

**Dem.**

$$\begin{aligned}
\frac{1}{u}\int_0^u(\phi(0)-\mathrm{Re}\phi(v))\ dv &= \frac{1}{u}\int_0^u\int_{-\infty}^{\infty}(1-\cos vx)\ dF(x)\ dv\\
&= \int_{-\infty}^{\infty}\frac{1}{u}\int_0^u(1-\cos vx)\ dv\ dF(x)\ \text{(Fubini)}\\
&= \int_{-\infty}^{\infty}(1-\operatorname{sen} ux/ux)\ dF(x)\\
&\geq \operatorname{ínf}_{|t|\geq 1}(1-\operatorname{sen} t/t)\int_{|x|\geq 1/u}\ dF(x)\\
&= \frac{1}{k}\int_{|x|\geq 1/u}\ dF(x).
\end{aligned}$$

De hecho, $\operatorname{ínf}_{|t|\geq 1}(1-\operatorname{sen} t/t) = 1-\operatorname{sen} 1 \geq 1/7$ (pues para cualquier $t$, con $|t|\geq 1$, $\operatorname{sen} t/t \leq \operatorname{sen} 1 < 6/7$: En efecto, como $\operatorname{sen} t/t$ es par, basta considerar $t \geq 0$. Si $t \geq 1{,}3$, $\operatorname{sen} t/t \leq |\operatorname{sen} t|/1{,}3 \leq 1/1{,}3 = 0{,}76923... < \operatorname{sen} 1 \approx 0{,}84$, mientras que para $1 \leq t \leq 1{,}3$, $\operatorname{sen} t/t$ es decreciente, probándose la afirmación), de manera que podríamos tomar $k = 7$. ∎

**Teorema 4.3.2** (Teorema de continuidad de Lévy.)**.** *Sea $\{F_n\}$ una sucesión acotada de funciones de distribución sobre $\mathbb{R}$, y sea $\{\phi_n\}$ la sucesión correspondiente de funciones características.*

*(a) Si $F_n \xrightarrow{d} F$, en donde $F$ es una función de distribución con f.c. $\phi$, entonces $\phi_n(t) \to \phi(t)$ para todo $t \in \mathbb{R}$.*

*(b) Recíprocamente, si $\phi_n(t) \to \phi(t)$ para todo $t \in \mathbb{R}$, en donde $\phi$ es una función (con valores en $\mathbb{C}$) continua en $t = 0$, entonces $\phi$ es la función característica de alguna función de distribución $F$, y además $F_n \xrightarrow{d} F$. (Nótese que si las $F_n$ son f.d.p., entonces $F$ es f.d.p.).*

**Dem.**

La parte (a) se sigue de la definición de convergencia débil. Para probar (b), demostraremos primero que, bajo las hipótesis dadas, $\{F_n\}$ es compacta. Por el lema 4.3.1, existe una constante $k > 0$ independiente de las $F_n$, tal que para todo $u > 0$

$$F_n(\mathbb{R}-(-1/u,1/u)) \leq \frac{k}{u}\int_0^u(\phi_n(0)-\mathrm{Re}\phi_n(v))\ dv,$$

y por convergencia dominada, el lado derecho converge a

$$h(u) = \frac{k}{u}\int_0^u(\phi(0)-\mathrm{Re}\phi(v))\ dv$$

cuando $n \to \infty$. Así, dado $\epsilon > 0$, escojamos $\delta > 0$ tal que $|\phi(0) - \mathrm{Re}\phi(v)| < \epsilon/(3k)$ para todo $v$, con $|v| \le \delta$. Eligiendo arbitrariamente $0 < u_0 \le \delta$, se tiene que $|h(u_0)| < \epsilon/2$. Por el teorema de la convergencia dominada de Lebesgue, existe $n_0 \in \mathbb{N}$ tal que

$$F_n(\mathbb{R} - (-1/u_0, 1/u_0)) \le \frac{k}{u_0} \int_0^{u_0} (\phi_n(0) - \mathrm{Re}\phi_n(v))\, dv \le \epsilon/2 + h(u_0) < \epsilon, \tag{4.3.1}$$

para todo $n \ge n_0$. Por otro lado, para $j = 1, \cdots, n_0 - 1$, tomemos $0 < u \le u_0$ suficientemente pequeño tal que

$$F_j(\mathbb{R} - (-1/u, 1/u)) < \epsilon.$$

Puesto que $\mathbb{R} - (-1/u, 1/u) \subset \mathbb{R} - (-1/u_0, 1/u_0)$, de esta última relación y de la ecuación (4.3.1) se sigue que dado $\epsilon > 0$, podemos escoger $u$ suficientemente pequeño tal que

$$F_n(\mathbb{R} - (-1/u, 1/u)) < \epsilon \qquad \forall n$$

lo que demuestra que $\{F_n\}$ es compacta. Luego, por la hipótesis presente, se sigue del corolario 4.2.3-(b), que $\{F_n\}$ converge débilmente a una función de distribución $F$, y por lo tanto, $\{\phi_n\}$ converge puntualmente a la f.c. de $F$. En consecuencia, como $\phi_n \to \phi$ puntualmente, concluimos que $\phi$ es la f.c. de $F$. ∎

**Observación 4.3.3.** *El teorema de continuidad de Lévy es quizás la herramienta más utilizada para demostrar convergencia débil. Esencialmente, el teorema nos dice que para verificar que $F_n \xrightarrow{d} F$, basta verificar que las correspondientes funciones características $\phi_n$ convergen a la función característica de $F$. Usaremos esta técnica para demostrar el teorema del límite central (TLC) en el siguiente capítulo. Otras aplicaciones se pueden ver en la lista de ejercicios de este capítulo.*

## 4.4. Ejercicios

1. Sean $X_1, X_2, \cdots$ vv.aa. i.i.d. con distribución uniforme sobre el intervalo $(0,1)$; es decir, para cada $n = 1, 2, \cdots$, $X_n$ tiene la distribución $F$

$$\begin{aligned} F(x) &= 0 \quad \text{si } x \leq 0 \\ &= x \quad \text{si } 0 \leq x \leq 1 \\ &= 1 \quad \text{si } x \geq 1. \end{aligned}$$

   Sea $S_n = X_1 + \cdots + X_n$. Use el teorema de Lévy para demostrar que

$$\frac{S_n - n/2}{\sqrt{n/12}} \xrightarrow{d} X, \qquad \text{cuando } n \to \infty,$$

   donde $X \sim N(0,1)$.

2. Sea $\mu_n$ la medida de probabilidad concentrada en el punto $x_n \in \mathbb{R}$ (es decir, $\mu_n(B) = 1$ si $x_n \in B$, 0 en caso contrario). Demuestre que $\mu_n \xrightarrow{d} \mu$ si y sólo si $x_n \to x$, y que $\mu$ debe ser la medida de probabilidad concentrada en $x$.

3. Sea $\Omega = [0,1]$, y sea $\mu_n$ la medida discreta que asigna la masa $1/(n+1)$ a cada uno de los puntos $0, 1/n, 2/n, \cdots, n/n$. Demuestre que $\mu_n \xrightarrow{d} \mu$, en donde $\mu$ es la medidad de Lebesgue restringida a $\Omega$.

4. Use el teorema de Lévy para demostrar el *teorema límite de Poisson*: Para cada $n \in \mathbb{N}$, sean $X_i^{(n)}$ vv.aa. independientes tales que

$$\mathbb{P}(X_i^{(n)} = 1) = 1 - \mathbb{P}(X_i^{(n)} = 0) = p_i^{(n)}, \qquad i = 1, 2, \cdots, k(n).$$

   Supóngase que

$$\sum_{i=1}^{k(n)} p_i^{(n)} \to \mu; \qquad \max_{1 \leq i \leq k(n)} p_i^{(n)} \to 0$$

   cuando $n \to \infty$. Entonces $X_1^{(n)} + \cdots X_{k(n)}^{(n)} \xrightarrow{d} X$, donde $X$ es una v.a. de Poisson con parámetro $\mu$.

# Capítulo 5

# EL TEOREMA LÍMITE CENTRAL

Sean $X_1, X_2, \cdots$ vv.aa. independientes, en donde cada $X_k$ tiene media y varianza finitas $m_k$ y $\sigma_k^2$, respectivamente. Sea $S_n = X_1 + \cdots + X_n$, $n = 1, 2, \cdots$, de manera que

$$\mathbb{E}S_n = \sum_{k=1}^{n} m_k, \qquad c_n^2 := \operatorname{var}(S_n) = \sum_{k=1}^{n} \sigma_k^2.$$

Consideremos la suma normalizada

$$T_n := (S_n - \mathbb{E}S_n)/c_n \qquad \forall n = 1, 2, \cdots,$$

la cual tiene media 0 y varianza 1. Para evitar casos degenerados, supondremos que $c_n > 0$ para todo $n$ suficientemente grande. Sea $X^*$ un v.a. con distribución normal $N(0,1)$, es decir, la f.d.p. de $X^*$ es

$$F^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}, \qquad x \in \mathbb{R}.$$

En este capítulo daremos condiciones bajo las cuales $T_n \xrightarrow{d} X^*$.

Los resultados sobre la convergencia $T_n \xrightarrow{d} X^*$ tienen una larga historia y se conocen con el nombre genérico de *teorema límite central* (TLC). El resultado más antiguo de este tipo se conoce actualmente como el teorema de De Moivre-Laplace y se refiere al caso en el que $X_1, X_2, \cdots$ son vv.aa.

i.i.d. de Bernoulli con parámetro $p$, digamos

$$\begin{aligned} X_n &= 1 \quad \text{con probabilidad } p \ (0 < p < 1) \\ &= 0 \quad \text{con probabilidad } 1 - p. \end{aligned}$$

En esta situación, se tiene que $S_n = X_1 + \cdots + X_n$ es una v.a. binomial con parámetros $n$ y $p$, de modo que $\mathbb{E}S_n = np$, y $c_n^2 = npq$, donde $q := 1 - p$, y tenemos:

**Teorema 5.0.1** (Teorema de De Moivre-Laplace.)**.** *Si $X_1, X_2, \cdots$ son vv.aa. i.i.d. de Bernoulli con parámetro $p$, entonces*

$$T_n := (S_n - np)/\sqrt{npq} \xrightarrow{d} X^* \qquad \textit{cuando } n \to \infty,$$

*es decir*

$$\lim_{n\to\infty} \mathbb{P}(T_n \leq x) = F^*(x), \qquad \textit{para todo } x \in \mathbb{R}.$$

Este teorema lo demostró A. De Moivre, alrededor de 1730, para $p = 1/2$, y Laplace, en 1812, para $0 < p < 1$. A continuación demostraremos una versión más general del teorema 5.0.1, pero que, a su vez, es un caso especial del teorema de Lindeberg 5.2.1 que consideraremos después.

## 5.1. Teorema Límite Central

**Teorema 5.1.1** (Teorema de Límite Central (caso i.i.d.).)**.** *Sean $X_1, X_2, \cdots$ variables i.i.d. con media $\mu$ y varianza $\sigma^2$ finitas ($\sigma^2 > 0$). Entonces*

$$T_n = \frac{(S_n - n\mu)}{\sigma\sqrt{n}} \xrightarrow{d} X^*,$$

*en donde $S_n = X_1 + \cdots + X_n$, $X^* \sim N(0, 1)$.*

**Dem.** Usando el teorema de continuidad de Lévy 4.3.2, demostraremos que la función característica $\Phi_n(t)$ de la v.a. $T_n$ converge, cuando $n \to \infty$, a la f.c. de $X^*$, a saber, $\mathrm{e}^{-t^2/2}$. Por principio de cuentas, para simplificar la demostración, podemos suponer sin pérdida de generalidad, que $\mu = 0$ y $\sigma^2 = 1$ (en caso contrario, aplicamos el teorema a las vv.aa. $Y_n = (X_n - \mu)/\sigma$, las cuales, por supuesto, tienen media 0 y varianza 1). Así pues, $T_n = S_n/\sqrt{n}$, y si $\phi$ es la f.c. común de $X_1, X_2, \cdots$, entonces la f.c. $\phi_n$ de $T_n$ está dada por

$$\begin{aligned} \Phi_n(t) &= \mathbb{E}\exp[itT_n] \\ &= \textstyle\prod_{k=1}^n \mathbb{E}\exp[i(\frac{t}{\sqrt{n}})X_k] \\ &= \phi(t/\sqrt{n})^n, \; n = 1, 2, \cdots \end{aligned}$$

Puesto que las vv.aa. $X_k$ tienen segundo momento finito, se sigue del ehjercicio 3 del capítulo de funciones características, que las derivadas $\phi'$, $\phi''$ son continuas y que

$$\phi'(0) = i\mathbb{E}(X_1) = 0, \qquad \phi''(0) = i^2\mathbb{E}(X_1^2) = -\sigma^2 = -1.$$

Luego usando la expansión de Taylor para $\phi$, obtenemos

$$\begin{aligned} \phi(t) &= 1 + \phi'(0)t + \tfrac{1}{2}\phi''(0)t^2 + \circ(t^2) \\ &= 1 - t^2/2 + \circ(t^2), \end{aligned}$$

y por lo tanto $\phi_n(t) = \phi(t/\sqrt{n})^n$ resulta:

$$\phi_n(t) = (1 - t^2/2n + \circ(t^2/n))^n$$

para cada $t$ fijo. De la última expresión, concluimos que

$$\phi_n(t) \longrightarrow \mathrm{e}^{-t^2/2}, \qquad \text{cuando } n \to \infty$$

para todo $t \in \mathbb{R}$. ■

**Observación 5.1.2.** *Para el caso de variables i.i.d. se tiene la siguiente* estimación de Cramer-Berry-Essen *sobre la rapidez de convergencia en el TLC:*

*Sean $X_1, X_2, \cdots$, variables i.i.d. con media 0, varianza $\sigma^2$ y tercer momento $\mathbb{E}|X_1|^3 < \infty$. Entonces existe una constante $C > 0$ tal que*

$$|F_n(x) - F^*(x)| \leq \frac{C\mathbb{E}|X_1|^3}{\sigma^3\sqrt{n}}, \qquad \textit{para todo } x \in \mathbb{R},$$

*en donde $F_n$ es la f.d.p. de $T_n$, y $F^*$ es la f.d.p. de $X^* \sim N(0,1)$. La demostración de este resultado se puede ver en Gnedenko y Kolmogorov (1968) [9, pág. 201]. En esta misma referencia se puede leer que, en* 1941*, Berry "demostró"que $C = 1{,}88$. Sin embargo, posteriormente, en* 1945*, P.L. Hsu encontró que la "demostración"de Berry era incorrecta. En todo caso, se sabe que $C < 3$.*

*Ahora pasaremos al siguiente resultado, demostrado por Y.W. Lindeberg en* 1922.

## 5.2. Teorema de Lindeberg

**Teorema 5.2.1** (Teorema de Lindeberg.)**.** *Sean $X_1, X_2, \cdots$ vv.aa. independientes con medias y varianzas finitas $m_k$ y $\sigma_k^2$, respectivamente. Sea*

$$T_n = \frac{(S_n - \mathbb{E}S_n)}{C_n},$$

*en donde $S_n = X_1 + \cdots + X_n$, $C_n = \mathrm{var}(S_n) = \sum_{k=1}^n \sigma_k^2$, y sea $F_k$ la f.d.p. de $X_k$. Si para cada $\epsilon > 0$ se cumple la* condición de Lindeberg*:*

$$L_n := C_n^{-2} \sum_{k=1}^{n} \int_{\{x:|x-m_k|\geq \epsilon C_n\}} (x-m_k)^2 \, dF_k(x) \to 0 \tag{5.2.1}$$

*cuando $n \to \infty$, entonces $T_n \xrightarrow{d} X^*$, en donde $X^* \sim N(0,1)$.*

**Observación 5.2.2.** *Antes de dmostrar el teorema 5.2.1, veremos alguna de sus implicaciones, o casos especiales.*

(a) *Caso i.i.d. Sea $X_1, X_2, \cdots$ variables i.i.d. con media $m$ y varianza $\sigma^2$ finitas; véanse, por ejemplo, los teoremas 5.0.1 y 5.1.1. Entonces se satisface la condición de Lindeberg (5.2.1) y, por lo tanto, $T_n \xrightarrow{d} X^*$. En efecto, si denotamos por $F$ la f.d.p. de $X_k$, entonces, puesto que $m_k = m$ para $k = 1, 2, \cdots$, y $C_n^2 = n\sigma^2$, vemos que $L_n$ resulta*

$$\begin{aligned} L_n &= \tfrac{1}{n\sigma^2} \textstyle\sum_{k=1}^n \int_{\{x:|x-m|\geq \epsilon\sigma\sqrt{n}\}} (x-m)^2 \, dF(x) \\ &= \tfrac{1}{\sigma^2} \textstyle\int_{\{x:|x-m|\geq \epsilon\sigma\sqrt{n}\}} (x-m)^2 \, dF(x) \to 0 \end{aligned}$$

*cuando $n \to \infty$, pues $\sigma^2 < \infty$, y $\{x : |x-m| \geq \epsilon\sigma\sqrt{n}\} \downarrow \emptyset$ cuando $n \to \infty$.*

(b) *Caso uniformemente acotado Sean $X_1, X_2, \cdots$ vv.aa. independientes tales que $|X_k| \leq M$ c.p.1. para todo $k$, y supóngase que $C_n \to \infty$. Entonces $T_n \xrightarrow{d} X^*$.*

*Verifiquemos que en este caso se cumple la condición de Lindeberg (5.2.1): Puesto que*

$$\begin{aligned} \textstyle\int_{\{x:|x-m_k|\geq \epsilon C_n\}} (x-m_k)^2 \, dF_k(x) &= \mathbb{E}\left\{(X_k - m_k)^2 1_{[|X_k - m_k|\geq \epsilon C_n]}\right\} \\ &\leq (2M)^2 \mathbb{P}(|X_k - m_k| \geq \epsilon C_n) \\ &\leq \tfrac{(2M)^2 \sigma_k^2}{(\epsilon C_n)^2} \ (\textit{desigualdad de Chebyshev}), \end{aligned}$$

*de aquí se sigue que*

$$\begin{aligned} L_n &\leq \tfrac{1}{C_n^2} \sum_{k=1}^n \tfrac{(2M)^2 \sigma_k^2}{(\epsilon C_n)^2} \\ &= \tfrac{(2M)^2}{\epsilon^2 C_n^2} \longrightarrow 0 \quad \textit{cuando } n \to \infty. \end{aligned}$$

(c) *Condición de Lyapunov. Sean $X_1, X_2, \cdots$ vv.aa. independientes con medias $m_k$ y varianzas $\sigma_k^2$ finitas. Si se satisface la condición de Lyapunov: para algún $\delta > 0$, se tiene que*

$$\frac{1}{C_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}|X_k - m_k|^{2+\delta} \longrightarrow 0 \qquad \textit{cuando } n \to \infty, \tag{5.2.2}$$

*entonces se satisface la condición de Lindeberg 5.2.1 y, por lo tanto, $T_n \xrightarrow{d} X^*$. En efecto, dado $\epsilon > 0$*

$$\begin{aligned} \mathbb{E}|X_k - m_k|^{2+\delta} &= \textstyle\int_{-\infty}^{\infty} |x - m_k|^{2+\delta} \, dF_k(x) \\ &\geq \textstyle\int_{\{x:|x-m_k|\geq \epsilon C_n\}} |x - m_k|^{\delta} |x - m_k|^2 \, dF_k(x) \\ &\geq \epsilon^{\delta} C_n^{\delta} \textstyle\int_{\{x:|x-m_k|\geq \epsilon C_n\}} (x - m_k)^2 \, dF_k(x), \end{aligned}$$

*y en consecuencia,*

$$\begin{aligned} L_n &= C_n^{-2} \textstyle\sum_{k=1}^n \int_{\{x:|x-m_k|\geq \epsilon C_n\}} (x - m_k)^2 \, dF_k(x) \\ &\leq \tfrac{1}{\epsilon^{\delta} C_n^{2+\delta}} \textstyle\sum_{k=1}^n \mathbb{E}|X_k - m_k|^{2+\delta}. \end{aligned}$$

*De esta última desigualdad vemos que (5.2.2) implica (5.2.1).*

**Demostración del teorema de Lindeberg 5.2.1**. Sin pérdida de generalidad, podemos suponer que $m_k = 0$ para todo $k$. Usaremos las siguientes estimaciones: Si $y$ es cualquier número real y $z$ es un número complejo con $|z| \geq 1/2$, entonces

$$\begin{array}{lll} (a) & \mathrm{e}^{iy} & = 1 + iy + \theta y^2/2 \\ (b) & \mathrm{e}^{iy} & = 1 + iy - y^2/2 + \theta_1 |y|^3/6 \\ (c) & \ln(1+z) & = z + \theta' |z|^2, \end{array} \tag{5.2.3}$$

en donde $\theta$ y $\theta_1$ dependen de $y$, con $|\theta| \leq 1$, $|\theta_1| \leq 1$, y $\theta'$ depende de $z$, con $|\theta'| \leq 1$; ln denota la rama principal del logaritmo.

Sea ahora $\phi_k$ la función característica de $X_k$, y sea $t$ un número real . La función característica $g_n$ de $T_n = S_n/C_n$ es

$$g_n(t) = \prod_{k=1}^{n} \phi_k(t/C_n).$$

Deseamos probar que

$$g_n(t) \longrightarrow \mathrm{e}^{-t^2/2}$$

o equivalentemente que

$$t^2/2 + \ln g_n(t) = t^2/2 + \sum_{k=1}^{n} \ln \phi_k(t/C_n) \longrightarrow 0 \tag{5.2.4}$$

cuando $n \to \infty$. Primero usaremos 5.2.3 (a), (b) para escribir $\phi_k(t)$ como sigue:

$$\begin{array}{rcl}
\phi_k(t) & = & \int_{-\infty}^{\infty} \mathrm{e}^{itx}\ dF_k(x) \\
 & = & \int_{|x|\geq \epsilon C_n} (1 + itx + \theta t^2 x^2/2)\ dF_k(x) \\
 & + & \int_{|x|<\epsilon C_n} (1 + itx - t^2x^2/2 + \theta_1 |t|^3|x|^3/6)\ dF_k(x) \\
 & & (\text{con } \theta \text{ y } \theta_1 \text{ dependiendo de } tx, \text{ y } |\theta|,\ |\theta_1| \leq 1) \\
 & = & 1 + (t^2/2)\int_{|x|\geq\epsilon C_n} \theta x^2\ dF_k(x) - (t^2/2)\int_{|x|<\epsilon C_n} x^2\ dF_k(x) \\
 & + & (|t|^3/6)\int_{|x|<\epsilon C_n} \theta_1 |x|^3\ dF_k(x),
\end{array}$$

puesto que $\int_{-\infty}^{\infty} 1 \cdot dF_k(x) = 1$, $\int_{-\infty}^{\infty} itx\ dF_k(x) = it\mathbb{E}(X_k) = 0$.

Por lo tanto,

$$\begin{array}{rcl}
\phi_k(\frac{t}{C_n}) & = & 1 + (\frac{t^2}{2C_n^2})\int_{|x|\geq\epsilon C_n} \theta x^2\ dF_k(x) - (\frac{t^2}{2C_n^2})\int_{|x|<\epsilon C_n} x^2\ dF_k(x) \\
 & + & (\frac{|t|^3}{6C_n^3})\int_{|x|<\epsilon C_n} \theta_1 |x|^3\ dF_k(x).
\end{array} \tag{5.2.5}$$

Ahora, puesto que

$$\left| \frac{1}{2}\int_{|x|\geq\epsilon C_n} \theta x^2\ dF_k(x) \right| \leq \frac{1}{2}\int_{|x|\geq\epsilon C_n} x^2\ dF_k(x),$$

podemos escribir

$$\begin{array}{rcl}
\frac{1}{2}\int_{|x|\geq\epsilon C_n} \theta x^2\ dF_k(x) & = & \theta_2 \int_{|x|\geq\epsilon C_n} x^2\ dF_k(x) \\
 & = & \theta_2 C_n^2 \alpha_{nk},
\end{array}$$

con $|\theta_2| \leq 1/2$, y en donde

$$\alpha_{nk} = \frac{1}{C_n^2} \int_{|x| \geq \epsilon C_n} x^2 \, dF_k(x).$$

Análogamente, tenemos:

$$\begin{aligned} \left| \tfrac{1}{6} \int_{|x|<\epsilon C_n} \theta_1 |x|^3 \, dF_k(x) \right| &\leq \tfrac{1}{6} \int_{|x|<\epsilon C_n} |x|^3 \, dF_k(x) \\ &\leq \tfrac{1}{6} \int_{|x|<\epsilon C_n} \epsilon C_n x^2 \, dF_k(x), \end{aligned}$$

y en consecuencia, podemos escribir

$$\begin{aligned} \tfrac{1}{6} \int_{|x|<\epsilon C_n} \theta_1 |x|^3 \, dF_k(x) &= \theta_3 \int_{|x|<\epsilon C_n} \epsilon C_n x^2 \, dF_k(x) \\ &= \theta_3 \epsilon C_n^3 \beta_{nk} \end{aligned}$$

con $|\theta_3| \leq 1/6$, y en donde

$$\beta_{nk} = \frac{1}{C_n^2} \int_{|x|<\epsilon C_n} x^2 \, dF_k(x) \qquad (\leq \epsilon^2).$$

Luego, la ecuación (5.2.5) resulta

$$\phi_k(t/C_n) = 1 + \mu_{nk}, \tag{5.2.6}$$

en donde

$$\mu_{nk} = \theta_2 t^2 \alpha_{nk} - \frac{t^2}{2} \beta_{nk} + |t|^3 \theta_3 \epsilon \beta_{nk}. \tag{5.2.7}$$

Ahora, por hipótesis, $L_n \sum_{k=1}^n \alpha_{nk} \to 0$ cuando $n \to \infty$, y por otra parte,

$$\sum_{k=1}^n (\alpha_{nk} + \beta_{nk}) = \frac{1}{C_n^2} \sum_{k=1}^n \sigma_k^2 = 1.$$

Luego, como $\beta_{nk} \leq \epsilon^2$, se sigue de (5.2.7) que

$$\max_{1 \leq k \leq n} |\mu_{nk}| \leq t^2 \epsilon^2 / 2 + |t|^3 \epsilon$$

y

$$\sum_{k=1}^n |\mu_{nk}| \leq t^2/2 + |t|^3 \epsilon$$

para todo $n$ suficientemente grande. Por otro lado, usando (5.2.3)-(c) y (5.2.6), podemos escribir el lado izquierdo de (5.2.4) como

$$t^2/2 + \sum_{k=1}^{n} \ln(1+\mu_{nk}) = t^2/2 + \sum_{k=1}^{n} (\mu_{nk} + \theta'|\mu_{nk}|^2),$$

con $|\theta'| \leq 1$. Finalmente, de (5.2.7) y usando el hecho de que

$$\sum_{k=1}^{n} |\mu_{nk}|^2 \leq \max_{1\leq k\leq n} |\mu_{nk}| \sum_{k=1}^{n} |\mu_{nk}|,$$

concluimos que dado cualquier $\delta > 0$ , podemos seleccionar $\epsilon > 0$ (el cual depende de $\delta$ y $t$) tal que

$$\left| t^2/2 + \sum_{k=1}^{n} \ln(1+\mu_{nk}) \right| < \delta$$

para todo $n$ suficientemente grande, lo cual implica (5.2.4), y por lo tanto, $T_n \xrightarrow{d} X^*$. ■

## 5.3. Teorema de Lindeberg-Feller

**Definición 5.3.1.** *Para dar una idea de lo que significa la condición de Lindeberg, introduciremos el concepto de* despreciabilidad asintótica uniforme *(d.a.u.). Sean $X_1, X_2, \cdots$ vv.aa. como en el teorema de Lindeberg 5.2.1. Decimos que las vv.aa. $(X_k - m_k)/C_n$ satisfacen la condición de d.a.u si para cualquier $\epsilon > 0$,*

$$\max_{1\leq k\leq n} \mathbb{P}(|X_k - m_k|/C_n \geq \epsilon) \longrightarrow 0 \qquad \textit{cuando } n \to \infty.$$

*Intuitivamente, esto significa que la contribución de cada término $(X_k - m_k)/C_n$ es pequeña comparada con la suma $T_n = (S_n - \mathbb{E}S_n)/C_n$.*

**Proposición 5.3.2.** *La condición de Lindeberg implica la condición de d.a.u.*

**Dem.**

$$\begin{array}{rcl} L_n & = & C_n^{-2} \sum_{k=1}^n \int_{\{x:|x-m_k|\geq \epsilon C_n\}} (x-m_k)^2 \, dF_k(x) \\ & \geq & C_n^{-2} \epsilon^2 C_n^2 \sum_{k=1}^n \mathbb{P}(|X_k - m_k| \geq \epsilon C_n) \\ & \geq & \epsilon^2 \max_{1\leq k\leq n} \mathbb{P}(|X_k - m_k|/C_n \geq \epsilon). \blacksquare \end{array}$$

**Observación 5.3.3.** *La condición de Lindeberg no es necesaria para convergencia a $X^*$. Para ver esto daremos un ejemplo de una sucesión de vv.aa. que no satisfacen la condición de d.a.u., pero que, sin embargo, $T_n \xrightarrow{d} X^*$. Sean $X_1, X_2, \cdots$ vv.aa. independientes, normalmente distribuidas con media 0 y varianzas $\sigma_1^2 = 1$, $\sigma_k^2 = 2^{k-2}$, $k \geq 2$. Entonces $C_n^2 = \sum_{k=1}^n \sigma_k^2 = 2^{n-1}$, y $X_n/C_n$ es normal con media cero y varianza $1/2$. Por lo tanto*

$$\max_{1\leq k\leq n} \mathbb{P}(|X_k|/C_n \geq \epsilon) \geq \mathbb{P}(|X_n|/C_n \geq \epsilon) = 2(1 - F^*(\epsilon\sqrt{2})),$$

*en donde $F^*$ es la f.d.p. de $X^* \sim N(0,1)$. Puesto que el último término es una constante positiva que no depende de $n$, concluimos que las vv.aa. $X_k/C_n$ no satisfacen la condición de d.a.u. Sin embargo, para todo $n$, $T_n = S_n/C_n$ es una v.a. normal $N(0,1)$, es decir, $T_n \sim X^*$ para todo $n$, de manera que $T_n \xrightarrow{d} X^*$.*

*Aunque la condición de Lindeberg no es necesaria para convergencia a $X^*$, resulta que si añadimos la condición de d.a.u. tenemos el siguiente resultado que enunciamos sin demostración.*

**Teorema 5.3.4.** *Sean $X_1, X_2, \cdots$ vv.aa. independientes con medias $m_k$ y varianza $\sigma_k^2$ finitas. Entonces la condición de Lindeberg 5.2.1 se satisface si y sólo si $T_n \xrightarrow{d} X^*$ y las vv.aa. $(X_k - m_k)/C_n$ satisfacen la condición de d.a.u.*

La suficiencia del teorema precedente se sigue del teorema de Lindeberg 5.2.1 y la proposición 5.3.2. La demostración de la necesidad se puede ver, por ejemplo, en Ash [1], Tucker [16] o Gnedenko y Kolmogorov [9].

## 5.4. Ejercicios

1. Sean $X_1, X_2, \cdots$ sucesiones arbitrarias de vv.aa. Dado cualquier número real $c$, muestre que siempre se pueden encontrar constantes $a_n$ y $b_n$ $(a_n > 0)$ de tal forma que $a_n^{-1}(X_1 + \cdots + X_n - b_n)$ convergen en distribución a la v.a. $X \equiv c$.

2. Sea $\{X_{nk} : n = 1, 2, \cdots,\ k = 1, \cdots, n\}$ una sucesión doble de vv.aa., y sea $h_{nk}$ la función característica de $X_{nk}$. Muestre que los $X_{nk}$ satisfacen la condición de d.a.u., es decir,

$$\max_{1\leq k\leq n} \mathbb{P}(|X_{nk}| \geq \epsilon) \to 0$$

cuando $n \to \infty$, para cada $\epsilon > 0$, si y solamente si

$$\max_{1\leq k\leq n} |h_{nk}(u) - 1| \to 0$$

cuando $n \to \infty$, y en este caso, la convergencia es uniforme sobre cada intervalo acotado.

*Indicación: Puede usar la desigualdad de truncamiento establecida en el lema 4.3.1.*

3. Sean $X_1, X_2, \cdots$ vv.aa. independientes , definidas como sigue:

$X_1 = \pm 1$, con igual probabilidad.

Si $k > 1$, y $c$ es un número real fijo mayor que 1,

$$\begin{aligned}
\mathbb{P}(X_k = 1) &= \quad \mathbb{P}(X_k = -1) &&= \frac{1}{2c} \\
\mathbb{P}(X_k = k) &= \quad \mathbb{P}(X_k = -k) &&= \frac{1}{2k^2}\left(1 - \frac{1}{c}\right) \\
\mathbb{P}(X_k = 0) &= 1 - \frac{1}{c} - \frac{1}{k^2}\left(1 - \frac{1}{c}\right).
\end{aligned}$$

De aquí definimos

$$X'_{nk} := \begin{cases} X_k & \text{si} \quad |X_k| \leq \sqrt{n}, \\ 0 & \text{si} \quad |X_k| > \sqrt{n}. \end{cases}$$

Establezca lo siguiente:

(a) Las $X_k/c_n$ satisfacen la condición de d.a.u., donde $c_n^2 := \mathrm{var}(S_n)$, con $S_n := X_1 + \cdots + X_n$.

(b) La condición de Lindeberg falla para las $X_k$, pero se cumple para las $X'_{nk}$. Más aún, si $S'_n := \sum_{k=1}^n X'_{nk}$, entonces $S'_n/c'_n \xrightarrow{d} N(0,1)$, donde ${c'_n}^2 := \mathrm{var}(S'_n) \sim_{n\to\infty} n/c$

(c) $\mathbb{P}(S_n \neq S'_n) \to 0$ cuando $n \to \infty$.

(d) $\sqrt{c}S_n/\sqrt{n} \xrightarrow{d} N(0,1)$, pero $S_n/\sqrt{n} \xrightarrow{d} N(0,1)$ no se tiene.

# Capítulo 6

# ESPERANZA CONDICIONAL

En este capítulo estudiaremos el concepto de *esperanza condicional* y sus propiedades más importantes. Este concepto extiende el de *probabilidad condicional*, siendo una herramienta teórica importante en el desarrollo de la teoría de probabilidades. En nuestro caso será útil en la definición de martingala.

Sean $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad, $C \in \mathfrak{F}$, con $\mathbb{P}(C) > 0$ . En los cursos básicos de probabilidad se define la *probabilidad condicional respecto o relativa a* $C$ como la medida de probabilidad $\mathbb{P}(\cdot|C) : \mathfrak{F} \longrightarrow [0,1]$ tal que

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} \quad \forall A \in \mathfrak{F}. \tag{6.0.1}$$

La integral respecto a esta medida de probabilidad es llamada la *"esperanza condicional relativa a* $C$*"* y la denotamos por $\mathbb{E}[\cdot|C]$. Así, si $X$ es v.a. integrable respeto a $\mathbb{P}(\cdot|C)$, entonces

$$\mathbb{E}[X|C] = \int_\Omega X(\omega)\mathbb{P}(d\omega|C).$$

**Observación 6.0.1.** *Notemos que*

$$\begin{aligned} \mathbb{P}(A|C) &= \frac{1}{\mathbb{P}(C)} \int_C I_A \quad d\mathbb{P} \quad \forall A \in \mathfrak{F}, \\ \mathbb{E}[X|C] &= \frac{1}{\mathbb{P}(C)} \int_C X \quad d\mathbb{P}. \end{aligned}$$

*Si $X$ es $\mathbb{P}$-integrable, entonces $X$ es $\mathbb{P}(\cdot|C)$-integrable.*
*En caso de que $\mathbb{P}(C) = 0$ convenimos en poner $\mathbb{E}[X|C] \equiv 0$.*

Ahora bien, sea $\mathcal{C}$ la sub $\sigma$-álgebra de $\mathfrak{F}$ generada por una partición numera-ble $\{C_n\}_n$ de conjuntos medibles de $\Omega$ tal que $\mathbb{P}(C_n) > 0$ para cada $n$ (por ejemplo, si $Y$ es v.a. discreta con valores $\{c_n\}_n$, $C_n := [Y = c_n]$ es una partición de $\Omega$ de conjuntos medibles y $\mathcal{C} = \sigma(Y)$).

Dada $X$ una v.a. $\mathbb{P}$-integrable, definimos la v.a. $\mathbb{E}[X|\mathcal{C}] : \Omega \longrightarrow \mathbb{R}$ por:

$$\begin{aligned}
\mathbb{E}[X|\mathcal{C}](\omega) &= \sum_{i=1}^{\infty} \mathbb{E}[X|C_i] I_{C_i}(\omega) \\
&= \sum_{i=1}^{\infty} \left( \frac{1}{\mathbb{P}(C_i)} \int_{C_i} X \quad d\mathbb{P} \right) I_{C_i}(\omega).
\end{aligned}$$

Puesto que $X$ es $\mathbb{P}$-integrable, se tiene que $\mathbb{E}[X|\mathcal{C}]$ es también $\mathbb{P}$-integrable, más aún

$$\begin{aligned}
\int_{\Omega} \mathbb{E}[X|\mathcal{C}] \quad d\mathbb{P} &= \sum_{i=1}^{\infty} \left( \frac{1}{\mathbb{P}(C_i)} \int_{C_i} X \quad d\mathbb{P} \right) \int_{\Omega} I_{C_i} \quad d\mathbb{P} \\
&= \int_{\Omega} X \quad d\mathbb{P}.
\end{aligned}$$

es decir, $\mathbb{E}[\mathbb{E}[X|\mathcal{C}]] = \mathbb{E}X$.
Más aún, si $C \in \mathcal{C}$, entonces $C$ es la unión numerable de una subfamilia de la partición $\{C_n\}_n$, $C = \cup_{i\in\Lambda} C_i$, luego

$$\begin{aligned}
\int_{C} \mathbb{E}[X|\mathcal{C}] \quad d\mathbb{P} &= \sum_{i\in\Lambda} \mathbb{E}[X|C_i] \int_{C} I_{C_i}(\omega) \quad d\mathbb{P} \\
&= \sum_{i\in\Lambda} \mathbb{E}[X|C_i]\mathbb{P}(C_i) \\
\text{por la observación 6.0.1} \quad &= \sum_{i\in\Lambda} \left( \frac{1}{\mathbb{P}(C_i)} \int_{C_i} X \quad d\mathbb{P} \right) \mathbb{P}(C_i) \\
&= \sum_{i\in\Lambda} \int_{C_i} X \quad d\mathbb{P} = \int_{\bigcup_{i\in\Lambda} C_i} X \quad d\mathbb{P} \\
&= \int_{C} X \quad d\mathbb{P}.
\end{aligned}$$

Resumiendo todo lo anterior, tenemos que la v.a. $\mathbb{E}[X|\mathcal{C}]$ satisface

(a) $\mathbb{E}[X|\mathcal{C}]$ es $\mathcal{C}$-medible.

(b) $\int_C \mathbb{E}[X|\mathcal{C}] \quad d\mathbb{P} = \int_C X \quad d\mathbb{P} \quad \forall C \in \mathcal{C}$.

El inciso $(b)$ establece que el promedio ó valor esperado de la v.a. $X$ sobre los eventos $C \in \mathcal{C}$ se puede obtener como el valor esperado de la v.a. $\mathcal{C}$-medible, $\mathbb{E}[X|\mathcal{C}]$ sobre los eventos $C \in \mathcal{C}$, lo que es de suma importancia, puesto que el comportamiento de $X$ sobre $\mathcal{C}$ puede ser obtenido a través de una v.a. "minimal" que contenga la mínima información requerida, es decir, que sea $\mathcal{C}$-medible.

**Observación 6.0.2.** *Si $Y$ es otra v.a. $\mathcal{C}$-medible integrable tal que*

$$\int_C Y \quad d\mathbb{P} = \int_C X \quad d\mathbb{P} \quad \forall C \in \mathcal{C}$$

*entonces $Y = \mathbb{E}[X|\mathcal{C}]$ c.p.1. (ó c.s.), así pues $\mathbb{E}[X|\mathcal{C}]$ es única salvo un conjunto de probabilidad cero.*

En lo que sigue extenderemos este concepto a un marco general de tal forma que tenga sentido $\mathbb{E}[X|\mathcal{C}]$, donde $\mathcal{C}$ sea cualquier sub $\sigma$-álgebra de $\mathfrak{F}$. Para ello necesitamos del teorema de Radon-Nikodym.

## 6.1. Teorema de Radon-Nikodym

En esta sección enunciaremos el teorema de Radon-Nikodym que establece condiciones bajo las cuales se puede tener la existencia de funciones de densidad para medidas absolutamente continuas respecto a una medida de probabilidad dada.

Extendemos el concepto de esperanza o promedio de una magnitud aleatoria respecto a un conjunto de condiciones o magnitudes aleatorias, haciendo uso del teorema de Radon-Nikodym, permitiendo formalizar y poner bajo un marco teórico preciso el concepto de esperanza condicional.

**Definición 6.1.1.** *Sean $(\Omega, \mathfrak{F})$ un espacio medible, $\nu : \mathfrak{F} \longrightarrow \mathbb{R}$ una función. Decimos que $\nu$ es una* medida signada finita *sobre $(\Omega, \mathfrak{F})$ si*

*(i)* $|\nu(\Omega)| < +\infty$,

*(ii)* $\nu(\emptyset) = 0$,

*(iii)* $\nu$ *es aditivamente numerable, es decir, si* $\{A_n\}_n$ *es una familia numerable disjunta de conjuntos, elementos de la* $\sigma$*-álgebra de* $\mathfrak{F}$*, entonces*

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \nu(A_n).$$

**Ejemplo 6.1.2.** *Sean* $(\Omega, \mathfrak{F}, \mu)$ *espacio medido,* $X : \Omega \longrightarrow \mathbb{R}$ *una función integrable respecto a* $\mu$*. Definimos* $\nu : \mathfrak{F} \longrightarrow \mathbb{R}$ *como*

$$\nu(F) = \int_F X \quad d\mu = \int_F X^+ \quad d\mu - \int_F X^- d\mu \quad \forall F \in \mathfrak{F}.$$

*Entonces* $\nu$ *es una medida signada finita.*

*En efecto, aplicando el teorema de Beppo-Levi para series de funciones medibles positivas, a los integrandos* $X^+$ *y* $X^-$*, vemos que* $\nu$ *es aditivamente numerable. También es claro que* $\nu(\emptyset) = 0$*. Así,* $\nu$ *es una medida signada finita sobre* $(\Omega, \mathfrak{F})$*.*

**Definición 6.1.3.** *Sea* $(\Omega, \mathfrak{F}, \mu)$ *un espacio medido,* $\nu$ *una medida signada finita sobre* $(\Omega, \mathfrak{F})$*. Decimos que* $\nu$ es absolutamente continua a la medida $\mu$ *y lo denotamos por*

$$\nu << \mu,$$

*si vale la implicación*

$$A \in \mathfrak{F} \ , \quad \mu(A) = 0 \Longrightarrow \nu(A) = 0.$$

**Ejemplo 6.1.4.** *En el ejemplo 6.1.2, la medida signada* $\nu$ *es absolutamente continua respecto a* $\mu$*. Así, la medida signada finita del ejemplo 6.1.2 es una medida signada absolutamente continua respecto a* $\mu$*.*

Se tiene la siguiente caracterización de continuidad absoluta:

**Proposición 6.1.5.** *Una condición necesaria y suficiente para que* $\nu << \mu$ *es que para cada* $\epsilon > 0$ *existe* $\delta > 0$ *tal que para cada* $A \in \mathfrak{F}$*, con* $\mu(A) < \delta$*, entonces* $\nu(A) < \epsilon$*.*

La demostración de esta proposición la podemos encontrar, por ejemplo, en [4]. De la proposición 6.1.5 y el ejemplo 6.1.2 se desprende el siguiente corolario.

**Corolario 6.1.6.** ***(Continuidad absoluta ó uniforme de la integral)*** *Si $X$ es integrable respecto a $\mu$, entonces para cada $\epsilon > 0$ existe $\delta > 0$ tal que*

$$\mu(A) < \delta \Longrightarrow \int_A |X| \quad d\mu < \epsilon.$$

Tenemos una especie de recíproco al ejemplo 6.1.4 anterior en el siguiente teorema conocido como el *teorema de Radon-Nikodym*.

**Teorema 6.1.7.** ***(de Radon-Nikodym)*** *Sea $(\Omega, \mathfrak{F}, \mu)$ un espacio de medida finita ó $\sigma$-finita, $\nu$ una medida signada finita sobre $(\Omega, \mathcal{C})$, (donde $\mathcal{C}$ es una sub $\sigma$-álgebra de $\mathfrak{F}$) tal que $\nu << \mu$. Entonces existe una v.a. $g : \Omega \longrightarrow \mathbb{R}$, $\mathcal{C}$-medible, de esperanza finita tal que*

$$\nu(F) = \int_F g \quad d\mu \quad \forall F \in \mathcal{C}.$$

*La función $g$ es única salvo igualdad $\mu$-casi en todas partes, y se llama* la derivada de Radon-Nikodym *de $\nu$ respecto a $\mu$, denotándola por $\frac{d\nu}{d\mu}$.*

Las demostración del teorema 6.1.7 se encuentra por ejemplo en [1, 4, 5, 15].

Gracias a este resultado, cobra sentido la definición dada en cursos precedentes sobre variableas aleatorias *absolutamente continuas*, el cual volvemos a recoger en la siguiente definición:

**Definición 6.1.8.** *Sean $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad, una variable aleatoria $X : \Omega \longrightarrow \mathbb{R}^n$ se dice ser absolutamente continua si la distribución de probabilidad $\mu_X$ es absolutamente continua respecto a la medida de Lebesgue de $\mathbb{R}^n$, es decir, si existe $f_X : \mathbb{R}^n \longrightarrow [0, +\infty)$ función Borel-medible tal que $\int_{\mathbb{R}^n} f_X(u_1, \cdots, u_n) \quad du_1 \cdots du_n = 1$ y*

$$\mu_X(B) = \mathbb{P}(X \in B) = \int_B f_X(u_1, \cdots, u_n) \quad du_1 \cdots du_n \quad \forall B \in \mathcal{B}(\mathbb{R}^n).$$

*Notemos que la función de distribución asociada a $X$ viene dada por*

$$F_X(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(u_1, \cdots, u_n) \quad du_1 \cdots du_n.$$

*A $f_X$ se le llama función de densidad de $X$.*

## 6.2. Esperanzas condicionales

La definición de esperanza condicional puede ser desarrollada como sigue.

Sean $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad, $\mathcal{C}$ una sub-$\sigma$-álgebra de $\mathfrak{F}$, $X \in \mathbf{L}^1$ (no necesariamente $\mathcal{C}$-medible). Definimos

$$\nu(C) = \int_C X \quad d\mathbb{P} = \mathbb{E}[XI_C] \quad \forall C \in \mathcal{C}. \tag{6.2.1}$$

Claramente $\nu$ es una medida signada sobre $(\Omega, \mathfrak{F})$ la cual es absolutamente continua respecto a $\mathbb{P}$. Por el teorema de Radon-Nikodym (ver teorema 6.1.7) existe $Y$ función $\mathcal{C}$-medible tal que

$$\int_C Y \quad d\mathbb{P} = \int_C X \quad d\mathbb{P} \quad \forall C \in \mathcal{C}.$$

Dicha función es única salvo un conjunto de probabilidad cero. A $Y$ la denotaremos por $\mathbb{E}[X|\mathcal{C}]$ y es llamada *la esperanza condicional de $X$ bajo la sub $\sigma$-álgebra* $\mathcal{C}$ de tal forma que

$$\nu(C) = \int_C \mathbb{E}[X|\mathcal{C}] \quad d\mathbb{P} \quad \forall C \in \mathcal{C}.$$

En resumen, la esperanza condicional de $X$ dado $\mathcal{C}$, $\mathbb{E}[X|\mathcal{C}]$, es una variable aleatoria $\mathcal{C}$-medible y está únicamente determinada salvo un conjunto de probabilidad cero, satisfaciendo

$$\int_C X \quad d\mathbb{P} = \int_C \mathbb{E}[X|\mathcal{C}] \quad d\mathbb{P} \quad \forall C \in \mathcal{C}.$$

En lo que sigue estableceremos otras propiedades importantes de esperanza condicional.

**Teorema 6.2.1.** *Sean $X$ una v.a. sobre el espacio medido $(\Omega, \mathfrak{F}, \mathbb{P})$, $\mathcal{C}$ una sub $\sigma$-álgebra de $\mathfrak{F}$. Las siguientes propiedades son válidas c.p.1. (o casi seguramente):*

*(a) Si $\mathcal{C} = \{\emptyset, \Omega\}$, entonces $\mathbb{E}[X|\mathcal{C}] = \mathbb{E}X$.*

*(b) Si $X \geq 0$, entonces $\mathbb{E}[X|\mathcal{C}] \geq 0$.*

*(c) Si $X$ es $\mathcal{C}$-medible, entonces $\mathbb{E}[X|\mathcal{C}] = X$.*

*(d) Si $X$ =constante= $a$, entonces $\mathbb{E}[X|\mathcal{C}] = a$.*

*(e) Si $X, Y \in \mathbf{L}^1$, entonces $\mathbb{E}[aX + bY|\mathcal{C}] = a\mathbb{E}[X|\mathcal{C}] + b\mathbb{E}[Y|\mathcal{C}]$ $a, b \in \mathbb{R}$.*

*(f) Si $X \leq Y$, entonces $\mathbb{E}[X|\mathcal{C}] \leq \mathbb{E}[Y|\mathcal{C}]$.*

*(g) Sean $X, Y, XY \in \mathbf{L}^1$, $X$ variable aleatoria $\mathcal{C}$-medible, entonces*

$$\mathbb{E}[XY|\mathcal{C}] = X\mathbb{E}[Y|\mathcal{C}].$$

*En particular $\mathbb{E}[\mathbb{E}[X|\mathcal{C}]Y|\mathcal{C}] = \mathbb{E}[X|\mathcal{C}]\mathbb{E}[Y|\mathcal{C}]$.*

*(h) Si $\mathcal{H}$ es independiente de las $\sigma$-álgebras $\sigma(X)$ y $\mathcal{C}$, entonces*

$$\mathbb{E}[X|\sigma(\mathcal{C}, \mathcal{H})] = \mathbb{E}[X|\mathcal{C}].$$

*En particular si $X, \mathcal{H}$ son independientes, entonces $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}X$.*

*(i) Sean $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathfrak{F}$ sub $\sigma$-álgebras de $\mathfrak{F}$, entonces $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_2]|\mathcal{C}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{C}_1]|\mathcal{C}_2] = \mathbb{E}[X|\mathcal{C}_1]$.*

**Dem.**

(a) Sea $h : \Omega \longrightarrow \mathbb{R}$ función constante tal que $h(\omega) = \mathbb{E}[X]$ para cada $\omega \in \Omega$. Veamos que $h$ es $\mathcal{C}$-medible; en efecto:

$$\{\omega \in \Omega : h(\omega) \geq r\} = \begin{cases} \Omega & , si \quad \mathbb{E}[X] \geq r \\ \emptyset & , si \quad \mathbb{E}[X] < r \end{cases}$$

$\forall r \in \mathbb{R}$, y $\{\Omega, \emptyset\} \subset \mathcal{C}$. Como $h$ es $\mathcal{C}$-medible, resta probar que

$$\int_C X \, d\mathbb{P} = \int_C h \, d\mathbb{P} \quad \forall C \in \mathcal{C} = \{\emptyset, \Omega\}.$$

En efecto , si $C = \Omega$

$$\mathbb{E}[X] = \int_\Omega X \, d\mathbb{P} = \int_\Omega h \, d\mathbb{P} = \mathbb{E}[X] \, \mathbb{P}(\Omega) = \mathbb{E}[X]$$

Si $C = \emptyset$

$$\int_\emptyset X \, d\mathbb{P} = 0 = \int_\emptyset h \, d\mathbb{P}$$

Por lo tanto

$$\int_C X \, d\mathbb{P} = \int_C h \, d\mathbb{P} \text{ para cada } C \in \mathcal{C}.$$

Así por unicidad de la esperanza condicional $\mathbb{E}[X|\mathcal{C}] = \mathbb{E}[X]$.

(b) Sea $h = \mathbb{E}[X|\mathcal{C}]$, como $X \geq 0$ entonces

$$\int_C h \, d\mathbb{P} = \int_C X \, d\mathbb{P} \geq 0 \quad \forall C \in \mathcal{C}.$$

Queremos ver que $h \geq 0$, en efecto, tomemos $C_1 = \{\omega \in \Omega : h(\omega) < 0\} \in \mathcal{C}$. Entonces

$$0 \leq \int_{C_1} h \, d\mathbb{P} \leq 0 \quad \Longrightarrow \quad \int_{C_1} h \, d\mathbb{P} = 0.$$

Note que

$$hI_{C_1} = hI_{\{\omega\in\Omega:h(\omega)<0\}} = -h^-$$

entonces

$$\int_{C_1} h \, d\mathbb{P} = \int_{C_1} -h^- = 0.$$

Como $h^- \geq 0$, necesariamente se tiene que $h^- = 0$ c. s. (c.p.1.), se sigue que $-h^- = 0$ c. s. y $h = h^+ - h^- = h^+$ de aquí que $h \geq 0$.

(c) Este inciso se sigue de la unicidad casi seguramente de la esperanza condicional.

(d) Note que $\int_C a \, d\mathbb{P} = \int_C X \, d\mathbb{P} = \int_C \mathbb{E}[X|\mathcal{C}] \, d\mathbb{P}$ y sabemos que la función constante es una función $\mathcal{C}$-medible, para toda $\mathcal{C}$ $\sigma$-álgebra; por unicidad de la esperanza condicional, $a = \mathbb{E}[X|\mathcal{C}]$.

(e) Note que

$$\begin{aligned}
\int_C \mathbb{E}[\alpha X + \beta Y|\mathcal{C}] \, d\mathbb{P} &= \int_C (\alpha X + \beta Y) \, d\mathbb{P} \\
&= \int_C \alpha X \, d\mathbb{P} + \int_C \beta Y \, d\mathbb{P} \\
&= \alpha \int_C \mathbb{E}[X|\mathcal{C}] \, d\mathbb{P} + \beta \int_C \mathbb{E}[Y|\mathcal{C}] \, d\mathbb{P} \\
&= \int_C (\alpha \mathbb{E}[X|\mathcal{C}] + \beta \mathbb{E}[Y|\mathcal{C}]) \, d\mathbb{P}.
\end{aligned}$$

La funciones $\alpha\mathbb{E}[X|\mathcal{C}]+\beta\mathbb{E}[Y|\mathcal{C}]$ y $\mathbb{E}[\alpha X+\beta Y|\mathcal{C}]$ son $\mathcal{C}$-medibles y por unicidad de la esperanza condicional

$$\mathbb{E}[\alpha X+\beta Y|\mathcal{C}]=\alpha\mathbb{E}[X|\mathcal{C}]+\beta\mathbb{E}[Y|\mathcal{C}] \text{ c.s. ó c.p.1.}$$

(f) Definamos $f:=Y-X$, $f\geq 0$. Usando (b) y (e) se sigue inmediatamente el resultado.

(g) Sea $\nu(C)=\int_C Y\ d\mathbb{P}=\int_C \mathbb{E}[Y|\mathcal{C}]\ d\mathbb{P}\quad \forall C\in\mathcal{C}$. Siendo $X$ $\mathcal{C}$-medible, por un teorema de cambio de variable, así como la definición de esperanza condicional, tenemos

$$\int_C \mathbb{E}[XY|\mathcal{C}]\ d\mathbb{P}=\int_C XY\ d\mathbb{P}=\int_C X\ d\nu=\int_C X\mathbb{E}[Y|\mathcal{C}]\ d\mathbb{P}\quad \forall C\in\mathcal{C}.$$

Por ser $X,\mathbb{E}[X|\mathcal{C}]$ funciones $\mathcal{C}$-medibles, se tiene que $\mathbb{E}[XY|\mathcal{C}]=X\mathbb{E}[Y|\mathcal{C}]$ con probabilidad uno.

(h) Supongamos primero que $X\geq 0$ y $\mathbb{E}[X]<+\infty$.
Notemos que $\sigma(\mathcal{C},\mathcal{H})$ es generada por el $\pi$-sistema formado por los conjuntos de la forma $C\cap H$, $C\in\mathcal{C}$, $H\in\mathcal{H}$. Por inciso (b) arriba $\mathbb{E}[X|\sigma(\mathcal{C},\mathcal{H})]\geq 0$ y $\mathbb{E}[X|\mathcal{C}]\geq 0$. Definamos las medidas finitas positivas

$$\begin{aligned}\mu(D)\ &:=\ \int_D \mathbb{E}[X|\sigma(\mathcal{C},\mathcal{H})]\ d\mathbb{P}\\ \nu(D)\ &:=\ \int_D \mathbb{E}[X|\mathcal{C}]\ d\mathbb{P}\quad \forall D\in\sigma(\mathcal{C},\mathcal{H})\end{aligned}$$

claramente

$$\mu(\Omega)=\nu(\Omega)=\int_\Omega X\ d\mathbb{P}=\mathbb{E}[X].$$

Por independencia de $\mathcal{H}$ respecto a $\sigma(X)$ y $\mathcal{C}$, tenemos que para cada

$C \in \mathcal{C}$, $H \in \mathcal{H}$

$$\begin{aligned}
\mu(C \cap H) &= \int_{C\cap H} \mathbb{E}[X|\sigma(\mathcal{C},\mathcal{H})]\ d\mathbb{P} \\
&= \int_{C\cap H} X\ d\mathbb{P} = \left(\int_C X\ d\mathbb{P}\right)\left(\int_H\ d\mathbb{P}\right) \\
&= \left(\int_C \mathbb{E}[X|\mathcal{C}]\ d\mathbb{P}\right)\int_H\ d\mathbb{P} \\
&= \int_{C\cap H} \mathbb{E}[X|\mathcal{C}]\ d\mathbb{P} = \nu(C\cap H).
\end{aligned}$$

Como consecuencia del lema de las clases de Dynkin (ver, por ejemplo, [4, Sección 1.6], se sigue que

$$\int_D \mathbb{E}[X|\sigma(\mathcal{C},\mathcal{H})]\ d\mathbb{P} = \int_D \mathbb{E}[X|\mathcal{C}]\ d\mathbb{P} \quad \forall D \in \sigma(\mathcal{C},\mathcal{H}).$$

De esta igualdad y del hecho de que $\mathbb{E}[X|\mathcal{C}]$ es $\mathcal{C}$-medible y $\mathcal{C} \subset \sigma(\mathcal{H},\mathcal{C})$ luego $\mathbb{E}[X|\mathcal{C}]$ es $\sigma(\mathcal{C},\mathcal{H})$-medible, por univocidad de la esperanza condicional se sigue

$$\mathbb{E}[X|\sigma(\mathcal{C},\mathcal{H})] = \mathbb{E}[X|\mathcal{C}].$$

Para el caso general de $X$ una v.a. arbitraria tal que $\mathbb{E}[X] < +\infty$ basta poner $X = X^+ - X^-$, considerar el inciso (e) precedente, y lo hecho para el caso de v.a. positivas:

$$\begin{aligned}
\mathbb{E}[X|\sigma(\mathcal{C},\mathcal{H})] &= \mathbb{E}[X^+ - X^-|\sigma(\mathcal{C},\mathcal{H})] \\
&= \mathbb{E}[X^+|\sigma(\mathcal{C},\mathcal{H})] - \mathbb{E}[X^-|\sigma(\mathcal{C},\mathcal{H})] \\
&= \mathbb{E}[X^+|\mathcal{C}] - \mathbb{E}[X^-|\mathcal{C}] \\
&= \mathbb{E}[X|\mathcal{C}].
\end{aligned}$$

(i) Sea $Y = \mathbb{E}[X|\mathcal{C}_2]$ y $C_1 \in \mathcal{C}_1$. Note que

$$\begin{aligned}\int_{C_1} \mathbb{E}[Y|\mathcal{C}_1]\ d\mathbb{P} &= \int_{C_1} Y\ d\mathbb{P} \\ &= \int_{C_1} \mathbb{E}[X|\mathcal{C}_2]\ d\mathbb{P} \\ &= \int_{C_1} X\ d\mathbb{P} \text{ (ya que } C_1 \in \mathcal{C}_2) \\ &= \int_{C_1} \mathbb{E}[X|\mathcal{C}_1]\ d\mathbb{P}.\end{aligned}$$

Por unicidad de esperanza condicional, dado que $\mathbb{E}[Y|\mathcal{C}_1], \mathbb{E}[X|\mathcal{C}_1]$ son $\mathcal{C}_1$-medibles, entonces $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_2]|\mathcal{C}_1] = \mathbb{E}[X|\mathcal{C}_1]$.
La igualdad $\mathbb{E}[\mathbb{E}[X|\mathcal{C}_1]|\mathcal{C}_2] = \mathbb{E}[X|\mathcal{C}_1]$ se sigue del inciso (c), pues $\mathbb{E}[X|\mathcal{C}_1]$ es $\mathcal{C}_2$ medible ya que $\mathcal{C}_1 \subset \mathcal{C}_2$. ■

Las demostraciones que hemos hecho son estándares, y podemos encontrarlas por ejemplo en [1, 5, 16].

**Observación 6.2.2.** *Sean $X \in \mathbf{L}_1(\Omega, \mathfrak{F}, P)$, $\mathfrak{C}$ una sub $\sigma$-álgebra de $\mathfrak{F}$. Si $\mathbb{E}[X|\mathfrak{C}] \leq \mathbb{E}[X]$ c.p.1., entonces $\mathbb{E}[X|\mathfrak{C}] = \mathbb{E}[X]$ c.p.1.*
*En efecto:*
*Sea $Y = \mathbb{E}[X] - \mathbb{E}[X|\mathfrak{C}] \geq 0$ c.s., notemos que*

$$\begin{aligned}\int_\Omega Y\ d\mathbb{P} &= \int_\Omega \mathbb{E}[X]\ d\mathbb{P} - \int_\Omega \mathbb{E}[X|\mathfrak{C}]\ d\mathbb{P} \\ &= \mathbb{E}[X]\mathbb{P}(\Omega) - \int_\Omega X\ d\mathbb{P} \\ &= 0 \ c.s.\end{aligned}$$

*Siendo $Y \geq 0$ c.p.1., se sigue sin más que $Y = 0$ c.s., es decir, $\mathbb{E}[X|\mathfrak{C}] = \mathbb{E}[X]$ c.p.1.*

A continuación veremos un teorema que será útil para variables aleatorias discretas.

**Teorema 6.2.3.** *Sean $\mathfrak{C}$ una sub $\sigma$-álgebra de $\mathfrak{F}$ generada por eventos disjuntos $\{B_1, B_2, ..\}$, donde $\mathbb{P}(B_n) > 0$ para cada $n \in \mathbb{N}$, tal que $\Omega = \cup_{n=1}^{\infty} B_n$, entonces para cada $n$ y para cada $\omega \in B_n$, $\mathbb{E}[X|\mathfrak{C}](\omega) = \frac{\mathbb{E}[XI_{B_n}]}{\mathbb{P}(B_n)}$.*

**Dem.** Sea $W = \sum_{j=1}^{\infty} \frac{\mathbb{E}[XI_{B_i}]}{\mathbb{P}(B_i)} I_{B_i}$. Sea $C \in \mathfrak{C}$ arbitrario, entonces existe una subsucesión de índices $\{n_j\}$ tal que $C = \bigcup_j B_{n_j}$. De aquí

$$\begin{aligned}\int_C W \, d\mathbb{P} &= \sum_j \int_{B_{n_j}} W \, d\mathbb{P} = \sum_j \int I_{B_{n_j}} W \, d\mathbb{P} \\ &= \sum_j \mathbb{E}[XI_{B_{n_j}}] = \mathbb{E}[XI_C] = \int_C X \, d\mathbb{P}.\end{aligned}$$

Puesto que $W$ es $\mathfrak{C}$-medible entonces $W = \mathbb{E}[X|\mathfrak{C}]$ c.p.1. ■

Como una aplicación inmediata de la esperanza condicional, podemos extender la noción de probabilidad condicional respecto a un evento, a la probabilidad condicional respecto a una $\sigma$-álgebra como sigue:

**Definición 6.2.4.** *Sean $\mathcal{C} \subset \mathcal{F}$ una sub $\sigma$-álgebra de $\mathfrak{F}$, $A$ un evento, elemento de $\mathfrak{F}$. Definimos la probabilidad condicional de $A$ respecto a la $\sigma$-álgebra $\mathcal{C}$, $\mathbb{P}(A|\mathcal{C})$, como sigue*

$$\mathbb{P}(A|\mathcal{C}) := \mathbb{E}[I_A|\mathcal{C}].$$

Como la esperanza condicional es una función $\mathcal{C}$-medible, entonces la probabilidad condicional es una función $\mathcal{C}$-medible.

En particular, si $\mathcal{C}$ es numerablemente generada por una familia $\{A_n\}_n$ de eventos disjuntos a pares tal que $\mathbb{P}(A_n) > 0$ para cada $n$, tenemos que

$$\mathbb{P}(A|\mathcal{C})(\omega) = \sum_{n=1}^{\infty} \frac{\mathbb{P}(A \cap A_n)}{\mathbb{P}(A_n)} I_{A_n}(\omega).$$

Notemos que estamos extendiendo la noción de probabilidad condicional dada en (6.0.1) a través de una variable aleatoria $\mathcal{C}$-medible.
En seguida mencionaremos algunas propiedades de la probabilidad condicional, cuyas demostraciones se encuentran en las referencias ya citadas [1, 5, 16].

**Propiedades de la probabilidad condicional**

(a) $\mathbb{P}(A|\mathcal{C}) \geq 0$ para todo $A \in \mathfrak{F}$.

(b) $\mathbb{P}(\Omega|\mathcal{C}) = 1$, $\mathbb{P}(\emptyset|\mathcal{C}) = 0$.

(c) Sean $A_1, A_2, ...$ eventos disjuntos en $\mathfrak{F}$, entonces

$$\mathbb{P}(\bigcup_n A_n|\mathcal{C}) = \sum_n \mathbb{P}(A_n|\mathcal{C}).$$

### 6.2.1. Formas condicionales de los teoremas de la convergencia monótona, convergencia dominada y lema de Fatou

En esta parte se establecerán las versiones análogas del teorema de la convergencia monótona, lema de Fatou, teorema de la convergencia dominada de Lebesgue para esperanzas condicionales; herramientas que son necesarias para la teoría que se estudiará en el siguiente capítulo. Usaremos los siguientes textos como referencia, de los autores: Ash [1], Dudley [5] y Tucker [16].

**Teorema 6.2.5.** ***(Forma condicional del teorema de la convergencia monótona.)*** *Sea $\{X_n\}_n$ una sucesión creciente de variables aleatorias positivas tal que $X_n \xrightarrow[n\to\infty]{c.s.} X$, con $X, X_n \in \mathbf{L}^1$ para cada $n$. Entonces para cualquier $\mathcal{C}$ sub $\sigma$-álgebra de $\mathfrak{F}$*

$$\mathbb{E}[X_n|\mathcal{C}] \xrightarrow[n\to\infty]{c.s.} \mathbb{E}[X|\mathcal{C}].$$

**Dem.** Como $0 \leq X_n \leq X_{n+1} \leq X$ para cada $n \in \mathbb{N}$, y $X$ con esperanza finita, entonces $X_n$ tiene esperanza finita para cada $n \in \mathbb{N}$. Por el inciso (f) del teorema 6.2.1, tenemos

$$0 \leq \mathbb{E}[X_n|\mathcal{C}] \leq \mathbb{E}[X_{n+1}|\mathcal{C}] \leq \mathbb{E}[X|\mathcal{C}] \quad \forall n \text{ c. s.}$$

entonces $\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}]$ existe c.s. y además es $\mathcal{C}$-medible. Resta demostrar que para cada $C \in \mathcal{C}$

$$\int_C (\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}])\, d\mathbb{P} = \int_C \mathbb{E}[X|\mathcal{C}]\, d\mathbb{P}.$$

Sea $Y = \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}]$ y $Y_n = \mathbb{E}[X_n|\mathcal{C}]$, por el teorema de la convergencia monótona de Beppo-Levi

$$\begin{aligned}
\int_C Y \, d\mathbb{P} &= \lim_{n\to\infty} \int_C Y_n \, d\mathbb{P} = \lim_{n\to\infty} \int_C \mathbb{E}[X_n|\mathcal{C}] \, d\mathbb{P} \\
&= \lim_{n\to\infty} \int_C X_n \, d\mathbb{P} \text{ (por definición)} \\
&= \int_C X \, d\mathbb{P} \text{ (Teorema de Beppo-Levi)} \\
&= \int_C \mathbb{E}[X|\mathcal{C}] \, d\mathbb{P}
\end{aligned}$$

Por unicidad de la esperanza condicional se sigue que $\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}] = \mathbb{E}[X|\mathcal{C}]$. ■

**Teorema 6.2.6.** ***(Forma condicional del lema de Fatou.)***
*Sea $\{X_n\}_n$ una sucesión de variables aleatorias positivas en $\mathbf{L}^1$. Si $\mathcal{C}$ es una sub $\sigma$-álgebra de $\mathfrak{F}$, tal que $\mathbb{E}(\liminf_{n\to\infty} X_n) < +\infty$, entonces*

$$\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{C}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}] \quad \textit{c.p.1.}$$

**Dem.** Sea $Z_n = \inf_{k\geq n} X_k$. Así $Z_n$ es una sucesión creciente de variables aleatorias positivas, tales que

$$\liminf_n X_n = \lim_n Z_n \quad \text{c.p.1.}$$

Por el Teorema anterior (teorema 6.2.5) se tiene

$$\mathbb{E}[Z_n|\mathcal{C}] \xrightarrow[n\to\infty]{c.s.} \mathbb{E}[\liminf_n X_n|\mathcal{C}]. \tag{6.2.2}$$

Por otro lado, note que $Z_n \leq X_k$ para cada $k \geq n$. Por el inciso (f) del teorema 6.2.1

$$\mathbb{E}[Z_n|\mathcal{C}] \leq \mathbb{E}[X_k|\mathcal{C}].$$

Luego

$$\mathbb{E}[Z_n|\mathcal{C}] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{C}]. \tag{6.2.3}$$

De las relaciones (6.2.2) y (6.2.3) se sigue lo que se quiere demostrar. ■

**Teorema 6.2.7.** ***(Forma condicional del teorema de la convergencia dominada de Lebesgue.)*** *Sean $Y, \{X_n\}_n$ una sucesión de variables aleatorias, satisfaciendo $|X_n| \leq Y$ c.p.1. para cada $n$, con $Y, X_n \in \mathbf{L}^1$ para cada $n$. Supongamos además que $X_n \xrightarrow[n\to\infty]{c.s.} X$. Luego si $\mathcal{C}$ es una sub $\sigma$-álgebra de $\mathfrak{F}$, entonces*

$$\mathbb{E}[X_n|\mathcal{C}] \xrightarrow[n\to\infty]{c.s.} \mathbb{E}[X|\mathcal{C}].$$

**Dem.** Supongamos que las hipótesis anteriores se cumplen para toda $\omega \in \Omega$, y también que $Y \geq 0$. Entonces $\{Y + X_n\}_n$ es una sucesión de funciones medibles positivas tal que

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

Entonces por el teorema anterior (teorema 6.2.6)

$$\mathbb{E}[\liminf_{n\to\infty} (Y + X_n)|\mathcal{C}] \leq \liminf_{n\to\infty} \mathbb{E}[Y + X_n|\mathcal{C}] \text{ c. s.}$$

De aquí se sigue que

$$\mathbb{E}[X|\mathcal{C}] = \mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{C}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}] \text{ c. s.} \tag{6.2.4}$$

De manera similar para la sucesión $\{Y - X_n\}_n$ se demuestra que

$$\mathbb{E}[\liminf_{n\to\infty} (Y - X_n)|\mathcal{C}] \leq \liminf_{n\to\infty} \mathbb{E}[Y - X_n|\mathcal{C}] \text{ c. s.}$$

lo que implica que

$$\mathbb{E}[X|\mathcal{C}] = \mathbb{E}[\limsup_{n\to\infty} X_n|\mathcal{C}] \geq \limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{C}] \text{ c. s.} \tag{6.2.5}$$

De las relaciones (6.2.4) y (6.2.5) se tiene el resultado. ■

### 6.2.2. Forma condicional de la desigualdad de Jensen

En este apartado, daremos un breve repaso sobre las funciones convexas, para después dar lugar a la desigualdad de Jensen para esperanzas condicionales.

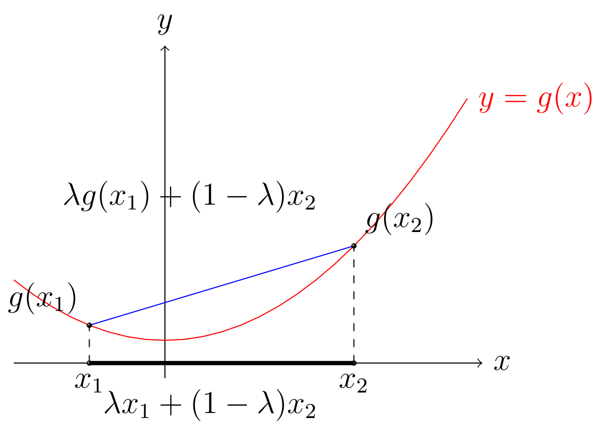


Figura 6.1: Representación geométrica de una función convexa.

**Definición 6.2.8.** *Sea $I$ un intervalo y $g : I \to \mathbb{R}$ una función. Se dice que $g$ es una* función convexa *si para cada $x, y \in I$, $\lambda \in (0,1)$ se satisface la desigualdad*

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y).$$

Enunciamos algunas propiedades de las funciones convexas, estudiadas en los cursos de cálculo y análisis. Un estudio sistemático de las funciones convexas se tiene en el texto de Royden [15].

**Propiedades de las funciones Convexas**

- Una función $g$ es cóncava si satisface la desigualdad contraria de la definición; es decir, si $g$ es convexa entonces $-g$ es cóncava.
- Toda combinación lineal con coeficientes positivos de funciones convexas es una función convexa.
- Toda función convexa es continua.
- Toda función convexa $g : (a,b) \longrightarrow \mathbb{R}$ es monótona sobre $\mathbb{R}$, ó bien existe un $x_0 \in (a,b)$ tal que $g$ es decreciente sobre $(a, x_0]$ y creciente sobre $[x_0, b)$.
- Si $g : (a,b) \longrightarrow \mathbb{R}$ es una función convexa, entonces para cada $x_0 \in (a,b)$ existe una recta $y = m(x - x_0) + g(x_0)$ que siempre está por

debajo de la gráfica de $g$; es decir,

$$m(x - x_0) + g(x_0) \leq g(x) \quad \forall x \in (a, b).$$

Dicha recta la llamaremos *recta de soporte en el punto* $x_0$ y siempre estará por debajo de la gráfica de $g$.

**Teorema 6.2.9.** ***(Forma condicional de la desigualdad de Jensen)*** *Sea* $g : \mathbb{R} \to \mathbb{R}$ *una función convexa, y sea* $X$ *una variable aleatoria tal que* $X, g(X) \in \mathbf{L}^1$. *Si* $\mathcal{C}$ *es una sub* $\sigma$*-álgebra de* $\mathfrak{F}$*, entonces*

$$g(\mathbb{E}[X|\mathcal{C}]) \leq \mathbb{E}[g(X)|\mathcal{C}] \quad c.s.$$

La demostración de este teorema podemos encontrarla por ejemplo en [5, 15, 16]. Sin embargo haremos un bosquejo de dicha prueba.

**Dem.** Considere la recta de soporte en el punto $x_0$ que debe estar por debajo de la gráfica de $g$ de tal forma que:

$$g(x_0) + m(x - x_0) \leq g(x) \quad \forall x \in \mathbb{R} \tag{6.2.6}$$

donde $m$ es la pendiente de dicha recta de soporte que pasa por el punto $(x_0, g(x_0))$.

Consideremos particularmente $x_0 := \mathbb{E}[X|\mathcal{C}]$ y $x = X$, sustituyendo en (6.2.6) obtenemos

$$g\left(\mathbb{E}[X|\mathcal{C}]\right) + m(X - \mathbb{E}[X|\mathcal{C}]) \leq g(X). \tag{6.2.7}$$

Tomando la esperanza condicional respecto a $\mathcal{C}$ de ambos lados de la desigualdad (6.2.7), para el lado izquierdo se tiene

$$g\left(\mathbb{E}[X|\mathcal{C}]\right) + m\mathbb{E}[\left(X - \mathbb{E}[X|\mathcal{C}]\right)|\mathcal{C}] = g\left(\mathbb{E}[X|\mathcal{C}]\right) \tag{6.2.8}$$

ya que $\mathbb{E}[(X - \mathbb{E}[X|\mathcal{C}])|\mathcal{C}] = 0$. Por otro lado al calcular la esperanza condicional del lado derecho de la desigualdad (6.2.7) se tiene $\mathbb{E}[g(X)|\mathcal{C}]$ y por lo tanto de las ecuaciones (6.2.7) y (6.2.8) se tiene que

$$g\left(\mathbb{E}[X|\mathcal{C}]\right) \leq \mathbb{E}[g(X)|\mathcal{C}].$$

■

**Consecuencia:** $\big|\mathbb{E}[X|\mathcal{C}]\big| \leq \mathbb{E}\big[|X|\big|\mathcal{C}\big]$.

**Notación:** Sean $X, Y$ variables aleatorias, $X$ con esperanza finita. Denotamos $\mathbb{E}[X|Y]$ a la esperanza condicional $\mathbb{E}[X|\sigma(Y)]$. En general, si $Y_1$,$Y_2$,...,$Y_n$ son variables aleatorias, $\mathbb{E}[X|Y_1, \cdots, Y_n]$ denotará a $\mathbb{E}[X|\sigma(Y_1, \cdots, Y_n)]$. Si $\{Y_\alpha\}_{\alpha\in\Lambda}$ es una familia de variables aleatorias, $\mathbb{E}[X|\{Y_\alpha\}_{\alpha\in\Lambda}]$ denotará a $\mathbb{E}[X|\sigma(\{Y_\alpha\}_{\alpha\in\Lambda})]$.

Concluimos este apartado considerando un caso particular para la existencia de *densidades condicionales*.

Sean $X \in \mathbf{L}^1$, $Y$ una variable aleatoria. Para cada $B \in \mathcal{B}(\mathbb{R})$, definamos a $\varphi$ por

$$\varphi(B) = \int_{[Y\in B]} X \quad d\mathbb{P} \tag{6.2.9}$$

Es claro que $\varphi$ es una medida signada finita boreliana. Consideremos la medida de distribución de $Y$, $\mu_Y$, es decir,

$$\mu_Y(B) = \mathbb{P}[Y \in B] \quad \text{para cada } B \in \mathcal{B}(\mathbb{R}).$$

Claramente $\varphi$ es una medida absolutamente continua respecto a $\mu_Y$ por lo que podemos aplicar el Teorema de Radon-Nikodym (teorema 6.1.7). Definimos $\mathbb{E}[X|Y = y]$ como la función Borel medible que está únicamente determinada excepto sobre un conjunto $A$ con $\mu_Y(A) = 0$ por

$$\varphi(B) = \int_B \mathbb{E}[X|Y = y] \quad d\mu_Y(y). \tag{6.2.10}$$

Notemos que $\mathbb{E}[X|Y = y]$ es la derivada de Radon-Nikodym de $\varphi$ con respecto a $\mu_Y$, es decir,

$$\mathbb{E}[X|Y = y] = \frac{d\varphi}{d\mu_Y}(y).$$

A $\mathbb{E}[X|Y = y]$ se llamará la *densidad condicional* de $X$ dada la condición $Y = y$.

Ahora mostraremos la conección entre esta definición de esperanza condicional y la precedente.

**Teorema 6.2.10.** *Sea* $X \in \mathbf{L}^1$ *y sea* $Y$ *una variable aleatoria. Si* $g$ *es la función definida en (6.2.10), es decir,* $g(y) = \mathbb{E}[X|Y=y]$*, entonces* $g(Y) = \mathbb{E}[X|Y]$ *c.p.1.*

**Dem.** Como $g$ es una función Borel medible y además tenemos que $g$, $\mathbb{E}[X|Y]$ son $\sigma(Y)$-medibles. Necesitamos demostrar

$$\int_C g(Y) \ \ d\mathbb{P} = \int_C \mathbb{E}[X|Y] \ \ d\mathbb{P} \quad \forall C \in \sigma(Y).$$

De la definición de $\mathbb{E}[X|Y]$, únicamente necesitamos demostrar que

$$\int_C g(Y) \ \ d\mathbb{P} = \int_C X \ \ d\mathbb{P} \quad \forall C \in \sigma(Y).$$

Para cada $C \in \sigma(Y)$ existe un $B \in \mathcal{B}(\mathbb{R})$ tal que

$$C = [Y \in B].$$

Entonces por (6.2.9) y (6.2.10)

$$\begin{aligned} \int_C g(Y) \ \ d\mathbb{P} &= \int_{[Y \in B]} g(Y) \ \ d\mathbb{P} \\ &= \int_B g(y) \ \ d\mu_Y(y) \\ &= \int_C X \ \ d\mathbb{P}. \end{aligned}$$

Por lo tanto $g(Y) = \mathbb{E}[X|Y]$ c.p.1. ■

De manera semejante podemos generalizar el resultado anterior estableciendo:

**Teorema 6.2.11.** *Sean* $(\Omega, \mathfrak{F}, \mathbb{P})$ *un espacio de probabilidad,*

$$Y_1, Y_2, ..., Y_n, X : \Omega \longrightarrow \mathbb{R}$$

*variables aleatorias. Supongamos que* $\mathbb{E}|X| < +\infty$*, entonces existe*

$$g : \mathbb{R}^n \longrightarrow \mathbb{R}$$

*una función Borel-medible tal que*

$$\mathbb{E}[X|Y_1, ..., Y_n] = g(Y_1, .., Y_n).$$

*Dicha $g$ es única $\mu_Y$-casi seguramente, donde $Y$ es el vector aleatorio definido como $Y = (Y_1, \cdots, Y_n)$, $\mu_Y = \mathbb{P} \circ Y^{-1}$.*
*Dicha $g$ se denotará por $g(y_1, .., y_n) = \mathbb{E}[X|Y_1 = y_1, ..., Y_n = y_n]$.*

**Dem.** Recuerde que $\sigma(Y) = Y^{-1}(\mathcal{B}(\mathbb{R}^n))$. Sea $B \in \mathcal{B}(\mathbb{R}^n)$ y sea

$$\nu(B) := \int_{Y^{-1}(B)} X \; d\mathbb{P} \text{ para cada } B \in \mathcal{B}(\mathbb{R}^n)$$

$\nu$ es medida signada sobre $\mathcal{B}(\mathbb{R}^n)$ y es finita pues $\mathbb{E}[|X|] < \infty$, veamos que $\nu << \mu_Y$, en efecto, sea $B_0 \in \mathcal{B}(\mathbb{R}^n)$ tal que $\mu_Y(B_0) = 0$ lo que implica que

$$\begin{aligned} 0 = \mu_Y(B_0) &= \mathbb{P}(Y^{-1}(B_0)) \\ &\Longrightarrow Y^{-1}(B_0) \text{ es } \mathbb{P} \text{ despreciable} \\ &\Longrightarrow \nu(B_0) = 0 \end{aligned}$$

por lo tanto, $\nu << \mu_Y$; por el teorema de Radon Nikodym (teorema 6.1.7) existe una única $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ tal que

$$\begin{aligned} \nu(B) &= \int_B g(y_1, .., y_n) \; \mu_Y d(y_1, .., y_n) \\ &= \int g(y_1, .., y_n) I_B(y_1, .., y_n) \; d\mathbb{P} \circ Y^{-1} \\ &= \int g(Y) I_B(Y) \; d\mathbb{P} \\ &= \int g(Y) I_{Y^{-1}(B)} \; d\mathbb{P} \\ &= \int_{Y^{-1}(B)} g(Y) \; d\mathbb{P} \end{aligned}$$

por lo tanto $\int_{Y^{-1}(B)} X \; d\mathbb{P} = \int_{Y^{-1}(B)} g(Y) \; d\mathbb{P}$. De aquí se sigue que

$$\int_{Y^{-1}(B)} \mathbb{E}[X|Y_1, ..., Y_n] \; d\mathbb{P} = \int_{Y^{-1}(B)} g(Y) \; d\mathbb{P}$$

es decir,

$$\int_C \mathbb{E}[X|Y_1, ..., Y_n] \; d\mathbb{P} = \int_C g(Y) \; d\mathbb{P} \text{ para cada } C \in \sigma(Y).$$

Como $\mathbb{E}[X|Y_1, ..., Y_n], g(Y)$ son $\sigma(Y)$-medibles, entonces

$$\mathbb{E}[X|Y_1, ..., Y_n] = g(Y) = g(Y_1, ..., Y_n).$$

■

**Ejemplo 6.2.12.** *Sean $X, Y : \Omega \longrightarrow \mathbb{R}$ variables aleatorias, $X$ función integrable, $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ absolutamente continua, con funciones de densidad $f_Y, f_{X,Y}$ respectivamente. Mediante un cálculo elemental usando la relación (6.2.10) obtenemos que*

$$\mathbb{E}[X|Y = y] = \frac{\int_{-\infty}^{+\infty} x f_{X,Y}(x, y) \quad dx}{f_Y(y)}$$

*para aquellas $y \in \mathbb{R}$ tal que $f_Y(y) > 0$.*

## 6.3. Ejercicios

1. Sea $X \in \mathbf{L}_1(\Omega, \mathfrak{F}, \mathbb{P})$ v.a. integrable, y sea $\mathfrak{B}$ una sub-sigma álgebra de $\mathfrak{F}$. Definamos

$$\varphi(B) := \int_B X \ d\mathbb{P}, \qquad \text{para todo } B \in \mathfrak{B}.$$

   Pruebe que $\varphi$ es una medida signada finita sobre $(\Omega, \mathfrak{B}, \mathbb{P})$, la cual es absolutamente continua con respecto a la restricción de $\mathbb{P}$ sobre $(\Omega, \mathfrak{B})$.

2. Sean $X, Y \in \mathbf{L}_1(\Omega, \mathfrak{B}, \mathbb{P})$. Pruebe que $X = Y$ c.s. si y sólo si

$$\int_B X \ d\mathbb{P} = \int_B Y \ d\mathbb{P}, \qquad \text{para todo } B \in \mathfrak{B}.$$

3. Pruebe que toda función convexa es continua.

4. Pruebe que si $g : \mathbb{R} \to \mathbb{R}$ es una función convexa, entonces ó bien $g$ es monótona, ó existe un punto $x_0 \in \mathbb{R}$ tal que $g$ es decreciente en el intervalo $(-\infty, x_0]$, y creciente en el intervalo $[x_0, \infty)$.

5. Sea $(\Omega, \mathfrak{A}, \mathbb{P})$ el espacio de probabilidad del intervalo unitario, y sea $\mathfrak{B}$ la sub-sigma álgebra generada por los intervalos $\{[0, \frac{1}{4}], (\frac{1}{4}, \frac{2}{3}], (\frac{2}{3}, 1]\}$, y sea $X(\omega) = \omega^2$ para todo $\omega \in \Omega = [0, 1]$. Demuestre que $\mathbb{E}[X|\mathfrak{B}]$ se puede expresar como

$$\mathbb{E}[X|\mathfrak{B}] = a_1 I_{[0,\frac{1}{4}]} + a_2 I_{(\frac{1}{4},\frac{2}{3}]} + a_3 I_{(\frac{2}{3},1]}.$$

   Calcule las constantes $a_1, a_2, a_3$.

6. Si $Y$ es una v.a. discreta tal que $\mathbb{P}(Y = y_n) > 0$ para cada $n = 1, 2, \cdots$, y $\sum_{n=1}^{\infty} \mathbb{P}(Y = y_n) = 1$. Dado $X \in \mathbf{L}_1(\Omega, \mathfrak{F}, \mathbb{P})$, definamos $\varphi : \mathbb{R} \to \mathbb{R}$ como

$$\varphi(y) = \begin{cases} \int_\Omega X \ d\mathbb{P}(\cdot|[Y = y_n]) & \text{si} \quad y = y_n; \\ \\ 0 & \text{si} \quad y \notin \{y_n\}. \end{cases}$$

   Pruebe que $\varphi(y) = \mathbb{E}[X|Y = y] \ \mu_Y$-c.s.

7. Pruebe las afirmaciones hechas en el ejemplo 6.2.12.

# Bibliografía


[1] ASH, R., DOLÉANS-DADE C., *Probability* & *Measure Theory*, Academic Press, Reimpreso 2008.

[2] BERGSTRÖM, H., *Weak Convergence of Measures*, Academic Press, 2014.

[3] BILLINGSLEY, P., *Convergence of Probability Measures*, Wiley, New York, 1999.

[4] COHN, D. L., *Measure Theory*, Birkhäuser, segunda edición, 2010.

[5] DUDLEY, R. M., *Real Analysis and Probability*, Cambridge University Press, segunda edición, 2011.

[6] DURRETT, R., *Probability. Theory and examples*, Cambridge University Press; cuarta edición, 2010.

[7] ETHIER, S., KURTZ, T., *Markov Processes: Characterization and Convergence*, Wiley-Interscience; 2nd edition, 2005.

[8] FÖLLMER, H., SCHIED, A., *Stochastic Finance: An Introduction in Discrete Time*, de Gruyter; Edición: 4th Rev., ed. 2016.

[9] GNEDENKO, B.V., KOLMOGOROV, A.N., *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, 1968.

[10] IGLEHART, D. L. *Weak convergence in applied probability*, Stoch. Proc. Appl. 2, pp. 211 – 241, 1974.

[11] KALLENBERG, O., *Foundations of Modern Probability*, Springer, Probability and its Applications, 2001.

[12] KALLENBERG, O., *Random Measures, Theory and Applications*, Springer, Probability Theory and Stochastic Modelling, 2017.

[13] KUSHNER, H., *Approximation and Weak Convergence Methods with Applications to Stochastic Systems Theory*, MIT Press, 2008.

[14] KUSHNER, H., *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, serie: Stochastic Modelling and Applied Probability (Book 24), 2000.

[15] ROYDEN, H. L., *Real Analysis*, Prentice Hall College, tercera edición 1988.

[16] TUCKER, H. G., *A graduate Course in Probability*, Dover Publications, Inc., 2014.