Teoría de la medida
Notas del curso

Carlos Chávez
Imanol Nuñez

1 de septiembre de 2020

Resumen

Algunas notas de teoría de la medida. En construcción…

Índice general

Capítulo 1

σ-álgebras y medidas

Nuestro primer objetivo será desarrollar una teoría de integración mucho más general y a la vez flexible que la teoría clásica. Pensemos en la integral de Riemann de una función integrable, f, sobre el intervalo $[a,b]$. La integral no es otra cosa que el límite de sumas de la forma,

$$\sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}),$$

donde $a = x_0 \leq \cdots \leq x_n = b$ conforman una partición del intervalo $[a,b]$. Esta definición de integral depende de nuestra capacidad para medir la longitud de cualquier intervalo: $l\big([c,d]\big) = d - c$. Si tuviéramos una manera precisa de asignar una longitud a cualquier subconjunto de $\mathbb{R}$ ¿por qué no definir la integral en términos de particiones arbitrarias del intervalo $[a,b]$? Incluso podríamos extrapolar esta idea a cualquier espacio arbitrario. De esta manera, si denotamos por $\mathcal{P}(X)$ al conjunto potencia de X (todos los subconjuntos de X), nos gustaría contar con una *medida* $\mu : \mathcal{P}(X) \to [0,\infty]$ y a través de esta definir la integral de cualquier función $f : X \to \mathbb{R}$. Para que la teoría sea consistente con todas las propiedades que deseamos que cumpla una integral será necesario pedirle ciertas condiciones a μ. Sin embargo, como veremos a lo largo del capítulo, para que de manera general μ se comporte de manera “adecuada", será necesario restringirse a subclases de $\mathcal{P}(X)$ que llamaremos *σ-álgebras*.

1.1. σ-álgebras y clases monótonas

Definición 1.1. Decimos que $\Sigma \subset \mathcal{P}(X)$ es un **álgebra** si:

1) $X \in \Sigma$,

2) para todo $A \in \Sigma$, también $A^c \in \Sigma$,

3) si $A, B \in \Sigma$, entonces $A \cup B \in \Sigma$,

Si en vez de 3) para $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ se tiene $\bigcup_{n\in\mathbb{N}} A_n \in \Sigma$, entonces decimos que Σ es una **$\boldsymbol{\sigma}$-álgebra**.

Nótese que $\emptyset = X^c \in \Sigma$. Además si Σ es una σ-álgebra es trivial que es un álgebra ya que para $A, B \in \Sigma$ podemos tomar $A_1 = A$, $A_2 = B$ y $A_n = \emptyset$ para $n > 2$. Así, $A \cup B = \bigcup_{n\in\mathbb{N}} A_n \in \Sigma$.

Si Σ es σ-álgebra, la dupla (X, Σ) será llamada un *espacio medible* y los elementos de Σ los *conjuntos medibles*.

Proposición 1.2. *Sea $\Sigma \subset \mathcal{P}(X)$ un álgebra y $A_1, \ldots, A_n \in \Sigma$. Entonces,*

1) $\bigcup_{i=1}^{n} A_i \in \Sigma$,

2) $\bigcap_{i=1}^{n} A_i \in \Sigma$, *3)* $A_1 - A_2 \in \Sigma$.

Si Σ *es* σ*-álgebra y* $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ *entonces* $\bigcap_{n\in\mathbb{N}} A_n \in \Sigma$.

Demostración. 1) Se sigue fácilmente de un argumento inductivo. 2) Tenemos la identidad $\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c\right)^c \in \Sigma$. 3) Por lo anterior, $A_1 - A_2 = A_1 \cap A_2^c \in \Sigma$. Finalmente, si Σ es σ-álgebra y $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$, de manera análoga a 2) tenemos $\bigcap_{n\in\mathbb{N}} A_n = \left(\bigcup_{n\in\mathbb{N}} A_n^c\right)^c \in \Sigma$. □

Usualmente las medidas están definidas explícitamente en clases de conjuntos que no son σ-álgebras. Más adelante nos encontraremos con el problema de extender una medida a partir de una clase de conjuntos $\mathcal{C}$ en la que si tenemos una noción de medida. Así que tiene sentido hablar de una *σ-álgebra generada por una clase* $\mathcal{C} \subset \mathcal{P}(X)$. Definimos a esta σ-álgebra como la más pequeña que contiene a $\mathcal{C}$ y su existencia está dada por la siguiente proposición:

Proposición 1.3. *Dada una clase* $\mathcal{C} \subset \mathcal{P}(X)$*, existe una mínima* σ*-álgebra en* X *que contiene a* $\mathcal{C}$ *y la denotaremos por* $\sigma(\mathcal{C})$*. Es decir, si* $\Sigma \subset \mathcal{P}(X)$ *es una* σ*-álgebra que contiene a* $\mathcal{C}$*, entonces* $\sigma(\mathcal{C}) \subset \Sigma$.

Demostración. Denotemos por $\mathcal{G}$ a la familia de σ-álgebras en X que contienen a $\mathcal{C}$. Esta familia no es vacía ya que $\mathcal{P}(X) \in \mathcal{G}$. Sea $\sigma(\mathcal{C})$ la intersección de todas las σ-álgebras en $\mathcal{G}$. Así, para toda $\Sigma \in \mathcal{G}$ tenemos $\sigma(\mathcal{C}) \subset \Sigma$ y claramente $\mathcal{C} \subset \sigma(\mathcal{C})$. Sólo queda probar que $\sigma(\mathcal{C})$ es una σ-álgebra.

Para todo $\Sigma \in \mathcal{G}$, $X \in \Sigma$. Por lo tanto $X \in \sigma(\mathcal{C})$. Supongamos ahora que $A \in \sigma(\mathcal{C})$. Entonces para todo $\Sigma \in \mathcal{G}$, $A \in \Sigma$. Esto a su vez implica que $A^c \in \Sigma$ y en consecuencia $A^c \in \sigma(\mathcal{C})$. De manera análoga podemos ver que si $\{A_n\}_{n\in\mathbb{N}} \subset \sigma(\mathcal{C})$, entonces $\bigcup_{n\geq 1} A_n \in \sigma(\mathcal{C})$. Por lo tanto $\sigma(\mathcal{C})$ es σ-álgebra. □

De ahora en andelante identificaremos a $\sigma(\mathcal{C})$ como la σ-álgebra generada por $\mathcal{C}$. Nótese que si $\mathcal{C}' \subset \mathcal{C}$, entonces $\sigma(\mathcal{C}') \subset \sigma(\mathcal{C})$.

Supongamos que (X, τ) es un espacio topológico. Definimos a la *σ-álgebra de Borel* en (X, τ) como la σ-álgebra generada por τ y la denotamos por $\mathcal{B}(X, \tau)$ ó simplemente $\mathcal{B}(X)$ si no se genera alguna confusión. Si (X, d) es un espacio métrico, $\mathcal{B}(X)$ denota a la σ-álgebra generada por los conjuntos que son abiertos con respecto d (la topología inducida por d).
De manera equivalente podríamos definir $\mathcal{B}(X) = \sigma\left(\{F \subset X : F \textit{ es cerrado}\}\right)$. Probemos esta igualdad:

Para todo cerrado F, F^c es abierto y en consecuencia $F \in \mathcal{B}(X)$. Se sigue la contención $\sigma\left(\{F \subset X : F \textit{ es cerrado}\}\right) \subset \mathcal{B}(X)$.
Análogamente, si G es abierto, G^c es cerrado y por lo tanto $G \in \sigma\big(\{F \subset X : F \textit{ es cerrado}\}\big)$. Podemos concluir que $\mathcal{B}(X) \subset \sigma\left(\{F \subset X : F \textit{ es cerrado}\}\right)$ y obtenemos la igualdad.

Los espacios de medida $\big(X, \mathcal{B}(X)\big)$ serán nuestro principal objeto de estudio. En particular trabajaremos con el espacio $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$ donde estamos considerando que $\mathbb{R}^n$ está equipado con la métrica euclidiana. A continuación daremos algunas equivalencias de $\mathcal{B}(\mathbb{R}^n)$. Para esto necesitaremos el siguiente lema:

Lema 1.4. *Todo abierto* $G \subset \mathbb{R}^n$ *es la unión numerable de rectángulos abiertos.*

Demostración. La demostración es similar a la de la última equivalencia que probamos. Para todo $x = (x_1, \ldots, x_n)$ en $\mathbb{R}^n$ y $r > 0$, denotamos por

$$C_r(x) := (x_1 - r, x_1 + r) \times \cdots \times (x_n - r, x_n + r).$$

Utilizaremos el hecho de que $\mathbb{Q}^n$ es un subconjunto denso numerable de $\mathbb{R}^n$. Sea $A = G \cap \mathbb{Q}^n$ y para todo $q \in A$ sea $R_q = \{r \in \mathbb{Q}^+ : C_r(q) \subset G\}$. Queremos probar que

$$G = \bigcup_{q \in A} \bigcup_{r \in R_q} C_r(q).$$

La contención hacia la izquierda es trivial. Ahora, dado $x \in G$, sea $r_0 > 0$ tal que $B_{r_0}(x) \subset G$. Luego tomamos $r' \in \mathbb{Q}^+$ tal que $0 < r' \leq \frac{r_0}{2\sqrt{n}}$ y $q_0 \in B_{r'}(x)$. Claramente $x \in B_{r'}(q_0) \subset C_{r'}(q_0)$. Solo queda probar que $C_{r'}(q_0) \subset G$. Pero notando que para todo $z \in C_{r'}(q_0)$ se tiene $d(q_0, z) < \sqrt{n}r'$, tenemos

$$d(x,z) \leq d(x,q_0) + d(q_0,z) < r' + \sqrt{n}r' \leq \frac{1+\sqrt{n}}{2\sqrt{n}} r_0 \leq r_0.$$

Es decir, $C_{r'}(q_0) \subset B_{r_0}(x) \subset G$ y podemos concluir que $x \in \bigcup_{q \in A} \bigcup_{r \in R_q} C_r(q)$.
Por lo tanto $G = \bigcup_{q \in A} \bigcup_{r \in R_q} C_r(q)$, que es lo que buscábamos demostrar. $\square$

Notación: Para $a = (a_1, \ldots, a_n)$ y $b = (b_1, \ldots, b_n)$ en $\mathbb{R}^n$ decimos que $a < b$ $(\leq)$ si y sólo si $a_i < b_i$ $(\leq)$ para $1 \leq i \leq n$. Si $a < b$, denotaremos al rectángulo abierto generado por a y b como

$$(a,b) := \{x \in \mathbb{R}^n : a < x < b\} = (a_1, b_1) \times \cdots \times (a_n, b_n).$$

De manera análoga definimos a los rectángulos $(a,b]$, $[a,b)$ y $[a,b]$. Por otro lado

$$(a, \infty) := (a_1, \infty) \times \cdots \times (a_n, \infty)$$

y definimos de manera análoga a $(-\infty, a)$, $(-\infty, a]$ y $[a, \infty)$.

Proposición 1.5. *Tenemos las siguientes equivalencias:*

1) $\mathcal{B}(\mathbb{R}^n) = \sigma\big(\{(a,b) : a,b \in \mathbb{R}^n, a < b\}\big)$,

2) $\mathcal{B}(\mathbb{R}^n) = \sigma\left(\{(a,b] : a,b \in \mathbb{R}^n, a < b\}\right)$
$= \sigma\left(\{[a,b) : a,b \in \mathbb{R}^n, a < b\}\right)$
$= \sigma\left(\{[a,b] : a,b \in \mathbb{R}^n, a < b\}\right)$,

3) $\mathcal{B}(\mathbb{R}^n) = \sigma\big(\{(a,\infty) : a \in \mathbb{R}^n\}\big) = \sigma\big(\{[a,\infty) : a \in \mathbb{R}^n\}\big)$,

4) $\mathcal{B}(\mathbb{R}^n) = \sigma\big(\{(-\infty,a) : a \in \mathbb{R}^n\}\big) = \sigma\big(\{(-\infty,a] : a \in \mathbb{R}^n\}\big)$.

Demostración. 1) $\{(a,b) : a,b \in \mathbb{R}^n, a < b\}$ es una subclase de conjuntos abiertos y por lo tanto $\sigma\left(\{(a,b) : a,b \in \mathbb{R}^n, a < b\}\right) \subset \mathcal{B}(\mathbb{R}^n)$. Por otro lado, por el lema 1.4 todo abierto $G \subset \mathbb{R}^n$ se puede ver como una unión numerable de rectángulos abiertos; es decir, $G = \bigcup_{k \in \mathbb{N}} \big(a^{(k)}, b^{(k)}\big)$. En consecuencia $G \in \sigma\left(\{(a,b) : a,b \in \mathbb{R}^n, a < b\}\right)$. Como $\mathcal{B}(\mathbb{R}^n)$ es la mínima σ-álgebra que contiene a los abiertos de $\mathbb{R}^n$, tenemos $\mathcal{B}(\mathbb{R}^n) \subset \sigma\left(\{(a,b) : a,b \in \mathbb{R}^n, a < b\}\right)$ y se sigue la igualdad.

2) Dado $c < d$ sea $d^{(k)} = (d_1 - 1/k, \cdots, d_n - 1/k)$. Entonces

$$(c,d) = \bigcup_{k \in \mathbb{N}} \left(c, d^{(k)}\right] \in \sigma\left(\{(a,b] : a,b \in \mathbb{R}^n, a < b\}\right).$$

Por otro lado, si $d^{(k)} = (d_1 + 1/k, \cdots, d_n + 1/k)$,

$$(c,d] = \bigcap_{k \in \mathbb{N}} \left(c, d^{(k)}\right) \in \sigma\left(\{(a,b) : a,b \in \mathbb{R}^n, a < b\}\right).$$

Por lo tanto $\sigma\left(\{(a,b] : a,b\in\mathbb{R}^n, a<b\}\right) = \sigma\left(\{(a,b) : a,b\in\mathbb{R}^n, a<b\}\right) = \mathcal{B}(\mathbb{R}^n)$. Las otras dos igualdades se demuestran de manera análoga.

3) Para todo $c\in\mathbb{R}^n$ sea $c^{(k)} = (c_1+k,\ldots,c_n+k)$. Tenemos

$$(c,\infty) = \bigcup_{k\in\mathbb{N}} \left(c, c^{(k)}\right] \in \sigma\left(\{(a,b] : a,b\in\mathbb{R}^n, a<b\}\right).$$

Ahora, sea $d^{(i)} = (c_1,\ldots,c_{i-1},d_i,c_{i+1},\ldots,c_n)$ para $1\le i\le n$. Entonces,

$$(c,d] = (c,\infty) - \bigcup_{i=1}^{n} \left(d^{(i)},\infty\right) \in \sigma\left(\{(a,\infty) : a\in\mathbb{R}^n\}\right).$$

Podemos concluir que $\sigma\big(\{(a,\infty) : a\in\mathbb{R}^n\}\big) = \sigma\left(\{(a,b] : a,b\in\mathbb{R}^n, a<b\}\right) = \mathcal{B}(\mathbb{R}^n)$. Las equivalencias restantes se prueban de manera muy similar. $\square$

Supongamos que $\{X_\alpha\}_{\alpha\in A}$ es una colección de conjuntos no vacíos. Definimos a $X = \prod_{\alpha\in A} X_\alpha$ como el conjunto de funciones $f : A\to\bigcup_{\alpha\in A} X_\alpha$ tales que $f(\alpha)\in X_\alpha$ para todo $\alpha\in A$. El **axioma de elección** postula que X no es vacío. Nótese que en el caso que A es finito, esta definición no es precisamente la misma que la del producto cartesiano. Para nosotros será conveniente considerarlos como lo mismo. De esta manera una n-ada $(x_1,\ldots,x_n)\in X_1\times\ldots\times X_n$ puede identificarse con la función $f : \{1,\ldots,n\}\to\bigcup_{i=1}^n X_i$ tal que $f(i) = x_i$ para $1\le i\le n$. Recíprocamente, para todo elemento $x\in X$ nos podemos referir a $x(\alpha)$ como la α-ésima coordenada de x y la podemos denotar por x_α. Conviniendo con lo anterior, no habrá ningún problema en adoptar la siguiente notación:

$$\left(\prod_{\beta\in B} X_\beta\right)\times\left(\prod_{\alpha\in A-B} X_\alpha\right) := \prod_{\alpha\in A} X_\alpha \quad \textit{para todo} \quad B\subset A.$$

Para todo $\alpha\in A$ definimos al α-ésimo **mapeo proyección**, $\pi_\alpha : X\to X_\alpha$, por medio de la identidad $\pi_\alpha(x) = x(\alpha)\equiv x_\alpha$.

Si Σ_α es σ-álgebra en X_α para todo $\alpha\in A$, definimos a la **$\boldsymbol{\sigma}$-álgebra producto** en X como

$$\bigotimes_{\alpha\in A}\Sigma_\alpha := \sigma\big(\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha\in\Sigma_\alpha,\ \alpha\in A\}\big).$$

Proposición 1.6. *Supongamos que $\Sigma_\alpha = \sigma(\mathcal{C}_\alpha)$ para cada $\alpha\in A$. Entonces,*

$$\bigotimes_{\alpha\in A}\Sigma_\alpha = \sigma\big(\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha\in\mathcal{C}_\alpha,\ \alpha\in A\}\big).$$

Si además A es finito o numerable y $X_\alpha\in\mathcal{C}_\alpha$ para todo $\alpha\in A$, entonces

$$\bigotimes_{\alpha\in A}\Sigma_\alpha = \sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha\in\mathcal{C}_\alpha\right\}\right).$$

Demostración. Denotemos por $\Sigma = \sigma\big(\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha\in\mathcal{C}_\alpha,\ \alpha\in A\}\big)$. Claramente $\Sigma\subset\bigotimes_{\alpha\in A}\Sigma_\alpha$. Por otro lado sea $\mathcal{M}_\alpha = \{E\in\mathcal{P}(X_\alpha) : \pi_\alpha^{-1}(E)\in\Sigma\}$. Es fácil ver que $\mathcal{M}_\alpha$ es una σ-álgebra en X_α y además contiene a $\mathcal{C}_\alpha$. Por lo tanto $\Sigma_\alpha\subset\mathcal{M}_\alpha$ que a su vez implica que $\pi_\alpha^{-1}(E)\in\Sigma$ para todo

$E \in \Sigma_\alpha$. Podemos concluir que $\bigotimes_{\alpha\in A} \Sigma_\alpha \subset \Sigma$ obteniendo así la igualdad.
Para la segunda equivalencia simplemente nótese que para todo $\beta \in A$ fijo y $E_\beta \in \mathcal{C}_\beta$,

$$\pi_\beta^{-1}(E_\beta) = E_\beta \times \prod_{\alpha\in A-\{\beta\}} X_\alpha \in \sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{C}_\alpha\right\}\right).$$

Por la primera parte del enunciado se sigue que $\bigotimes_{\alpha\in A} \Sigma_\alpha \subset \sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{C}_\alpha\right\}\right)$.
Por otro lado, para cualesquiera $E_\alpha \in \Sigma_\alpha$, dado que A es a lo más numerable,

$$\prod_{\alpha\in A} E_\alpha = \bigcap_{\alpha\in A} \pi_\alpha^{-1}(E_\alpha) \in \bigotimes_{\alpha\in A} \Sigma_\alpha.$$

Por lo tanto $\sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{C}_\alpha\right\}\right) \subset \bigotimes_{\alpha\in A} \Sigma_\alpha$ y podemos concluir. $\square$

Proposición 1.7. *Sean $X_1, \ldots, X_n$ espacios métricos separables y consideremos a $X = \prod_{i=1}^n X_i$ equipado con la métrica producto. Entonces $\mathcal{B}(X) = \bigotimes_{i=1}^n \mathcal{B}(X_i)$.*

Demostración. Por la proposición anterior,

$$\bigotimes_{i=1}^n \mathcal{B}(X_i) = \sigma\left(\left\{\prod_{i=1}^n U_i : U_i \subset X_i \textit{ es abierto}, 1 \leq i \leq n\right\}\right)$$

Dado que el producto de abiertos $U_i \subset X_i$ es abierto bajo la métrica producto, tenemos $\bigotimes_{i=1}^n \mathcal{B}(X_i) \subset \mathcal{B}(X)$. Ahora, como X_i es separable, existe una base numerable B_i. Luego,

$$B = \left\{\prod_{i=1}^n G_i : G_i \in B_i\right\},$$

es una base numerable en X. En consecuencia todo abierto $G \subset X$ es la unión numerable de abiertos (bajo la métrica producto) de la forma $\prod_{i=1}^n G_i \in B$ por lo que $G \in \bigotimes_{i=1}^n \mathcal{B}(X_i)$. Se sigue que $\mathcal{B}(X) \subset \bigotimes_{i=1}^n \mathcal{B}(X_i)$. $\square$

Corolario 1.8. $\mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}(\mathbb{R})$.

Existen otras subclases importantes de $\mathcal{P}(X)$ que podemos considerar, entre ellas estan son los π-sistemas y λ-sistemas:

Definición 1.9. Decimos que Π es un **π-sistema** si para todo $A, B \in \Pi$ tenemos $A \cap B \in \Pi$. Diremos que Λ es un **λ-sistema** si:

1) $X \in \Lambda$,

2) si $A, B \in \Lambda$ y $A \subset B$, entonces $B - A \in \Lambda$,

3) si $\{A_n\}_{n\in\mathbb{N}} \subset \Lambda$ y $A_n \subset A_{n+1}$ para todo $n \in \mathbb{N}$, entonces $\bigcup_{n\in\mathbb{N}} A_n \in \Lambda$.

Como ejemplos importantes de π-sistemas están los rectángulos semiabiertos en $\mathbb{R}^n$,

$$\Pi_1 = \{(a,b] : a, b \in \mathbb{R}^n, a \leq b\},$$

y los productos $\prod_{\alpha \in A} E_\alpha \in \bigotimes_{\alpha \in A} \Sigma_\alpha$ cuando A es a lo más numerable,

$$\Pi_2 = \left\{ \prod_{\alpha \in A} E_\alpha : E_\alpha \in \Sigma_\alpha \right\}.$$

Lema 1.10. *Una clase $\Sigma \subset \mathcal{P}(X)$ es σ-álgebra si y sólo si también es un λ-sistema y un π-sistema.*

Demostración. La implicación necesaria es trivial. Ahora, si Σ es un λ-sistema, por un lado $X \in \Sigma$ y para todo $A \in \Sigma$, $A^c = X - A \in \Sigma$. Luego, notamos que también por ser π sistema, para todo arreglo finito $A_1, \ldots, A_n \in \Sigma$, se tiene

$$\bigcup_{i=1}^n A_i = \left(\bigcap_{i=1}^n A_i^c \right)^c \in \Sigma.$$

Dada una sucesión $\{A_n\}_{n \in \mathbb{N}} \subset \Sigma$, sea $B_n = \bigcup_{i=1}^n A_i$. Entonces $B_n \in \Sigma$ y $B_n \subset B_{n+1}$. De nuevo, porque Σ es λ-sistema se sigue que $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n \in \Sigma$. $\square$

De manera análoga que en el caso de σ-álgebras, podemos definir al λ-*sistema* generado por una clase $\mathcal{C} \subset \mathcal{P}(X)$ como el mínimo λ-sistema que contiene a $\mathcal{C}$ y lo denotamos por $\lambda(\mathcal{C})$. El siguiente resultado es conocido como el **teorema de clases monótonas** y será de mucha utilidad:

Teorema 1.11. *Sea $\Pi \subset \mathcal{P}(X)$ un π-sistema. Entonces $\lambda(\Pi) = \sigma(\Pi)$.*

Demostración. Para cualquier $B \in \mathcal{P}(X)$ sea

$$\Sigma_B = \{A \in \mathcal{P}(X) : A \cap B \in \lambda(\Pi)\}.$$

Es fácil ver que si $B \in \lambda(\Pi)$ entonces Σ_B es λ-sistema. Si en particular $B \in \Pi$, entonces $\Pi \subset \Sigma_B$ y en consecuencia $\lambda(\Pi) \subset \Sigma_B$. Se sigue que para todo $A \in \lambda(\Pi)$ y $B \in \Pi$, $A \cap B \in \lambda(\Pi)$. En este caso, también podemos concluir que $\Pi \subset \Sigma_A$ que a su vez implica $\lambda(\Pi) \subset \Sigma_A$. Por lo tanto, de lo anterior tenemos que para todo $A, B \in \lambda(\Pi)$, $A \cap B \in \lambda(\Pi)$. Es decir, $\lambda(\Pi)$ también es un π-sistema y por el lema 1.10 tenemos que $\lambda(\Pi)$ es σ-álgebra. Así, obtenemos la contención $\sigma(\Pi) \subset \lambda(\Pi)$. Por otro lado la contención $\lambda(\Pi) \subset \sigma(\Pi)$ es trivial. $\square$

1.2. Medidas

Definición 1.12. Sea (X, Σ) un espacio medible. $\mu : \Sigma \to [0, \infty]$ es una **medida** si:

1) $\mu(\emptyset) = 0$,

2) si $\{A_n\}_{n \in \mathbb{N}} \subset \Sigma$ y $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces $\mu\big(\bigcup_{n \in \mathbb{N}} A_n\big) = \sum_{n=1}^{\infty} \mu(A_n)$.

La propiedad 2) es llamada σ**-aditividad** e implica **aditividad finita**:

Si $A_1, \ldots, A_n$ son disjuntos por pares entonces $\mu\big(\bigcup_{i=1}^n A_i\big) = \sum_{i=1}^n \mu(A_i)$.

La tripleta (X, Σ, μ) es llamada un **espacio de medida**. Un ejemplo de espacio de medida es $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ con $\mu(A) = |A|$. En este caso μ es llamada la *medida de conteo*.

Si (X, Σ) es un espacio medible, para cualquier $x \in X$ definimos a la *medida de Dirac centrada en* x a través de la identidad

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

Las siguientes son propiedades que cumplen las medidas:

Proposición 1.13. *Sea (X, Σ, μ) un espacio de medida. Entonces μ cumple:*

1) *(monotonía) Si $A, B \in \Sigma$ y $A \subset B$, entonces $\mu(A) \leq \mu(B)$,*

2) *(subaditividad) si $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$, entonces $\mu\big(\bigcup_{n\in\mathbb{N}} A_n\big) \leq \sum_{n=1}^\infty \mu(A_n)$,*

3) *(Continuidad por abajo) si $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ y $A_n \subset A_{n+1}$ para todo $n \in \mathbb{N}$, entonces $\mu\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \lim_{n\to\infty} \mu(A_n)$,*

4) *(Continuidad por arriba) si $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$, $A_{n+1} \subset A_n$ para todo $n \in \mathbb{N}$ y $\mu(A_1) < \infty$, entonces $\mu\big(\bigcap_{n\in\mathbb{N}} A_n\big) = \lim_{n\to\infty} \mu(A_n)$.*

Demostración. 1) $\mu(B) = \mu(B - A) + \mu(A) \geq \mu(A)$.

2) Sea $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$. Claramente $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$ y $B_i \cap B_j = \emptyset$ para $i \neq j$. Luego, por monotonía $\mu(B_n) \leq \mu(A_n)$ y por lo tanto,

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n=1}^\infty \mu(B_n) \leq \sum_{n=1}^\infty \mu(A_n).$$

3) Con $A_0 = \emptyset$, se tiene

$$\begin{aligned} \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} (A_n - A_{n-1})\right) &= \sum_{n=1}^\infty \mu(A_n - A_{n-1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i - A_{i-1}) = \lim_{n\to\infty} \mu(A_n). \end{aligned}$$

4) Sea $B_n = A_1 - A_n$, entonces $B_n \subset B_{n+1}$ y

$$\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} (A_1 \cap A_n^c) = A_1 \cap \left(\bigcup_{n\in\mathbb{N}} A_n^c\right) = A_1 - \bigcap_{n\in\mathbb{N}} A_n.$$

Dado que $\mu(A_1) < \infty$, $\mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mu(A_1) - \mu\left(\bigcap_{n\in\mathbb{N}} A_n\right)$. Luego, por la continuidad por abajo,

$$\begin{aligned} \mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) &= \mu(A_1) - \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) \\ &= \mu(A_1) - \lim_{n\to\infty} \mu(B_n) = \mu(A_1) - \lim_{n\to\infty} (\mu(A_1) - \mu(A_n)) = \lim_{n\to\infty} \mu(A_n). \end{aligned}$$ $\square$

Las siguientes son nociones importantes que tendremos que considerar en el futuro:

Definición 1.14. Decimos que μ es *finita* si $\mu(X) < \infty$. Es *σ-finita* si existen $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ tales que $\mu(E_n) < \infty$ para todo $n \in \mathbb{N}$ y $X = \bigcup_{n\in\mathbb{N}} E_n$.

Dado un espacio de medida (X, Σ, μ) diremos que un conjunto $A \in \Sigma$ es *nulo* (o μ-nulo) si $\mu(A) = 0$. Por subaditividad se sigue que toda unión numerable de conjuntos nulos es un conjunto nulo. Además, si A es un conjunto nulo, por monotonía se sigue que para todo $B \in \Sigma$ tal que $B \subset A$, $\mu(B) = 0$. Cuando Σ contiene a todos los subconjuntos de conjuntos nulos decimos μ es *completa* y que (X, Σ, μ) es un espacio de medida *completo*. Una caracterización alternativa de completez, la cual aclará el término, es la siguiente: si $A_1 \subset B \subset A_2$ con $A_1, A_2 \in \Sigma$ tales que $\mu(A_2 - A_1) = 0$ entonces $B \in \Sigma$. Para ver que las definiciones son equivalentes, notemos que si $A_1 = \emptyset$ entonces recuperamos la primera definición. Conversamente es fácil ver que $B - A_1 \subset A_2 - A_1$, por lo que $\mu(A_2 - A_1) = 0$ implica que $B - A_1 \in \Sigma$ y por lo tanto $B = A_1 \cup (B - A_1) \in \Sigma$.

Un concepto intímamente relacionado con los conjuntos nulos es el siguiente: en un espacio (X, Σ, μ), diremos que una proposición $P(x)$ referente a $x \in X$ es cierta **μ-casi donde sea**, **μ-casi dondequiera** o **μ-casi seguramente** (abreviados por μ-c.d.s, μ-c.d. y μ-c.s. respectivamente o c.d.s, c.d., c.s. cuando no haya confusión) si existe $A \in \Sigma$ con $\mu(A) = 0$ tal que $P(x)$ es cierta si $x \in X - A$. Una aclaración pertinente es que el conjunto $\{x \in X : P(x) \text{ es falso}\}$ no necesariamente pertenece a Σ. Sin embargo si (X, Σ, μ) es completo entonces $\{x \in X : P(x) \text{ es falso}\} \in \Sigma$ cuando $P(x)$ es cierta casi dondequiera.

Por cuestiones técnicas resulta conveniente trabajar con medidas completas y en caso de que una medida no lo sea siempre se puede extender a una medida que sí sea completa:

Teorema 1.15. *Sea $\mathcal{N} = \{A \in \Sigma : \mu(A) = 0\}$ y $\overline{\Sigma} = \{E \cup B : E \in \Sigma,\ B \subset A \in \mathcal{N}\}$. Entonces $\overline{\Sigma}$ es σ-álgebra y existe una única extensión $\overline{\mu}$ de μ que es completa en $\overline{\Sigma}$.*

Demostración. Trivialmente $X \in \overline{\Sigma}$. Por otro lado, como Σ y $\mathcal{N}$ son cerrados bajo uniones numerables, entonces también lo es $\overline{\Sigma}$. Ahora, sean $E \in \Sigma$ y $B \subset A \in \mathcal{N}$. Podemos asumir que $E \cap A = \emptyset$ ya que podemos remplazar a A y B por $A - E$ y $B - E$ respectivamente. Entonces $E \cup B = (E \cup A) \cap (A^c \cup B)$ y tenemos $(E \cup B)^c = (E^c \cap A^c) \cup (A \cap B^c) \in \overline{\Sigma}$. Por lo tanto $\overline{\Sigma}$ es σ-álgebra.
Dado $E \cup B \in \overline{\Sigma}$ definimos $\overline{\mu}(E \cup B) = \mu(E)$. Veamos que $\overline{\mu}$ está bien definida. Si $E \cup B = \tilde{E} \cup \tilde{B}$ con $\tilde{B} \subset \tilde{A}$ para algún $\tilde{A} \in \mathcal{N}$, entonces $E \subset \tilde{E} \cup \tilde{A}$. Por lo tanto

$$\overline{\mu}(E \cup B) = \mu(E) \leq \mu(\tilde{E}) + \mu(\tilde{A}) = \mu(\tilde{E}) = \overline{\mu}(\tilde{E} \cup \tilde{B}).$$

Siguiendo el mismo razonamiento obtenemos $\mu(\tilde{E} \cup \tilde{B}) \leq \overline{\mu}(E \cup B)$ obteniendo así la igualdad.
Supongamos que $E \in \Sigma$, $B \subset A \in \mathcal{N}$ y $\overline{\mu}(E \cup B) = 0$. Entonces $E \in \mathcal{N}$ y en consecuencia todo subconjunto $F \subset E \cup B$ esta contenido en $E \cup A \in \mathcal{N}$. Por lo tanto $F \in \overline{\Sigma}$ y concluimos que $\overline{\mu}$ es completa.

Si $\{E_n \cup B_n\}_{n\in\mathbb{N}} \subset \overline{\Sigma}$ son disjuntos por pares, entonces

$$\begin{aligned}\overline{\mu}\left(\bigcup_{n\in\mathbb{N}}(E_n \cup B_n)\right) &= \overline{\mu}\left(\left(\bigcup_{n\in\mathbb{N}} E_n\right) \cup \left(\bigcup_{n\in\mathbb{N}} B_n\right)\right) \\ &= \mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) \\ &= \sum_{n=1}^{\infty} \mu(E_n) \\ &= \sum_{n=1}^{\infty} \overline{\mu}(E_n \cup B_n).\end{aligned}$$

Finalmente, supongamos que $\tilde{\mu}$ es una extensión completa de μ en $\overline{\Sigma}$. Para $E \in \Sigma$ y $B \subset A \in \mathcal{N}$, claramente $E, B \in \overline{\Sigma}$ y en consecuencia $\tilde{\mu}(E \cup B) \leq \tilde{\mu}(E) + \tilde{\mu}(B) \leq \tilde{\mu}(E) + \tilde{\mu}(A) = \tilde{\mu}(E)$. Por monotonía se sigue que $\tilde{\mu}(E \cup B) = \tilde{\mu}(E) = \mu(E)$. □

1.3. Medidas exteriores y teorema de extensión de Carathéodory

Supongamos que en un espacio X existe una clase $\mathcal{C} \subset \mathcal{P}(X)$ para la cual tenemos una noción de medida a través de una función $\mu_0 : \mathcal{C} \to [0, \infty]$. Por ahora solo suponemos que $\emptyset, X \in \mathcal{C}$ y $\mu_0(\emptyset) = 0$. Definiremos la **medida exterior** de cualquier $A \in \mathcal{P}(X)$ inducida por μ_0 como

$$\mu^*(A) := \text{ínf}\left\{\sum_{n=1}^{\infty} \mu_0(E_n) : E_n \in \mathcal{C},\ A \subset \bigcup_{n\in\mathbb{N}} E_n\right\}.$$

La idea es simple; aproximaremos la medida de cualquier conjunto $A \in \mathcal{P}(X)$ a través de conjuntos que si podemos medir. Nótese que μ^* está bien definida ya que $X \in \mathcal{C}$.

Proposición 1.16. *La medida exterior $\mu^* : \mathcal{P}(X) \to [0, \infty]$ cumple las propiedades de monotonía, subaditividad y $\mu^*(\emptyset) = 0$.*

Demostración. Como $\emptyset \in \mathcal{C}$ y $\mu_0(\emptyset) = 0$, es inmediato que $\mu^*(\emptyset) = 0$. Por otro lado, si $A \subset B$ entonces

$$\left\{\sum_{n=1}^{\infty} \mu_0(E_n) : E_n \in \mathcal{C},\ B \subset \bigcup_{n\in\mathbb{N}} E_n\right\} \subset \left\{\sum_{n=1}^{\infty} \mu_0(E_n) : E_n \in \mathcal{C},\ A \subset \bigcup_{n\in\mathbb{N}} E_n\right\}.$$

En consecuencia $\mu^*(A) \leq \mu^*(B)$. Para probar subaditividad tómense $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ y $\epsilon > 0$. Para cada $n \in \mathbb{N}$ existe $\{E_k^n\}_{k\in\mathbb{N}} \subset \mathcal{C}$ tal que $A_n \subset \bigcup_{k\in\mathbb{N}} E_k^n$ y $\sum_{k=1}^{\infty} \mu_0(E_k^n) \leq \mu^*(A_n) + \epsilon/2^n$. Luego, notamos que $\bigcup_{n\in\mathbb{N}} A_n \subset \bigcup_{n,k\in\mathbb{N}} E_k^n$ y en consecuencia

$$\mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n,k\in\mathbb{N}} \mu_0(E_k^n) = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mu_0(E_k^n) \leq \sum_{n=1}^{\infty}(\mu^*(A_n) + \epsilon/2^n) = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon.$$

Como ϵ es arbitrario se sigue la subaditividad. □

Diremos que $A \in \mathcal{P}(X)$ es μ^**-medible* si y sólo si para todo $E \in \mathcal{P}(X)$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Denotemos a la clase de conjuntos μ^*-medibles por $\mathcal{M}$.

Proposición 1.17. *$\mathcal{M}$ es una σ-álgebra y la restricción de μ^* a $\mathcal{M}$ es una medida completa.*

Demostración. Claramente $\mathcal{M}$ es cerrado bajo complemento. Ahora probemos que $\mathcal{M}$ es un álgebra. Sean $A, B \in \mathcal{M}$ y $E \subset X$. Primero tenemos

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c)\\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c).\end{aligned}$$

Por otro lado, $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$ así que por subaditividad,

$$\mu^*(E \cap (A \cup B)) \leq \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B).$$

Agregando $\mu^*(E \cap (A \cup B)^c)$ en ambos lados de la última desigualdad se sigue que

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).$$

Por subaditividad tenemos la otra desigualdad. Por lo tanto $A \cup B \in \mathcal{M}$. Además, si $A \cap B = \emptyset$,

$$\mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B);$$

es decir, μ^* es aditiva en $\mathcal{M}$.
Veamos que $\mathcal{M}$ es σ-álgebra. Dada una sucesión $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{M}$ queremos probar que $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$. Podemos suponer que los conjuntos A_n son disjuntos por pares (de no ser así consideremos $B_n = A_n - \bigcup_{i=1}^{n-1} A_i \in \mathcal{M}$). Entonces para todo $E \subset X$,

$$\begin{aligned}\mu^*\left(E \cap \bigcup_{i=1}^{n} A_i\right) &= \mu^*\left(\left(E \cap \bigcup_{i=1}^{n} A_i\right) \cap A_n\right) + \mu^*\left(\left(E \cap \bigcup_{i=1}^{n} A_i\right) \cap A_n^c\right)\\ &= \mu^*\left(E \cap A_n\right) + \mu^*\left(E \cap \bigcup_{i=1}^{n-1} A_i\right).\end{aligned}$$

Inductivamente se sigue que $\mu^*\left(E \cap \bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu^*(E \cap A_i)$. Luego, como μ^* es monótona,

$$\mu^*(E) = \mu^*\left(E \cap \bigcup_{i=1}^{n} A_i\right) + \mu^*\left(E \cap \left(\bigcup_{i=1}^{n} A_i\right)^c\right) \geq \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*\left(E \cap \left(\bigcup_{n\in\mathbb{N}} A_n\right)^c\right).$$

Dado que la desigualdad anterior se cumple para todo $n \in \mathbb{N}$, de la subaditivdad de μ^* obtenemos

$$\begin{aligned}\mu^*(E) &\geq \sum_{n=1}^{\infty} \mu^*(E \cap A_n) + \mu^*\left(E \cap \left(\bigcup_{n\in\mathbb{N}} A_n\right)^c\right)\\ &\geq \mu^*\left(E \cap \bigcup_{n\in\mathbb{N}} A_n\right) + \mu^*\left(E \cap \left(\bigcup_{n\in\mathbb{N}} A_n\right)^c\right) \geq \mu^*(E).\end{aligned}$$

Por lo tanto se tiene la igualdad y vemos que $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$. También de la última desigualdad se puede ver que μ^* es σ-aditiva (tomando $E = \bigcup_{n\in\mathbb{N}} A_n$). Para finalizar, si $\mu^*(A) = 0$, para todo $E \subset X$,

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E)$$

y podemos concluir que $A \in \mathcal{M}$. Por lo tanto μ^* es completa en $\mathcal{M}$. □

Definición 1.18. Sea $\mathcal{A} \subset \mathcal{P}(X)$ un álgebra. $\mu_0 : \mathcal{A} \to [0, \infty]$ es una **premedida** si,

1) $\mu_0(\emptyset) = 0$,

2) para toda sucesión, $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, de conjuntos disjuntos por pares tales que $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$, se tiene $\mu_0\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \sum_{n=1}^{\infty} \mu_0(A_n)$. Es decir, μ_0 es σ-aditiva.

Lema 1.19. *Sea μ_0 una premedida en un álgebra $\mathcal{A} \subset \mathcal{P}(X)$ y $\mu^* : \mathcal{P}(X) \to [0, \infty]$ la medida exterior inducida por μ_0; es decir,*

$$\mu^*(E) = \text{ínf} \left\{ \sum_{n=1}^{\infty} \mu_0(A_n) : A_n \in \mathcal{A},\ E \subset \bigcup_{n\in\mathbb{N}} A_n \right\}.$$

Entonces, $\mu^* = \mu_0$ en $\mathcal{A}$ y todo $A \in \mathcal{A}$ es μ^*-medible.

Demostración. Sea $A \in \mathcal{A}$ y $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ tal que $A \subset \bigcup_{n\in\mathbb{N}} A_n$. Si tomamos $B_n = A \cap \big(A_n - \bigcup_{i=1}^{n-1} A_i\big) \in \mathcal{A}$, entonces los conjuntos B_n son disjuntos por pares y $A = \bigcup_{n\in\mathbb{N}} B_n$. Por otro lado, es fácil ver que una premedida cumple la propiedad de monotonía y en consecuencia,

$$\mu_0(A) = \mu_0\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n=1}^{\infty} \mu_0(B_n) \leq \sum_{n=1}^{\infty} \mu_0(A_n).$$

Se sigue que $\mu_0(A) \leq \mu^*(A)$. La otra desigualdad la podemos obtener fácilmente tomando $A_1 = A$ y $A_n = \emptyset$ para todo $n > 1$. Por lo tanto $\mu^*(A) = \mu_0(A)$ para todo $A \in \mathcal{A}$.
Ahora, dados $A \in \mathcal{A}$, $E \subset X$ y $\epsilon > 0$ sea $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ tal que $E \subset \bigcup_{n\in\mathbb{N}} A_n$ y $\sum_{n=1}^{\infty} \mu_0(A_n) \leq \mu^*(E) + \epsilon$. Entonces,

$$\begin{aligned}\mu^*(E) + \epsilon \geq \sum_{n=1}^{\infty} \mu_0(A_n) &= \sum_{n=1}^{\infty} \mu_0(A_n \cap A) + \sum_{n=1}^{\infty} \mu_0(A_n \cap A^c) \\ &\geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \geq \mu^*(E).\end{aligned}$$

Como ϵ es arbitrario podemos concluir que $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$; es decir, A es μ^*-medible. □

Teorema 1.20. *(Teorema de extensión de Carathéodory) Sea μ_0 una premedida en un álgebra $\mathcal{A} \subset \mathcal{P}(X)$ y μ^* la medida exterior inducida por μ_0. Entonces $\mu = \mu^*|\sigma(\mathcal{A})$ es una medida en $\sigma(\mathcal{A})$ que extiende a μ_0. Más aún, si ν es otra medida en $\sigma(\mathcal{A})$ que extiende a μ_0, entonces $\nu(E) \leq \mu(E)$ para todo $E \in \sigma(\mathcal{A})$ y la igualdad se cumple si $\mu(E) < \infty$. Si μ_0 es σ-finita, entonces μ es la única extensión de μ_0 en $\sigma(\mathcal{A})$.*

Demostración. Por la proposición 1.17, μ^* define una medida en la σ-álgebra de conjuntos μ^*-medibles, $\mathcal{M}$. Por el lema anterior $\mathcal{A} \subset \mathcal{M}$ y $\mu^*(A) = \mu_0(A)$ para todo $A \in \mathcal{A}$. En consecuencia $\sigma(\mathcal{A}) \subset \mathcal{M}$ y $\mu = \mu^*|\sigma(\mathcal{A})$ es una medida que extiende a μ_0.
Probemos la segunda parte del enunciado. Sea ν una medida en $\sigma(\mathcal{A})$ que extiende a μ_0. Si $E \in \sigma(\mathcal{A})$ y $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ es tal que $E \subset \bigcup_{n\in\mathbb{N}} A_n$, entonces $\nu(E) \leq \sum_{n=1}^{\infty} \nu(A_n) = \sum_{n=1}^{\infty} \mu_0(A_n)$. Por lo tanto $\nu(E) \leq \mu^*(E) \equiv \mu(E)$.
Ahora supongamos que $\mu(E) < \infty$. Dado $\epsilon > 0$, sea $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ tal que $E \subset \bigcup_{n\in\mathbb{N}} A_n$ y $\sum_{n=1}^{\infty} \mu_0(A_n) \leq \mu(E) + \epsilon$. Denotemos por $A = \bigcup_{n\in\mathbb{N}} A_n$, entonces

$$\mu\left(A - E\right) + \mu(E) = \mu(A) \leq \sum_{n=1}^{\infty} \mu_0(A_n) \leq \mu(E) + \epsilon.$$

Como $\mu(E) < \infty$, obtenemos $\mu(A - E) \leq \epsilon$. Por otro lado,

$$\nu(A) = \lim_{n\to\infty} \nu\left(\bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \mu_0\left(\bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \mu\left(\bigcup_{i=1}^{n} A_i\right) = \mu(A).$$

En consecuencia,

$$\mu(E) \leq \mu(A) = \nu(A) = \nu(A - E) + \nu(E) \leq \mu(A - E) + \nu(E) \leq \epsilon + \nu(E).$$

Como ϵ es arbitrario se sigue $\nu(E) = \mu(E)$. Finalmente, sea $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ tal que $X = \bigcup_{n\in\mathbb{N}} A_n$ y $\mu_0(A_n) < \infty$ para todo $n \in \mathbb{N}$. Sin mayor perdida de generalidad podemos suponer que los conjuntos A_n son disjuntos por pares. Entonces para todo $E \in \sigma(\mathcal{A})$,

$$\nu(E) = \sum_{n=1}^{\infty} \nu(E \cap A_n) = \sum_{n=1}^{\infty} \mu(E \cap A_n) = \mu(E).$$

Por lo tanto $\nu = \mu$. □

1.4. Construcción de medidas

Las medidas exteriores y el teorema de extensión de Carathéodory son clave para poder definir medidas no triviales. Sin embargo hay un paso previo que debemos considerar. Por ejemplo, pensemos en la longitud de un subconjunto de $\mathbb{R}$. Tenemos una noción precisa de longitud de un intervalo $(a, b]$; es decir $l\big((a, b]\big) = b - a$. Naturalmente podríamos tomar la medida exterior con respecto l definido en

$$\mathcal{E} = \{(a, b] : -\infty \leq a \leq b < \infty\} \cup \{(a, \infty) : -\infty \leq a < \infty\}.$$

El problema es que la clase $\mathcal{E}$ no es un álgebra y no podemos usar directamente el Teorema de extensión de Carathéodory. Resulta que $\mathcal{E}$ es una clase elemental:

Definición 1.21. Dado un espacio X, decimos que $\mathcal{E} \subset \mathcal{P}(X)$ es una **clase elemental** si:

1) $\emptyset \in \mathcal{E}$,

2) si $A, B \in \mathcal{E}$, entonces $A \cap B \in \mathcal{E}$,

3) si $A \in \mathcal{E}$, entonces existen $B_1, \ldots, B_n \in \mathcal{E}$ disjuntos por pares y tales que $A^c = \bigcup_{i=1}^n B_i$.

Como veremos a continuación, extender medidas a partir de una clase elemental no requerirá mucho más trabajo del que ya hemos hecho.

Proposición 1.22. *Sea $\mathcal{E}$ una clase elemental. Entonces*

$$\mathcal{A} := \left\{ \bigcup_{i=1}^n A_i : \{A_i\}_{i=1}^n \subset \mathcal{E},\ A_i \cap A_j = \emptyset \textit{ para } i \neq j \right\}$$

es un álgebra. Más aún, supongamos que $\rho : \mathcal{E} \to [0, \infty]$ es σ-aditiva y $\rho(\emptyset) = 0$. Si definimos $\mu_0 : \mathcal{A} \to [0, \infty]$ para todo $A = \bigcup_{i=1}^n A_i \in \mathcal{A}$ por medio de

$$\mu_0\left(\bigcup_{i=1}^n A_i\right) := \sum_{i=1}^n \rho(A_i),$$

entonces μ_0 está bien definida y es una premedida.

Demostración. Veamos que $\mathcal{A}$ es cerrado bajo complemento. Sean $A_1, \ldots, A_n \in \mathcal{E}$ disjuntos por pares. Como $\mathcal{E}$ es clase elemental, para $1 \leq i \leq n$ existen $B_1^i, \ldots, B_{J_i}^i \in \mathcal{E}$ disjuntos y tales que $A_i^c = \bigcup_{j=1}^{J_i} B_j^i$. Entonces,

$$\left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n \bigcup_{j=1}^{J_i} B_j^i = \bigcup \left\{ B_{j_1}^1 \cap \cdots \cap B_{j_n}^n : 1 \leq j_i \leq J_i,\ 1 \leq i \leq n \right\} \in \mathcal{A}.$$

Es fácil ver que $\mathcal{A}$ es cerrado bajo intersección. Si $A_1, \ldots, A_n \in \mathcal{E}$ y $B_1, \ldots, B_m \in \mathcal{E}$ son arreglos de conjuntos respectivamente disjuntos, entonces

$$\left(\bigcup_{i=1}^n A_i\right) \bigcap \left(\bigcup_{j=1}^m B_j\right) = \bigcup \left\{ A_i \cap B_j : 1 \leq i \leq n,\ 1 \leq j \leq m \right\} \in \mathcal{A}.$$

En consecuencia si $A, B \in \mathcal{A}$ podemos concluir que $A \cup B = \left(A^c \cap B^c\right)^c \in \mathcal{A}$. Por lo tanto $\mathcal{A}$ es álgebra.
Probemos que μ_0 está bien definida. Sean $A_1, \ldots, A_n \in \mathcal{E}$ y $B_1, \ldots, B_m \in \mathcal{E}$ arreglos de conjuntos disjuntos tales que $\bigcup_{i=1}^n A_i = \bigcup_{j=1}^m B_j$. Dado que ρ es aditiva, $A_i = \bigcup_{j=1}^m (A_i \cap B_j)$ y $B_j = \bigcup_{i=1}^n (A_i \cap B_j)$ (los conjuntos $A_i \cap B_j$ son disjuntos por pares), entonces

$$\sum_{i=1}^n \rho(A_i) = \sum_{i=1}^n \sum_{j=1}^m \rho(A_i \cap B_j) = \sum_{j=1}^m \sum_{i=1}^n \rho(A_i \cap B_j) = \sum_{j=1}^m \rho(B_j).$$

Solo nos queda probar que μ_0 es σ-aditiva. Sea $A = \bigcup_{i=1}^n A_i \in \mathcal{A}$ para conjuntos $A_1, \ldots, A_n \in \mathcal{E}$ disjuntos. Supongamos que $A = \bigcup_{k \in \mathbb{N}} B_k$, donde $\{B_k\}_{k \in \mathbb{N}} \subset \mathcal{A}$ es una sucesión de conjuntos disjuntos. Para cada $k \in \mathbb{N}$, $B_k = \bigcup_{j=1}^{J_k} C_j^k$ con $C_1^k, \ldots, C_{J_k}^k \in \mathcal{E}$ disjuntos. Ahora, para $1 \leq i \leq n$ tenemos

$$A_i = \bigcup_{k \in \mathbb{N}} \bigcup_{j=1}^{J_k} \left(C_j^k \cap A_i\right).$$

Mientras que para cada $k \in \mathbb{N}$,

$$B_k = \bigcup_{i=1}^{n} \bigcup_{j=1}^{J_k} (C_j^k \cap A_i).$$

Dado que los conjuntos $C_j^k \cap A_i \in \mathcal{E}$ son disjuntos por pares y ρ es σ-aditiva,

$$\begin{aligned} \mu_0(A) = \sum_{i=1}^{n} \rho(A_i) &= \sum_{i=1}^{n} \sum_{\substack{k\in\mathbb{N}, \\ 1\leq j\leq J_i}} \rho\big(C_j^k \cap A_i\big) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{\infty} \sum_{j=1}^{J_k} \rho\big(C_j^k \cap A_i\big) = \sum_{k=1}^{\infty} \sum_{i=1}^{n} \sum_{j=1}^{J_k} \rho\big(C_j^k \cap A_i\big) = \sum_{k=1}^{\infty} \mu_0(B_k). \end{aligned}$$

Por lo tanto μ_0 es premedida. $\square$

Teorema 1.23. *Sea $\mathcal{E} \subset \mathcal{P}(X)$ una clase elemental. Supongamos que $\rho : \mathcal{E} \to [0, \infty]$ es σ-aditiva, σ-finita y $\rho(\emptyset) = 0$. Si μ^* es la medida exterior inducida por ρ, entonces $\mu = \mu^*|\sigma(\mathcal{E})$ es la única medida en $\sigma(\mathcal{E})$ que extiende a ρ; es decir, $\mu(A) = \rho(A)$ para todo $A \in \mathcal{E}$.*

Demostración. Definamos μ_0 y $\mathcal{A}$ como en la proposición anterior. Para todo $E \subset X$, es fácil ver que

$$\left\{ \sum_{n=1}^{\infty} \mu_0(A_n) : A_n \in \mathcal{A},\ E \subset \bigcup_{n\in\mathbb{N}} A_n \right\} = \left\{ \sum_{n=1}^{\infty} \rho(B_n) : B_n \in \mathcal{E},\ E \subset \bigcup_{n\in\mathbb{N}} B_n \right\}.$$

Por lo tanto μ^* no solo es la medida exterior inducida por ρ sino también por la premedida μ_0. También resulta muy sencillo ver que $\sigma(\mathcal{A}) = \sigma(\mathcal{E})$ y que μ_0 es σ-finita. Por el teorema de extensión de Caratheodory, $\mu = \mu^*|\sigma(\mathcal{E})$ es la única medida que extiende a μ_0 en $\sigma(\mathcal{E})$. Por otro lado, si $\nu : \sigma(\mathcal{E}) \to [0, \infty]$ es una medida que extiende a ρ, en particular es una medida que extiende a μ_0 y se sigue que $\nu = \mu$. $\square$

Con el teorema 1.23 ya estamos en posición de definir medidas en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

Teorema 1.24. *Sea $F : \mathbb{R} \to \mathbb{R}$ una función no-decreciente y continua por la derecha. Entonces existe una única medida $\mu_F : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ tal que para todo $a \leq b$ en $\mathbb{R}$, $\mu_F\big((a, b]\big) = F(b) - F(a)$. Más aún, para todo $E \in \mathcal{B}(\mathbb{R})$,*

$$\mu_F(E) = \inf \left\{ \sum_{n=1}^{\infty} \big(F(b_n) - F(a_n)\big) : b_n, a_n \in \mathbb{R},\ E \subset \bigcup_{n\in\mathbb{N}} (a_n, b_n] \right\}.$$

Demostración. Consideremos a la clase elemental

$$\mathcal{E} = \big\{(a, b] : -\infty \leq a \leq b < \infty\big\} \cup \big\{(a, \infty) : -\infty \leq a < \infty\big\}.$$

Sean $M^+ = \lim_{x\to\infty} F(x)$ y $M^- = \lim_{x\to-\infty} F(x)$ y definimos $\rho : \mathcal{E} \to [0, \infty]$ por medio de las identidades

$$\begin{aligned} \rho\big((a, b]\big) &= F(b) - F(a) \quad \textit{si} \quad -\infty < a \leq b < \infty, \\ \rho\big((-\infty, b]\big) &= F(b) - M^- \quad \textit{si} \quad -\infty < b < \infty, \\ \rho\big((a, \infty)\big) &= M^+ - F(a) \quad \textit{si} \quad -\infty < a < \infty, \\ \rho\big((-\infty, \infty)\big) &= M^+ - M^-. \end{aligned}$$

Claramente ρ es σ-finita ya que $\mathbb{R} = \bigcup_{k\in\mathbb{Z}}(k, k+1]$ y $\rho(\emptyset) = 0$. Veamos que ρ es σ-aditiva. Supongamos que para $(a, b] \in \mathcal{E}$ existen intervalos disjuntos, $\{(a_k, b_k]\}_{k\in\mathbb{N}} \subset \mathcal{E}$, tales que $(a, b] = \bigcup_{k\in\mathbb{N}}(a_k, b_k]$. Para todo $k \in \mathbb{N}$ podemos reordenar el arreglo $(a_1, b_1], \dots, (a_k, b_k]$ para que $a \le a_1 \le b_1 \le \cdots \le a_k \le b_k \le b$. En este caso, resulta evidente que

$$\begin{aligned}\rho\big((a,b]\big) &= \rho\big((a,a_1]\big) + \rho\big((a_1,b_1]\big) + \rho\big((b_1,a_2]\big) + \cdots + \rho\big((a_k,b_k]\big) + \rho\big((b_k,b]\big)\\ &\ge \rho\big((a_1,b_1]\big) + \rho\big((a_2,b_2]\big) + \cdots + \rho\big((a_{k-1},b_{k-1}]\big) + \rho\big((a_k,b_k]\big)\\ &= \sum_{i=1}^{k} \rho\big((a_i,b_i]\big).\end{aligned}$$

Como la desigualdad anterior se cumple para todo $k \in \mathbb{N}$, podemos concluir que $\rho\big((a,b]\big) \ge \sum_{k=1}^{\infty} \rho\big((a_k,b_k]\big)$. Ahora, supongamos que $(a, b] \subset \bigcup_{i=1}^{n}(c_i, d_i]$. Veremos por inducción que

$$\rho\big((a,b]\big) \le \sum_{i=1}^{n} \rho\big((c_i,d_i]\big).$$

Para $n = 1$ está claro que se cumple. Supongamos que se cumple para $n-1$ y probemos el resultado para n. Debe existir algún intervalo $(c_j, d_j]$ tal que $c_j < b \le d_j$. Aplicando la hipótesis inductiva al intervalo $(a, c_j]$ se sigue que

$$\rho\big((a,b]\big) = \rho\big((a,c_j]\big) + \rho\big((c_j,b]\big) \le \sum_{i\ne j} \rho\big((c_i,d_i]\big) + \rho\big((c_j,d_j]\big).$$

Probemos que se cumple la desigualdad restante. Supongamos que $a > -\infty$. Dado $\epsilon > 0$, sea δ tal que $F(a+\delta) \le F(a) + \epsilon$ y para todo $k \in \mathbb{N}$, sea δ_k tal que $F(b_k + \delta_k) \le F(b_k) + \epsilon/2^k$. Dado que los intervalos $(a_k, b_k + \delta_k)$ son una cubierta abierta para el intervalo compacto $[a+\delta, b]$, existe un arreglo finito, $(a_{k_1}, b_{k_1} + \delta_{k_1}], \dots, (a_{k_n}, b_{k_n} + \delta_{k_n}]$, que lo cubre. Por el resultado previo tenemos entonces que

$$\begin{aligned}\rho\big((a,b]\big) = F(b) - F(a) &\le F(b) - F(a+\delta) + \epsilon\\ &\le \sum_{i=1}^{n} \big(F(b_{k_i} + \delta_{k_i}) - F(a_{k_i})\big) + \epsilon\\ &\le \sum_{k=1}^{\infty} \big(F(b_k + \delta_k) - F(a_k)\big) + \epsilon\\ &\le \sum_{k=1}^{\infty} \big(F(b_k) - F(a_k) + \epsilon/2^k\big) + \epsilon = \sum_{k=1}^{\infty} \rho\big((a_k,b_k]\big) + 2\epsilon.\end{aligned}$$

Como ϵ es arbitrario se sigue la desigualdad $\rho\big((a,b]\big) \le \sum_{k=1}^{\infty} \rho\big((a_k,b_k]\big)$. En el caso que $a = -\infty$, podemos considerar al intervalo $(N, b]$ para cualquier $N > -\infty$. Como los intervalos $(a_k, b_k + \delta_k)$ son una cubierta, el mismo razonamiento nos lleva a la desigualdad $\rho\big((N,b]\big) \le \sum_{k=1}^{\infty} \rho\big((a_k,b_k]\big)$. Tomando el limite cuando $N \to \infty$ obtenemos la desigualdad que buscamos. Por lo tanto $\rho\big((a,b]\big) = \sum_{k=1}^{\infty} \rho\big((a_k,b_k]\big)$. Probar que la σ-aditividad se cumple para un intervalo de la forma (a, ∞) con $a \ge -\infty$, es análogo. Podemos concluir que ρ es σ-aditiva.

Por el teorema 1.23, si μ^* es la medida exterior inducida por ρ, entonces $\mu_F = \mu^*|\mathcal{B}(\mathbb{R})$ es la única medida que extiende a ρ en $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$ y para todo $E \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\mu_F(E) &= \operatorname{ínf}\left\{\sum_{n=1}^{\infty}\rho(A_n) : A_n \in \mathcal{E},\ E \subset \bigcup_{n\in\mathbb{N}} A_n\right\}\\ &= \operatorname{ínf}\left\{\sum_{n=1}^{\infty}\big(F(b_n) - F(a_n)\big) : b_n, a_n \in \mathbb{R},\ E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n]\right\}.\end{aligned}$$

La última igualdad se ve fácilmente de que $(-\infty, a] = \bigcup_{k\in\mathbb{N}}(a-k, a-k+1]$ y $\rho\big((-\infty,a]\big) = F(a) - M^- = \sum_{k=1}^{\infty}\big(F(a-k+1) - F(a-k)\big)$. También tenemos identidades análogas para intervalos de la forma (a, ∞) para $a \geq -\infty$. $\square$

Nótese que si $G = F + c$ para alguna constante $c \in \mathbb{R}$, entonces F y G inducen la misma medida.

Observemos que la medida μ_F resulta ser la restricción de una medida exterior a $\mathcal{B}(\mathbb{R})$; esta restricción en general no será una medida completa. Por el teorema 1.15, podemos completar la medida μ_F y denotaremos por $\mathcal{M}_F$ a la σ-álgebra donde ésta es completa, que no es otra cosa que la clase de conjuntos medibles con respecto a la medida exterior generada por ρ. No es difícil ver que para todo $E \in \mathcal{M}_F$ se sigue teniendo

$$\mu_F(E) = \operatorname{ínf}\left\{\sum_{n=1}^{\infty}\big(F(b_n) - F(a_n)\big) : b_n, a_n \in \mathbb{R},\ E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n]\right\}.$$

La extensión de μ_F en $\mathcal{M}_F$ será llamada la *medida de Lebesgue-Stieltjes* con respecto a F y a menos que haya confusión la seguiremos denotando por μ_F. En el caso que $F(x) = x$, simplemente será llamada la *medida de Lebesgue*. La medida de Lebesgue suele denotarse por λ y por convención denotaremos a $\mathcal{M}_F$ por $\mathcal{L}(\mathbb{R})$. Algo interesante es que toda medida de Lebesgue-Stieltjes cumple ciertas propiedades de regularidad. Para ver esto primero probemos el siguiente lema:

Lema 1.25. *Para todo $E \in \mathcal{M}_F$,*

$$\mu_F(E) = \operatorname{ínf}\left\{\sum_{n=1}^{\infty}\mu_F\big((a_n, b_n)\big) : b_n, a_n \in \mathbb{R},\ E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n)\right\}.$$

Demostración. Si $E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n)$, entonces

$$\mu_F(E) \leq \mu_F\left(\bigcup_{n\in\mathbb{N}}(a_n, b_n)\right) \leq \sum_{n=1}^{\infty}\mu_F\big((a_n, b_n)\big).$$

Se sigue que $\mu_F(E) \leq \operatorname{ínf}\left\{\sum_{n=1}^{\infty}\mu_F\big((a_n, b_n)\big) : b_n, a_n \in \mathbb{R},\ E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n)\right\}$. Por otro lado, dado $\epsilon > 0$, deben existir intervalos $(a_n, b_n]$ tales que $E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n]$ y $\sum_{n=1}^{\infty}\big(F(b_n) - F(a_n)\big) \leq \mu_F(E) + \epsilon$. Por la continuidad por derecha de F, para todo $n \in \mathbb{N}$ debe existir $c_n > b_n$ tal que $F(c_n) \leq F(b_n) + \epsilon/2^n$. En consecuencia $E \subset \bigcup_{n\in\mathbb{N}}(a_n, c_n)$ y

$$\sum_{n=1}^{\infty}\mu_F\big((a_n, c_n)\big) \leq \sum_{n=1}^{\infty}\big(F(c_n) - F(a_n)\big) \leq \sum_{n=1}^{\infty}\big(F(b_n) - F(a_n)\big) + \epsilon \leq \mu_F(E) + 2\epsilon.$$

Como ϵ es arbitrario, se sigue la desigualdad restante y podemos concluir. $\square$

Teorema 1.26. *Para todo $E \in \mathcal{M}_F$,*

$$\mu_F(E) = \operatorname{ínf}\left\{\mu_F\big(U\big) : E \subset U \text{ y } U \textit{ es abierto}\right\}$$
$$= \sup\left\{\mu_F\big(K\big) : K \subset E \text{ y } K \textit{ es compacto}\right\}.$$

Demostración. Claramente

$$\operatorname{ínf}\left\{\mu_F\big(U\big) : E \subset U \text{ y } U \textit{ es abierto}\right\} \geq \mu_F(E)$$
$$\geq \sup\left\{\mu_F\big(K\big) : K \subset E \text{ y } K \textit{ es compacto}\right\}.$$

Ahora, por el lema anterior, para todo $\epsilon > 0$ existen intervalos (a_n, b_n) tales que $E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n)$ y $\sum_{n=1}^{\infty} \mu_F\big((a_n, b_n)\big) \leq \mu_F(E) + \epsilon$. Si tomamos $U = \bigcup_{n\in\mathbb{N}}(a_n, b_n)$, entonces $\mu_F(U) \leq \mu_F(E) + \epsilon$ y como ϵ es arbitrario obtenemos la primera igualdad. Para la segunda igualdad, primero supongamos que E es acotado. Dado $\epsilon > 0$, sea U abierto tal que $\overline{E} - E \subset U$ y $\mu(U) \leq \mu\big(\overline{E} - E\big) + \epsilon$. Entonces $K = \overline{E} - U$ es compacto, $K \subset E$ y

$$\mu_F(E) = \mu_F\big(\overline{E}\big) - \mu_F\big(\overline{E} - E\big) \leq \mu_F(K) + \mu_F(U) - \mu_F\big(\overline{E} - E\big) \leq \mu_F(K) + \epsilon.$$

Podemos concluir que se cumple la segunda igualdad cuando E es acotado. Para el caso general sea $E_j = E \cap (j, j+1]$ para todo $j \in \mathbb{Z}$. Por el caso acotado, para todo $\epsilon > 0$ existen compactos $K_j \subset E_j$ tales que $\mu_F(E_j) \leq \mu_F(K_j) + \epsilon/2^{|j|}$. Entonces

$$\mu_F(E) = \lim_{n\to\infty} \mu_F\left(\bigcup_{i=-n}^{n} E_i\right) \leq 3\epsilon + \lim_{n\to\infty} \mu_F\left(\bigcup_{i=-n}^{n} K_i\right).$$

Como ϵ es arbitrario y $\bigcup_{i=-n}^{n} K_i$ es compacto para todo $n \in \mathbb{N}$ podemos concluir. □

Retomemos a la medida de Lebesgue. Es fácil ver que para $a < b$ en $\mathbb{R}$,

$$\lambda\big((a, b)\big) = \lambda\big([a, b)\big) = \lambda\big([a, b]\big) = b - a$$

y para todo $a \in \mathbb{R}$, $\lambda\big(\{a\}\big) = 0$. En consecuencia, la medida de Lebesgue de cualquier conjunto numerable es cero. También, la medida de Lebesgue cumple las propiedades de escalamiento e invarianza ante traslaciones:
Dados $E \subset \mathbb{R}$ y $s, r \in \mathbb{R}$ definimos

$$E + s = \{y + s : y \in E\}, \qquad rE = \{ry : y \in E\}.$$

Proposición 1.27. *Para todo $E \in \mathcal{L}(\mathbb{R})$ y $r, s \in \mathbb{R}$, $E + s$, $rE \in \mathcal{L}(\mathbb{R})$ y*

$$\lambda(E + s) = \lambda(E), \qquad \lambda(rE) = |r|\lambda(E).$$

Demostración. Probaremos el enunciado para rE con $r \neq 0$. Probarlo para $E + s$ es análogo y el caso $r = 0$ es trivial. Sea

$$\Sigma_r = \{E \in \mathcal{B}(\mathbb{R}) : rE \in \mathcal{B}(\mathbb{R})\}.$$

Si $\{E\}, \{E_n\}_{n\in\mathbb{N}} \subset \Sigma_r$, entonces

$$r\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \left\{ry : y \in \bigcup_{n\in\mathbb{N}} E_n\right\} = \bigcup_{n\in\mathbb{N}} \{ry : y \in E_n\} = \bigcup_{n\in\mathbb{N}} (rE_n) \in \mathcal{B}(\mathbb{R}),$$
$$r(E^c) = \left\{rz : z \notin E\right\} = \left\{rz : rz \notin rE\right\} = \left\{y : y \notin rE\right\} = (rE)^c \in \mathcal{B}(\mathbb{R}).$$

Entonces Σ_r es σ-álgebra. Por otro lado, es fácil ver que para todo intervalo (a, b), $r(a, b)$ es otro intervalo ((ra, rb) si $r > 0$ ó (rb, ra) si $r < 0$). Por lo tanto $\Sigma_r = \mathcal{B}(\mathbb{R})$.
Ahora, sea $E \in \mathcal{B}(\mathbb{R})$ y $\{(a_n, b_n]\}_{n\in\mathbb{N}}$ tal que $E \subset \bigcup_{n\in\mathbb{N}}(a_n, b_n)$. Tenemos $rE \subset \bigcup_{n\in\mathbb{N}} r(a_n, b_n)$ y en consecuencia

$$\lambda(rE) \leq \sum_{n=1}^{\infty} \lambda\big(r(a_n, b_n)\big) = |r| \sum_{n=1}^{\infty}(b_n - a_n).$$

Por el lema 1.25 se sigue que $\lambda(rE) \leq |r|\lambda(E)$. Siguiendo el mismo razonamiento, podemos ver que $\lambda(E) = \lambda(r^{-1}rE) \leq |r^{-1}|\lambda(rE)$. Por lo tanto $\lambda(E) = |r|\lambda(E)$.
Finalmente, sea $A \in \mathcal{B}(\mathbb{R})$ un conjunto nulo. Si $B \subset A$, entonces $rB \subset rA$ y $\lambda(rA) = |r|\lambda(A) = 0$. Dado que $\mathcal{L}(\mathbb{R}) = \big\{E \cup B : E \in \mathcal{B}(\mathbb{R}),\ B \subset A \in \mathcal{N}\big\}$, podemos concluir que para todo $E \in \mathcal{L}(\mathbb{R})$, $rE \in \mathcal{L}(\mathbb{R})$ y $\lambda(rE) = |r|\lambda(E)$. □

Una pregunta natural que puede surgir es si $\mathcal{L}(\mathbb{R}) = \mathcal{P}(\mathbb{R})$. De acuerdo al axioma de elección este no es el caso. Diremos que $x \sim y$ si y sólo si $x - y \in \mathbb{Q}$. Es fácil ver que $\sim$ es una relación de equivalencia en $\mathbb{R}$. Dado que para todo $x \in \mathbb{R}$ existe $q \in \mathbb{Q}$ tal que $x - q \in (0, 1)$, podemos encontrar al menos un representante de cada clase de equivalencia en el intervalo $(0, 1)$. Así, por el axioma de elección podemos tomar un conjunto $E \subset (0, 1)$ que contenga un único punto de cada clase de equivalencia. Veremos que E no puede estar en $\mathcal{L}(\mathbb{R})$. Para esto requerimos las siguientes afirmaciones:

1) Si $x \in (0, 1)$ existe $q \in (-1, 1) \cap \mathbb{Q}$ tal que $x \in E + q$.

2) Si $q, r \in \mathbb{Q}$ y $q \neq r$, entonces $(E + q) \cap (E + r) = \emptyset$.

1) Existe $y \in E$ tal que $x - y \in \mathbb{Q}$. Claramente $q = x - y \in (-1, 1)$ y $x \in E + q$.
2) Si $x \in (E + q) \cap (E + r)$, entonces existen $y, z \in E$ tales que $x = y + q = z + r$. Entonces $y - z = q - r$ por lo que $y \neq z$ pero $y \sim z$. Esto contradice la manera en que E fue seleccionado.
Ahora, sea $A = (-1, 1) \cap \mathbb{Q}$ y $C = \bigcup_{q\in A}(E + q)$. Si tuviéramos $E \in \mathcal{L}(\mathbb{R})$, entonces también $C \in \mathcal{L}(\mathbb{R})$ y $\lambda(E + q) = \lambda(E)$ para todo $q \in A$. Sin embargo $(0, 1) \subset C \subset (-1, 2)$ y en consecuencia $1 \leq \lambda(C) \leq 3$. Esto es imposible ya que por un lado es necesario que $\lambda(E) > 0$ para que $1 \leq \lambda(C)$. Por otro lado, dado que los conjuntos $E + q$ son disjuntos, $\lambda(C) = \sum_{q\in A} \lambda(E + q)$ y la única manera de que este valor esté acotado es que $\lambda(E) = 0$. Por lo tanto E no puede ser un conjunto Lebesgue-medible.

Otra pregunta que puede hacerse es si existe una generalización de la medida de Lebesgue-Stieltjes para dimensiones mayores a uno. La respuesta es que si.

Definición 1.28. Sea $F : \mathbb{R}^n \to \mathbb{R}$,

1) Decimos que F es continua por la derecha si para todo x y $\{x_k\}_{k\in\mathbb{N}}$ en $\mathbb{R}^n$, tales que $x_{k+1} \leq x_k$ y $x_k \downarrow x$, se tiene $\lim_{k\to\infty} F(x_k) = F(x)$.

2) Diremos que F es n-creciente si para $a \leq b$, $F(a) \leq F(b)$ y

$$\sum_{k=0}^{n}(-1)^k \sum_{c\in\Delta_{a,b}^k} F(c) \geq 0,$$

donde $\Delta_{a,b}^k$ es el conjunto de vectores en $\mathbb{R}^n$ tales que k entradas coinciden con las respectivas entradas de a y el resto coinciden con las respectivas entradas de b.

Por ejemplo, en el caso $n = 2$, $\Delta_{a,b}^0 = \{(b_1, b_2)\}$, $\Delta_{a,b}^1 = \{(a_1, b_2), (b_1, a_2)\}$, $\Delta_{a,b}^2 = \{(a_1, a_2)\}$ y se tiene

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

Siguiendo la misma notación, tenemos:

Teorema 1.29. *Sea $F : \mathbb{R}^n \to \mathbb{R}$ una función n-creciente y continua por la derecha. Entonces existe una única medida $\mu_F : \mathcal{B}(\mathbb{R}^n) \to [0, \infty]$ tal que para todo $a \leq b$ en $\mathbb{R}^n$, $\mu_F\big((a, b]\big) = \sum_{k=0}^n (-1)^k \sum_{c \in \Delta_{a,b}^k} F(c)$. Más aún, para todo $E \in \mathcal{B}(\mathbb{R}^n)$,*

$$\mu_F(E) = \text{ínf}\left\{ \sum_{j=1}^{\infty} \mu_F\big((a_j, b_j]\big) : b_j, a_j \in \mathbb{R}^n,\ E \subset \bigcup_{j \in \mathbb{N}} (a_j, b_j] \right\}.$$

El teorema 1.29 no se probará; la demostración es en esencia la misma que del caso unidimensional. Sin embargo, las complicaciones son producto de las variantes que se generan al considerar más dimensiones. Nótese que para el caso $F(x) = \prod_{i=1}^n x_i$,

$$\mu_F\big((a, b]\big) = \prod_{i=1}^n (b_i - a_i).$$

Se sigue que F induce a la medida de Lebesgue en $\mathbb{R}^n$. Más adelante tendremos las herramientas para extender la medida de Lebesgue a dimensiones mayores de manera mucho más sencilla.

Capítulo 2

Funciones medibles e integración

2.1. Medibilidad

Hasta ahora se han introducido los conceptos de espacios medibles y medidas, así como algunas de sus propiedades. El objetivo ahora es describir algunas funciones que van de un espacio medible a otro. Tomemos como analogía la definición topológica de una función continua; es decir, una funcion que va de un espacio topológico a otro y tal que la preimagen de conjuntos abiertos sea abierto. En el caso de espacios medibles la clase de funciones que nos inteserarán son las *funciones medibles*:

Definición 2.1. Sean (X, Σ_X) y (Y, Σ_Y) dos espacios medibles. Diremos que $f : X \to Y$ es $(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y})$**-medible**, o **medible** si no se presta a confusión, si y sólo si $f^{-1}(A) \in \Sigma_X$ cuando $A \in \Sigma_Y$.

Si es el caso que (X, Σ_X), (Y, Σ_Y) y (Z, Σ_Z) son espacios medibles y $f : X \to Y$ y $g : Y \to Z$ son funciones medibles, entonces es claro que $g \circ f : X \to Z$ también será una función medible. En general, puede ser muy complicado demostrar que la preimagen de cualquier conjunto en Σ_Y terminará siendo un elemento en Σ_X. La siguiente proposición nos da un criterio más simple para determinar si una función es medible.

Proposición 2.2. *Sean (X, Σ_X) y (Y, Σ_Y) dos espacios medibles, $\mathcal{C} \subset \mathcal{P}(Y)$ tal que $\sigma(\mathcal{C}) = \Sigma_Y$. Entonces $f : X \to Y$ es medible si y sólo si $f^{-1}(A) \in \Sigma_X$ para todo $A \in \mathcal{C}$.*

Demostración. Notemos que $\mathcal{C} \subset \Sigma_Y$, estableciendo la necesidad. Para la suficiencia supongamos que $f^{-1}[\mathcal{C}] \subset \Sigma_X$ y consideremos el conjunto

$$\mathcal{K} = \left\{ A \subset Y : f^{-1}(A) \in \Sigma_X \right\}.$$

Es claro que $\mathcal{C} \subset \mathcal{K}$ y por propiedades de la preimagen, $\mathcal{K}$ es cerrada bajo complementos y uniones contables, además de que $Y \in \mathcal{K}$. Por lo tanto $\mathcal{K}$ es una σ-álgebra. El hecho de que $\Sigma_Y = \sigma(\mathcal{C})$ implica que $\Sigma_Y \subset \mathcal{K}$, concluyendo la demostración. □

Recordando que se le dió una estructura de medibilidad a los espacios topológicos, podremos relacionar el conjunto de funciones continuas con el conjuntos de funciones medibles.

Corolario 2.3. *Si X y Y son espacios topológicos con borelianos $\mathcal{B}(X)$ y $\mathcal{B}(Y)$ respectivamente, entonces toda función $f : X \to Y$ continua es medible.*

Demostración. Como f es continua, para todo $U \subset Y$ abierto tendremos que $f^{-1}(U) \subset X$ será abierto y por tanto $f^{-1}(U) \in \mathcal{B}(X)$. Por la proposición anterior podemos concluir. □

En el contexto del corolario anterior, a las funciones que son $(\mathcal{B}(X), \mathcal{B}(Y))$-medibles las llamamos *funciones Borel-medibles*. Otro corolario directo de la proposición 2.2 es el siguiente.

Corolario 2.4. *Si (X, Σ) es un espacio medible y se considera la función $f : X \to \mathbb{R}^n$ entonces tendremos las siguientes equivalencias:*

1) f *es* $(\Sigma, \mathcal{B}(\mathbb{R}^n))$*-medible.*

2) $f^{-1}((a, b)) \in \Sigma$ *para todo* $a, b \in \mathbb{R}^n$ *con* $a < b$*.*

3) $f^{-1}((a, b]) \in \Sigma$ *para todo* $a, b \in \mathbb{R}^n$ *con* $a < b$*.*

4) $f^{-1}([a, b)) \in \Sigma$ *para todo* $a, b \in \mathbb{R}^n$ *con* $a < b$*.*

5) $f^{-1}([a, b]) \in \Sigma$ *para todo* $a, b \in \mathbb{R}^n$ *con* $a < b$*.*

6) $f^{-1}([a, \infty)) \in \Sigma$ *para todo* $a \in \mathbb{R}^n$*.*

7) $f^{-1}((-\infty, b]) \in \Sigma$ *para todo* $b \in \mathbb{R}^n$*.*

Demostración. De las proposiciones 1.5 y 2.2 se sigue la equivalencia entre a) y cada uno de los demás inicisos. □

Proposición 2.5. *Sea (X, Σ) un espacio medible, $\{(Y_\alpha, \Sigma_\alpha)\}_{\alpha \in A}$ una colección de espacios medibles, $(Y, \Sigma') = \left(\prod_{\alpha \in A} Y_\alpha, \bigotimes_{\alpha \in A} \Sigma_\alpha\right)$ y $\pi_\alpha : Y \to Y_\alpha$ el mapeo proyección. Entonces $f : X \to Y$ será medible si y sólo si $f_\alpha = \pi_\alpha \circ f$ es medible para cada $\alpha \in A$.*

Demostración. Recordemos que $\bigotimes_{\alpha \in A} \Sigma_\alpha = \sigma\left(\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \Sigma_\alpha, \alpha \in A\}\right)$, de donde π_α es medible para cada α. Por lo tanto si f es medible entonces f_α será medible por ser composición de funciones medibles.

Por otra parte, si f_α es medible para cada α, entonces para cada $E_\alpha \in \Sigma_\alpha$ se tendrá que $f^{-1}\left(\pi_\alpha^{-1}(E_\alpha)\right) = f_\alpha^{-1}(E_\alpha) \in \Sigma$. Se concluye la demostración por la proposición 2.2. □

La proposición anterior va a resultar ser de extrema utilidad para probar que algunas operaciones elementales en espacios vectoriales serán medibles.

Corolario 2.6. *Sea (X, Σ) un espacio medible. Entonces $f : X \to \mathbb{C}$ es $(\Sigma, \mathcal{B}(\mathbb{C}))$-medible si y sólo si $\Re f$ y $\Im f$ son $(\Sigma, \mathcal{B}(\mathbb{R}))$-medibles.*

Demostración. Se sigue fácilmente de la proposición 2.5 notando que $\mathcal{B}(\mathbb{C}) = \mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ y que $f \equiv (\Re f, \Im f)$. □

Proposición 2.7. *Sean (X, Σ_X), (Y, Σ_Y) y (Z, Σ_Z) espacios medibles y $\{f_i : X \to Y\}_{1 \le i \le d}$ una colección de funciones medibles para $d \in \mathbb{N}$. Si $g : Y^d \to Z$ es $(\bigotimes_{n=1}^d \Sigma_Y, \Sigma_Z)$-medible entonces $g(f_1, \ldots, f_d) : X \to Z$, definida por $x \mapsto g(f_1(x), \ldots, f_d(x))$, es medible.*

Demostración. Sea $F : X \to Y^d$ la función dada por $x \mapsto (f_1(x), \ldots, f_d(x))$. Por la proposición anterior, F es $(\Sigma_X, \bigotimes_{n=1}^d \Sigma_Y)$-medible. Como la composición de funciones medibles es medible, $g(f_1, \ldots, f_d) = g \circ F$ es medible, como se quería demostrar. □

Proposición 2.8. *Sea (X, Σ) un espacio medible y $d \in \mathbb{N}$. Entonces se cumplen:*

1) Si $f, g : X \to \mathbb{R}^d$ son medibles, entonces para todo $a, b \in \mathbb{R}$, $af + bg$ es medible.

2) Si $f, g : X \to \mathbb{R}$ son medibles, entonces fg es medible.

2') Si $f, g : X \to \mathbb{C}$ son medibles, entonces para todo $\alpha, \beta \in \mathbb{C}$, $\alpha f + \beta g$ y fg son medibles.

Demostración. 1) El operador $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ dado por $T(x, y) = ax + by$ es lineal y por ende continuo. Del corolario 2.3 tenemos que T es $(\mathcal{B}(\mathbb{R}^{2d}), \mathcal{B}(\mathbb{R}^d))$-medible. Por otro lado, dado que $\mathcal{B}(\mathbb{R}^{2d}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)$, de la proposición 2.7 podemos concluir que $af + bg = T(f, g)$ es medible.
2) Es análogo al inciso 1) usando que el producto como operador es continuo.
2') También se puede probar de manera análoga a los incisos anteriores. $\square$

A veces será necesario considerar funciones con valores en $\overline{\mathbb{R}}$ por cuestiones técnicas. Consideremos

$$B = \{(a, b) : a, b \in \mathbb{R}\} \cup \{(a, \infty] : a \in \mathbb{R}\} \cup \{[-\infty, b) : b \in \mathbb{R}\}.$$

Se puede verificar fácilmente que B es una base topológica en $\overline{\mathbb{R}}$ por lo que las uniones arbitrarias de elementos en B definen una topología. En el apéndice C se prueba que esta colección de conjuntos abiertos coincide con la topología inducida por la métrica $d : \overline{\mathbb{R}} \times \overline{\mathbb{R}} \to \mathbb{R}^+$ dada por $(x, y) \mapsto 2|\arctan(x) - \arctan(y)|/\pi$ y la restricción de esta métrica en $\mathbb{R}$ es equivalente a la métrica euclidiana. También se puede verificar que bajo esta topología,

$$\mathcal{B}(\overline{\mathbb{R}}) = \left\{E \subseteq \overline{\mathbb{R}} : E \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\right\}.$$

Proposición 2.9. $\mathcal{B}(\overline{\mathbb{R}}) = \sigma\left(\{[-\infty, b) : b \in \mathbb{R}\}\right)$
$= \sigma\left(\{(a, \infty] : a \in \mathbb{R}\}\right).$

Demostración. Es claro que para $a, b \in \mathbb{R}$ tendremos que

$$[-\infty, b) = \overline{\mathbb{R}} - \bigcap_{n=1}^{\infty}(b - 1/n, \infty]$$

y

$$(a, \infty] = \overline{\mathbb{R}} - \bigcap_{n=1}^{\infty}[-\infty, a + 1/n),$$

de donde $\sigma\left(\{[-\infty, b) : b \in \mathbb{R}\}\right) = \sigma\left(\{(a, \infty] : a \in \mathbb{R}\}\right)$. Además tendremos que

$$(a, b) = [-\infty, b) \cap (a, \infty]$$

para $a, b \in \mathbb{R}$ arbitrarios. Es fácil ver que todo abierto en $\overline{\mathbb{R}}$ es la unión numerable de abiertos en la base B, de donde se sigue que $\mathcal{B}(\overline{\mathbb{R}}) = \sigma\left(\{[-\infty, b) : b \in \mathbb{R}\}\right)$. $\square$

Así cuando consideremos un espacio medible (X, Σ) podremos hablar de funciones medibles con valores en $\overline{\mathbb{R}}$, las cuales serán fundamentales más adelante. Algunos resultados de medibilidad para estas funciones se dan a continuación.

Proposición 2.10. *Sea (X, Σ) un espacio medible y $f : X \to \overline{\mathbb{R}}$. Serán equivalentes:*

1) f es $(\Sigma, \mathcal{B}(\overline{\mathbb{R}}))$-medible.

2) $f^{-1}([-\infty, b)) \in \Sigma$ para toda $b \in \mathbb{R}$.

3) $f^{-1}([-\infty, b]) \in \Sigma$ *para toda* $b \in \mathbb{R}$.

4) $f^{-1}((a, \infty]) \in \Sigma$ *para toda* $a \in \mathbb{R}$.

5) $f^{-1}([a, \infty]) \in \Sigma$ *para toda* $a \in \mathbb{R}$.

6) $f|_{f^{-1}(\mathbb{R})}$ *es* $(\Sigma \cap f^{-1}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$*-medible y* $f^{-1}(\{-\infty\}), f^{-1}(\{\infty\}) \in \Sigma$.

La demostración de esta proposición es análoga al corolario 2.4 y se deja como ejercicio al lector. Cabe mencionar que toda función $(\Sigma, \mathcal{B}(\mathbb{R}))$-medible es $(\Sigma, \mathcal{B}(\overline{\mathbb{R}}))$-medible ya que

$$f^{-1}\big((a, \infty]\big) = f^{-1}\big((a, \infty)\big) \cup f^{-1}\big(\{\infty\}\big).$$

Proposición 2.11. *Si* (X, Σ) *es un espacio medible y* $\{f_n : X \to \overline{\mathbb{R}}\}_{n\in\mathbb{N}}$ *es una colección de funciones* $(\Sigma, \mathcal{B}(\overline{\mathbb{R}}))$*-medibles, entonces las siguientes funciones también lo serán*

$$\sup_{n\in\mathbb{N}} f_n, \ \inf_{n\in\mathbb{N}} f_n, \ \limsup_{n\to\infty} f_n, \ \liminf_{n\to\infty} f_n.$$

Si es el caso que $f = \lim_{n\to\infty} f_n$ *existe, entonces* $f = \limsup_{n\to\infty} f_n = \liminf_{n\to\infty} f_n$ *y será medible.*

Demostración. Para $b \in \mathbb{R}$ tendremos que $g(x) := \sup_{n\in\mathbb{N}} f_n(x) \leq b$ si y sólo si $f_n(x) \leq b$ para toda $n \in \mathbb{N}$. Por lo tanto

$$g^{-1}([-\infty, b]) = \bigcap_{n\in\mathbb{N}} f_n^{-1}([-\infty, b])$$

y se concluye que g es medible por la proposición anterior. Por otra parte se tiene que $\inf_{n\in\mathbb{N}} f_n = -\sup_{n\in\mathbb{N}}(-f_n)$, $\limsup_{n\to\infty} f_n = \inf_{n\in\mathbb{N}} \sup_{m\geq n} f_m$ y $\liminf_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} \inf_{m\geq n} f_m$, estableciendo la primera parte del resultado. La segunda parte se establece por medio de la proposición A.1. $\square$

Corolario 2.12. *Si* $f, g : X \to \overline{\mathbb{R}}$ *son medibles entonces* $f \wedge g := \min(f, g)$ *y* $f \vee g := \max(f, g)$ *también lo son.*

Corolario 2.13. *Si* (X, Σ) *es un espacio medible y* $\{f_n : X \to \mathbb{C}\}_{n\in\mathbb{N}}$ *es una colección de funciones* $(\Sigma, \mathcal{B}(\mathbb{C}))$*-medibles tal que* $\lim_{n\to\infty} f_n$ *existe, entonces es medible.*

Demostración. Notemos que la existencia de $f = \lim_{n\to\infty} f_n$ implica la existencia de $u = \lim_{n\to\infty} \Re f_n$ y $v = \lim_{n\to\infty} \Im f_n$ y además $f = u + iv$. Por el corolario 2.6 y la proposición 2.11 podemos concluir. $\square$

Ahora introducimos la descomposición para funciones con valores en $\overline{\mathbb{R}}$ en *parte positiva* y *parte negativa* de f dadas por

$$f^+ := f \vee 0 \quad \text{y} \quad f^- := -f \vee 0 = -(f \wedge 0),$$

respectivamente. Su medibilidad se sigue del corolario 2.12. Estas funciones nos permiten descomponer tanto f como $|f|$ de la siguiente manera:

$$f = f^+ - f^- \quad \text{y} \quad |f| = f^+ + f^-.$$

Para finalizar la sección introducimos una clase de funciones que será el bloque básico de la teoría de integración, además de ser una clase que nos permitirá demostrar algunos resultados de forma más sencilla.

Definición 2.14. Sea (X, Σ) un espacio medible. Para $E \subset X$ la **función indicadora** o **función característica** de E está dada por

$$\mathbb{1}_E(x) \equiv \chi_E(x) := \begin{cases} 1 & \textit{si } x \in E, \\ 0 & \textit{si } x \notin E. \end{cases}$$

Nota: la función indicadora se puede considerar como compleja o como función con valores en los reales extendidos; esta sutileza quedará clara del contexto.

Proposición 2.15. *La función indicadora $\mathbb{1}_E$ de $E \subset X$ es medible si y sólo si $E \in \Sigma$.*

Demostración. Si la función indicadora $\mathbb{1}_E$ es medible entonces $E = \mathbb{1}_E^{-1}(\{1\}) \in \Sigma$. Por otra parte, si $E \in \Sigma$ y $U \in \mathcal{B}(\overline{\mathbb{R}})$ o $U \in \mathcal{B}(\mathbb{C})$ entonces

$$\mathbb{1}_E^{-1}(U) = \begin{cases} \emptyset & \textit{si } 0, 1 \notin U, \\ E & \textit{si } 0 \notin U, 1 \in U, \\ X - E & \textit{si } 0 \in U, 1 \notin U, \\ X & \textit{si } 0, 1 \in U, \end{cases}$$

de donde se sigue la medibilidad de $\mathbb{1}_E$. □

Definición 2.16. Si (X, Σ) es un espacio medible y φ es una combinación lineal finita de indicadoras medibles con coeficientes complejos o reales (no extendidos) entonces diremos que φ es una **función simple** o función **Σ-simple**.

Alternativamente podemos definir a las funciones simples como funciones medibles que toman una cantidad finita de valores. Así si φ es una función medible tal que $\varphi(x) \in \{\alpha_1, \ldots, \alpha_n\}$ para $x \in X$, denotando por $E_j := \varphi^{-1}(\{\alpha_j\})$ para $j \in \{1, \ldots, n\}$, entonces $\varphi = \sum_{j=1}^{n} \alpha_j \mathbb{1}_{E_j}$. Ésta última es la *representación canónica* de φ y exhibe a la función simple como combinación lineal de indicadoras de conjuntos medibles ajenos tales que $X = \bigcup_{i=1}^{n} E_i$.

Proposición 2.17. *Sean (X, Σ) un espacio medible y $\{\varphi_i\}_{1 \leq i \leq n}$ una colección de funciones simples con $n \in \mathbb{N}$. Entonces $\sum_{i=1}^{n} \varphi_i$ y $\prod_{i=1}^{n} \varphi_i$ son funciones simples y si α es un escalar entonces también $\alpha\varphi_1$ será simple.*

Demostración. Que $\alpha\varphi_1$ es simple es trivial. Notemos que si $\varphi_1 = \sum_{i=1}^{n_1} \alpha_{1i} \mathbb{1}_{E_{1i}}$ y $\varphi_2 = \sum_{i=1}^{n_2} \alpha_{2i} \mathbb{1}_{E_{2i}}$ son simples y definimos $E_{ij} = E_{1i} \cap E_{2j}$ para $i \in \{1, \ldots, n_1\}$ y $j \in \{1, \ldots, n_2\}$ entonces

$$\varphi_1 + \varphi_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\alpha_{1i} + \alpha_{2j}) \mathbb{1}_{E_{ij}},$$

y

$$\varphi_1 \varphi_2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (\alpha_{1i} \alpha_{2j}) \mathbb{1}_{E_{ij}},$$

estableciendo el resultado para $n = 2$. El caso general se sigue directamente por medio de argumento inductivo. □

La siguiente proposición pondrá de manifiesto el papel de las funciones simples como las funciones aproximadoras por excelencia de funciones medibles.

Teorema 2.18. *Sea (X, Σ) un espacio medible.*

1) Si $f : X \to \overline{\mathbb{R}}$ es medible y $f \geq 0$ entonces existe una sucesión $\{\varphi_n\}_{n\in\mathbb{N}}$ de funciones simples tales que $0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$, $\lim_{n\to\infty} \varphi = f$ y $\{\varphi_n\}_{n\in\mathbb{N}}$ converge de forma uniforme a f en cualquier conjunto en el que f esté acotada.

2) Si $f : X \to \mathbb{C}$ es medible entonces existe una sucesión $\{\varphi_n\}_{n\in\mathbb{N}}$ de funciones simples tales que $0 \leq |\varphi_1| \leq |\varphi_2| \leq \cdots \leq |f|$, $\lim_{n\to\infty} \varphi = f$ y $\{\varphi_n\}_{n\in\mathbb{N}}$ converge de forma uniforme a f en cualquier conjunto en el que f esté acotada.

Demostración. 1) Para $n \in \mathbb{N} \cup \{0\}$ definamos $\varphi_n = (2^{-n}\lfloor 2^n f \rfloor) \wedge 2^n$. Para describir estas funciones, consideremos $k \in \{0, \ldots, 2^{2n} - 1\}$, definamos $E_n^k = f^{-1}([k2^{-n}, (k+1)2^{-n}))$, $F_n = f^{-1}([2^n, \infty])$ y notemos que

$$\varphi_n = \sum_{k=0}^{2^{2n}-1} k2^{-n}\mathbb{1}_{E_n^k} + 2^n \mathbb{1}_{F_n}.$$

Para ver que $\varphi_n \leq \varphi_{n+1}$ primero veamos que si $x \in X$ es tal que $f(x) \geq 2^n$ entonces $\varphi_n(x) = 2^n \leq 2^{n+1}$ y $2^n \leq 2^{-n-1}\lfloor 2^{n+1} f(x) \rfloor$, de donde $\varphi_n(x) \leq \varphi_{n+1}(x)$. Si ahora $x \in X$ es tal que $k \leq f(x)2^n < k+1 \leq 2^{2n}$ entonces $\varphi_n(x) = k2^{-n} = (2k)2^{-n-1} \leq 2^{-n-1}\lfloor 2^{n+1} f(x) \rfloor < (2k+2)2^{-n-1} = (k+1)2^{-n}$, de donde $\varphi_n(x) \leq \varphi_{n+1}(x)$. Además tendremos que $0 \leq f(x) - \varphi_n(x) = 2^{-n}(2^n f(x) - \lfloor 2^n f(x) \rfloor) \leq 2^{-n}$ para $x \in X$ tal que $f(x) \leq 2^n$ de donde se siguen los resultados de límites.

2) Al descomponer $f = (\Re f)^+ - (\Re f)^- + i(\Im f)^+ - i(\Im f)^-$, podemos aplicar el inciso (a) a $(\Re f)^+, (\Re f)^-, (\Im f)^+$ y $(\Im f)^-$, obteniendo las sucesiones de funciones simples $\{\psi_n^+\}, \{\psi_n^-\}, \{\varsigma_n^+\}$ y $\{\varsigma_n^-\}$ con las propiedades dadas en el inciso (a). Notando que

$$|f| = \left[\left((\Re f)^+\right)^2 + \left((\Re f)^-\right)^2 + \left((\Im f)^+\right)^2 + \left((\Im f)^-\right)^2\right]^{1/2},$$

se sigue el resultado. $\square$

Para finalizar la sección, presentaremos un resultado con el que comenzaremos a ilustrar la importancia de trabajar con medidas completas. Para esto recordamos el concepto de *casi donde sea*, aterrizado a funciones:
Sea (X, Σ, μ) un espacio de medida y $f : X \to Y$ cualquier función. Para cualquier propiedad P, diremos que f cumple P **μ-casi donde sea** (abreviado μ-c.d.s ó simplemente c.d.s.) si y sólo si el conjunto donde f no cumple P es subconjunto de algún conjunto nulo.

Proposición 2.19. *Sea (X, Σ, μ) un espacio de medida completo, entonces se cumplen*

1) Si $f : X \to Y$ es (Σ, Σ')-medible y $f = g$ μ-c.d.s., entonces g también es (Σ, Σ')-medible,

2) si $\{f_n : X \to \overline{\mathbb{R}}\}_{n\in\mathbb{N}}$ o $\{f_n : X \to \mathbb{C}\}_{n\in\mathbb{N}}$ es una sucesión de funciones Σ-medibles y $f_n \to f$ μ-c.d.s., entonces f es Σ-medible.

Demostración. 1) Por hipótesis $N = \{x \in X : f(x) \neq g(x)\}$ es subconjunto de un conjunto nulo por lo que $N \in \Sigma$. Entonces para $E \in \Sigma'$ tendremos que

$$g^{-1}(E) = (g^{-1}(E) \cap N^c) \cup (g^{-1}(E) \cap N) = (f^{-1}(E) \cap N^c) \cup (g^{-1}(E) \cap N),$$

que resulta ser medible porque f es medible y $g^{-1}(E) \cap N \subset N$.

2) Consideremos Y como $\mathbb{C}$ o $\overline{\mathbb{R}}$, dependiendo del caso. Sea N el conjunto nulo en el cual $\{f_n\}_{n\in\mathbb{N}}$ no converge a f. Definamos $g_n = f_n \mathbb{1}_{N^c}$ y notemos que

$$g(x) := \lim_{n\to\infty} g_n(x) = \begin{cases} f(x) & \text{si } x \notin N, \\ 0 & \text{si } x \in N. \end{cases}$$

Por la proposición 2.11 g es medible y por el inciso anterior se sigue que f es medible. □

Finalmente veremos que incluso si se nos olvida la hipótesis de que el espacio de medida sea completo no habrá mucho problema.

Proposición 2.20. *Sean (X, Σ, μ) un espacio de medida, $(X, \overline{\Sigma}, \overline{\mu})$ su compleción y f una función $\overline{\Sigma}$-medible. Entonces existirá una función Σ-medible, g, tal que $f = g$ $\overline{\mu}$-c.d.s.*

Demostración. Del teorema 1.15 sabemos que $\overline{\Sigma} = \{E \cup B : E \in \Sigma, B \subset A \in \mathcal{N}\}$, donde $\mathcal{N} = \{A \in \Sigma : \mu(A) = 0\}$. Es claro que $\mathbb{1}_{E\cup B}(x) = \mathbb{1}_E(x)$ para toda $x \notin B - E$ y $\overline{\mu}(B - E) = 0$. Así, bajo la representación canónica, las funciones $\phi = \sum_{i=1}^n \alpha_i \mathbb{1}_{E_i \cup B_i}$ y $\varphi = \sum_{i=1}^d \alpha_i \mathbb{1}_{E_i}$, $\overline{\Sigma}$-simple y Σ-simple respectivamente, serán iguales salvo en el conjunto nulo $\bigcup_{i=1}^n (B_i - E_i)$. Finalmente, por el teorema 2.18, para una función $\overline{\Sigma}$-medible f, podemos construir una sucesión de funciones $\overline{\Sigma}$-simples $\{\phi_n\}$ tal que $\lim_n \phi_n = f$. Sea φ_n una función Σ-simple que sea igual a ϕ_n salvo en un conjunto nulo N_n. Entonces existen $M_n \in \Sigma$ con $\mu(M_n) = 0$ tales que $N_n \subset M_n$. Definamos $M = \bigcup_{n=0}^\infty M_n$, para el que $\overline{\mu}(M) = 0$, y definamos $g := \lim_n \varphi_n \mathbb{1}_{M^c}$, que es una función Σ-medible, además de que $g(x) = f(x)$ para $x \notin M$. □

2.2. Integración

En las secciones anteriores se desarrollaron los elementos necesarios para definir la integral de funciones medibles con respecto a una medida. Con este fin consideraremos (X, Σ, μ) un espacio de medida y definimos

$$L^+(X, \Sigma) = \left\{ f : X \to \overline{\mathbb{R}} : f \textit{ es } (\Sigma, \mathcal{B}(\overline{\mathbb{R}}))\textit{-medible y } f \geq 0 \right\}.$$

Cuando no haya posibilidad de confusión se escribirá únicamente L^+.

Definición 2.21. Sea $\varphi \in L^+$ una función simple con representación canónica $\varphi = \sum_{j=1}^n \alpha_j \mathbb{1}_{E_j}$. Definimos la **integral** de φ con respecto a μ como el número real no negativo, posiblemente extendido, dado por

$$\int \varphi(x)\mu(dx) = \int \varphi(x) d\mu(x) = \int \varphi d\mu := \sum_{j=1}^n \alpha_j \mu(E_j).$$

Si además tenemos $E \in \Sigma$ es claro que $\varphi \mathbb{1}_E$ también es una función simple lo que nos permite definir la integral de φ sobre E:

$$\int_E \varphi d\mu := \int \varphi \mathbb{1}_E d\mu = \sum_{j=1}^{n} \alpha_j \mu(E_j \cap E).$$

Algunos resultados con respecto a la integral de funciones simples se resumen en la siguiente proposición.

Proposición 2.22. *Sean $\varphi, \psi \in L^+$ funciones simples.*

1) Si $c \geq 0$ entonces $\int c\varphi d\mu = c \int \varphi d\mu$.

2) $\int (\varphi + \psi) d\mu = \int \varphi d\mu + \int \psi d\mu$.

3) Si $\varphi \leq \psi$ entonces $\int \varphi d\mu \leq \int \psi d\mu$.

4) La función $\nu_\varphi : \Sigma \to [0, \infty]$ dada por $\nu_\varphi(A) := \int_A \varphi d\mu$ es una medida.

Demostración. 1) Si $c = 0$ entonces $c\varphi = 0 \cdot \mathbb{1}_X$ y así $\int c\varphi d\mu = 0 \cdot \mu(X) = 0 = c \int \varphi d\mu$. Si $c > 0$ y $\sum_{j=1}^{n} \alpha_i \mathbb{1}_{E_i}$ es la representación canónica de φ entonces la representación canónica de $c\varphi$ está dada por $c\varphi = \sum_{i=1}^{n} (c\alpha_i) \mathbb{1}_{E_i}$, por lo que

$$\int c\varphi d\mu = \sum_{i=1}^{n} (c\alpha_i) \mu(E_i) = c \sum_{i=1}^{n} \alpha_i \mu(E_i) = c \int \varphi d\mu.$$

2) Sean $\sum_{i=1}^{n} \alpha_i \mathbb{1}_{E_i}$ y $\sum_{j=1}^{m} \beta_j \mathbb{1}_{F_j}$ las representaciones canónicas de φ y ψ respectivamente. Entonces es claro que $\varphi + \psi = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) \mathbb{1}_{E_i \cap F_j}$ y, aunque puede que ésta no sea la representación canónica, dado que $\mathbb{1}_A + \mathbb{1}_B = \mathbb{1}_{A \cup B}$ si y sólo si $A \cap B = \emptyset$, basta agrupar los sumandos para los cuales $\alpha_i + \beta_j$ coinciden. Dado que $E_i = \bigcup_{j=1}^{m} (E_i \cap F_j)$ y $F_j = \bigcup_{i=1}^{n} (E_i \cap F_j)$,

$$\begin{aligned}\int (\varphi + \psi) d\mu &= \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) \mu(E_i \cap F_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \mu(E_i \cap F_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j \mu(E_i \cap F_j) \\ &= \sum_{i=1}^{n} \alpha_i \mu(E_i) + \sum_{j=1}^{m} \beta_j \mu(F_j) = \int \varphi d\mu + \int \psi d\mu.\end{aligned}$$

3) Usando las mismas representaciones canónicas que en el inciso anterior vemos que $\varphi \leq \psi$ si y sólo si $\alpha_i \leq \beta_j$ cuando $E_i \cap F_j \neq \emptyset$, por lo que

$$\int \varphi d\mu = \sum_{i=1}^{n} \alpha_i \mu(E_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \mu(E_i \cap F_j) \leq \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j \mu(E_i \cap F_j) = \sum_{j=1}^{m} \beta_j \mu(F_j) = \int \psi d\mu.$$

4) Que ν_φ es no negativa y $\nu_\varphi(\emptyset) = 0$ es claro. Si $\sum_{i=1}^{m} \alpha_i \mathbb{1}_{E_i}$ es la representación canónica de φ y $\{A_n\}_{n \in \mathbb{N}} \subset \Sigma$ es una colección de conjuntos ajenos dos a dos entonces,

$$\begin{aligned}\nu_\varphi\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \int_{\bigcup_{n \in \mathbb{N}} A_n} \varphi d\mu = \sum_{i=1}^{m} \alpha_i \mu\left(E_i \cap \bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{i=1}^{m} \sum_{n \in \mathbb{N}} \alpha_i \mu(E_i \cap A_n) \\ &= \sum_{n \in \mathbb{N}} \sum_{i=1}^{m} \alpha_i \mu(E_i \cap A_n) = \sum_{n \in \mathbb{N}} \int_{A_n} \varphi d\mu = \sum_{n \in \mathbb{N}} \nu_\varphi(A_n).\end{aligned}$$

□

Ahora extendemos la noción de integral para funciones $f \in L^+$.

Definición 2.23. Sea $f \in L^+$. La **integral** de f con respecto a μ se define como

$$\int f d\mu := \sup \left\{ \int \varphi d\mu : \varphi \textit{ es simple y } 0 \leq \varphi \leq f \right\}.$$

Si $E \in \Sigma$ se define la integral de f sobre E como

$$\int_E f d\mu := \int f \mathbb{1}_E d\mu.$$

Para que la definición dada tenga sentido se necesita probar que, en efecto, esto extiende la noción de integral que tenemos para funciones simples. Este es claramente el caso, pues por definición y del inciso 3) de la proposición anterior, tenemos que para una función simple $\varphi \in L^+$,

$$\sup \left\{ \int \psi d\mu : \psi \textit{ es simple y } 0 \leq \psi \leq \varphi \right\} \leq \int \varphi d\mu.$$

La otra desigualdad se sigue de que $\int \varphi d\mu \in \left\{ \int \psi d\mu : \psi \textit{ es simple y } 0 \leq \psi \leq \varphi \right\}$ estableciendo la extensión.

Para ejemplificar la conveniencia de definir la integral de esta manera consideremos al espacio de medida $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$. Un resultado básico de cálculo es que la función $f = \mathbb{1}_{\mathbb{I}}$, donde $\mathbb{I}$ denota a los irracionales, no es Riemann integrable en ningún intervalo. Sin embargo,

$$\int_{(0,1)} f d\lambda = \int \mathbb{1}_{(0,1)} f d\lambda = \lambda\big((0,1) \cap \mathbb{I}\big) = \lambda\big((0,1)\big) - \lambda\big((0,1) \cap \mathbb{Q}\big) = 1.$$

Ahora veremos que la integral, definida para funciones no negativas, respeta las propiedades que esperaríamos respetase una integral.

Proposición 2.24. *Sean $f, g \in L^+$.*

1) Si $f \leq g$ entonces $\int f d\mu \leq \int g d\mu$.

2) Si $E, F \in \Sigma$ y $E \subset F$ entonces $\int_E f d\mu \leq \int_F f d\mu$.

Demostración. 1) Si $\varphi \in L^+$ es una función simple tal que $\varphi \leq f$, entonces $\varphi \leq g$. Por lo tanto

$$\left\{ \int \varphi d\mu : \varphi \textit{ es simple y } 0 \leq \varphi \leq f \right\} \subset \left\{ \int \varphi d\mu : \varphi \textit{ es simple y } 0 \leq \varphi \leq g \right\},$$

de donde se concluye 1).

2) Notando que $f \mathbb{1}_E \leq f \mathbb{1}_F$ basta usar el inciso 1). □

El siguiente resultado fundamental que relaciona a las integrales y la convergencia de funciones se debe a Beppo Levi y Henri Lebesgue.

Teorema 2.25 (De convergencia monótona)**.** *Sea (X, Σ, μ) un espacio de medida y $\{f_n\}_{n \in \mathbb{N}} \subset L^+(X, \Sigma)$ una sucesión de funciones tal que $f_n \leq f_{n+1}$ y $f = \lim_{n \to \infty} f_n$ entonces*

$$\lim_{n \to \infty} \int f_n d\mu = \int f d\mu.$$

Demostración. Vemos que $f_n \leq f$ y por monotonía de la integral se tendrá que $\int f_n d\mu \leq \int f d\mu$ para toda $n \in \mathbb{N}$. Además $\int f_n d\mu \leq \int f_{n+1} d\mu$, por lo que $\lim_{n\to\infty} \int f_n d\mu$ existe y es menor o igual a $\int f d\mu$. Para establecer la desigualdad en el sentido contrario consideremos $0 < \epsilon < 1$ y $0 \leq \varphi \leq f$ simple arbitrarios. Definiendo

$$E_n := \{x \in X : f_n(x) \geq \epsilon\varphi(x)\} = (f_n - \epsilon\varphi)^{-1}\big([0,\infty]\big) \in \Sigma,$$

notamos que $E_n \subset E_{n+1}$ y $X = \bigcup_{n\in\mathbb{N}} E_n$ pues $f_n \to f$ y $\epsilon\varphi < f$. Además $f_n \geq f_n \mathbb{1}_{E_n} \geq \epsilon\varphi\mathbb{1}_{E_n}$. Por la monotonía de la integral tenemos que

$$\int f_n d\mu \geq \int f_n \mathbb{1}_{E_n} d\mu = \int_{E_n} f_n d\mu \geq \int_{E_n} \epsilon\varphi d\mu = \epsilon \int_{E_n} \varphi d\mu,$$

donde la última igualdad se da por el iniciso 1) de la proposición 2.22. Por el inciso 4) de la proposición 2.22 tenemos que $\nu_\varphi(E) = \int_E \varphi d\mu$ define una medida. Por lo tanto $\lim_{n\to\infty} \int_{E_n} \varphi d\mu = \int \varphi d\mu$ y se sigue que

$$\lim_{n\to\infty} \int f_n d\mu \geq \epsilon \int \varphi d\mu.$$

Tomando el supremo sobre todas las funciones simples tales que $0 \leq \varphi \leq f$ tenemos que

$$\lim_{n\to\infty} \int f_n d\mu \geq \epsilon \int f d\mu.$$

Como la desigualdad anterior se cumple para $0 < \epsilon < 1$ arbitrario entonces también se satisface para $\epsilon = 1$, concluyendo el resultado. □

En el teorema de convergencia monótona la hipótesis de que $f_n \leq f_{n+1}$ es esencial para que la conclusión se satisfaga. Si relajamos esta hipótesis y consideramos cualquier sucesión de funciones en $L^+(X,\Sigma)$ entonces un corolario del teorema de convergencia monótona nos permitirá obtener una conclusión más débil.

Corolario 2.26 (Lema de Fatou)**.** *Sea (X,Σ,μ) un espacio de medida y $\{f_n\}_{n\in\mathbb{N}} \subset L^+(X,\Sigma)$ arbitrario, entonces*

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Demostración. Definamos $g_n := \inf_{m\geq n} f_m$ para $n \in \mathbb{N}$. Entonces $g_n \leq g_{n+1}$ y $\lim_n g_n = \liminf_n f_n$ y por el teorema de convergencia monótona se tiene

$$\lim_{n\to\infty} \int g_n d\mu = \int \liminf_{n\to\infty} f_n d\mu.$$

Por otra parte tenemos que $g_n \leq f_m$ para $m \geq n$ y así $\int g_n d\mu \leq \int f_m d\mu$ para $m \geq n$, de donde $\int g_n d\mu \leq \inf_{m\geq n} \int f_m d\mu$. Tomando límites y usando la igualdad obtenida se tiene que

$$\int \liminf_{n\to\infty} f_n d\mu \leq \lim_{n\to\infty} \inf_{m\geq n} \int f_m d\mu = \liminf_{n\to\infty} \int f_n d\mu. \qquad \square$$

Otro corolario importante que podemos obtener del teorema de convergencia monótona es la linealidad de la integral para funciones no-negativas:

Proposición 2.27. *Sea* (X, Σ, μ) *un espacio de medida. Si* $f, g \in L^+(X, \Sigma)$ *y* $\alpha \geq 0$ *entonces*

1) $\int (f+g) d\mu = \int f d\mu + \int g d\mu$

2) $\int \alpha f d\mu = \alpha \int f d\mu$.

Demostración. Sean $\{\varphi_n\} \subset L^+(X, \Sigma)$ y $\{\psi_n\} \subset L^+(X, \Sigma)$ sucesiones de funciones simples tales que $\varphi_n \leq \varphi_{n+1}, \psi_n \leq \psi_{n+1}, \lim_{n\to\infty} \varphi_n = f$ y $\lim_{n\to\infty} \psi_n = g$. Entonces $\{\varphi_n + \psi_n\}_{n\in\mathbb{N}}$ y $\{\alpha\varphi_n\}_{n\in\mathbb{N}}$ son sucesiones de funciones simples no-negativas y crecientes tales que $\lim_{n\to\infty}(\varphi_n + \psi_n) = f + g$ y $\lim_{n\to\infty} \alpha\varphi_n = \alpha f$. Por el teorema de convergencia monótona y los incisos (1) y (2) de la proposición 2.22 por una parte tenemos que

$$\int (f+g) d\mu = \lim_{n\to\infty} \int (\varphi_n + \psi_n) d\mu = \lim_{n\to\infty} \left(\int \varphi_n d\mu + \int \psi_n d\mu \right) = \int f d\mu + \int g d\mu,$$

mientras que por otra parte,

$$\int \alpha f d\mu = \lim_{n\to\infty} \int \alpha\varphi_n d\mu = \lim_{n\to\infty} \alpha \int \varphi_n d\mu = \alpha \int f d\mu. \qquad \square$$

Corolario 2.28. *Sea* (X, Σ, μ) *un espacio de medida y* $\{f_n\}_{n\in\mathbb{N}} \subset L^+(X, \Sigma)$ *arbitrario, entonces si* $f = \sum_{n=1}^\infty f_n$*,*

$$\int f d\mu = \sum_{n=1}^\infty \int f_n d\mu.$$

Demostración. Para todo $k \in \mathbb{N}$, $\sum_{n=1}^k f_n$ define una sucesión monótona de funciones en L^+. Por convergencia monótona y linealidad de la integral,

$$\int f d\mu = \lim_{k\to\infty} \int \sum_{n=1}^k f_n d\mu = \lim_{k\to\infty} \sum_{n=1}^k \int f_n d\mu = \sum_{n=1}^\infty \int f_n d\mu. \qquad \square$$

Ahora extendemos el concepto de integral a cierta clase de funciones medibles $f : X \to \overline{\mathbb{R}}$. Recordando la descomposición de f en f^+ y f^-, si alguna de las cantidades $\int f^+ d\mu$ o $\int f^- d\mu$ es finita entonces definimos la *integral* de f con respecto a μ por

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu.$$

Si es el caso que $\int f d\mu \in \mathbb{R}$, equivalentemente $\int f^+ d\mu, \int f^- d\mu < \infty$ (ó $\int |f| d\mu < \infty$ en virtud de la proposición 2.27), entonces diremos que f es **integrable**.

Finalmente, si $f : X \to \mathbb{C}$ es medible entonces diremos que f es *integrable* si y sólo si $\Re f$ e $\Im f$ son integrables en el sentido real y, en este caso, definimos la integral de f con respecto a μ por

$$\int f d\mu = \int \Re f d\mu + i \int \Im f d\mu.$$

Dado que

$$|\Re f| \vee |\Im f| \leq |f| \leq |\Re f| + |\Im f|,$$

se sigue que f es integrable si y sólo si $\int |f|d\mu < \infty$. El espacio de funciones complejas, medibles e integrables es denotado por $L^1(X, \Sigma, \mu)$ ó simplemente L^1 cuando no haya posibilidad de confusiones. Nótese que L^1 es un espacio vectorial ya que para $f, g \in L^1$ y $\alpha, \beta \in \mathbb{C}$, $\alpha f + \beta g$ es medible y

$$\int |\alpha f + \beta g| d\mu \leq \int \big(|\alpha||f| + |\beta||g|\big) d\mu = |\alpha| \int |f| d\mu + |\beta| \int |g| d\mu < \infty.$$

Proposición 2.29. *Sea (X, Σ, μ) un espacio de medida. Si $f, g \in L^1(X, \Sigma, \mu)$ y $\alpha \in \mathbb{C}$ entonces*

1) $\int (f+g) d\mu = \int f d\mu + \int g d\mu$

2) $\int \alpha f d\mu = \alpha \int f d\mu$.

Demostración. 1) Dado que $\Re(f+g) = \Re f + \Re g$ y $\Im(f+g) = \Im f + \Im g$, basta considerar el caso cuando f y g son reales. Entonces $f^+, f^-, g^+, g^- \in L^+(X, \Sigma)$ todas tienen integral finita. Notemos que $f^+ - f^- + g^+ - g^- = f + g = (f+g)^+ - (f+g)^-$ o equivalentemente $f^+ + g^+ + (f+g)^- = f^- + g^- + (f+g)^+$. Por la proposición 2.27,

$$\int f^+ d\mu + \int g^+ d\mu + \int (f+g)^- d\mu = \int f^- d\mu + \int g^- d\mu + \int (f+g)^+ d\mu,$$

donde todos los elementos son finitos. Reacomodando términos obtenemos que

$$\begin{aligned}\int (f+g) d\mu &= \int (f+g)^+ d\mu - \int (f+g)^- d\mu \\ &= \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu = \int f d\mu + \int g d\mu.\end{aligned}$$

2) Para $\alpha \geq 0$ se sigue de la definición y la proposición 2.27. Si $f = u + iv$, entonces u y v son integrables. De esta manera si $\alpha \leq 0$, entonces $-\alpha \geq 0$ y

$$\begin{aligned}\int (\alpha u) d\mu &= \int (\alpha u)^+ d\mu - \int (\alpha u)^- d\mu \\ &= \int (-\alpha) u^- d\mu - \int (-\alpha) u^+ d\mu = -(-\alpha) \int u d\mu = \alpha \int u d\mu.\end{aligned}$$

La misma igualdad se cumple para αv y se sigue el resultado. Por otro lado si $\alpha = i$, utilizando lo anterior vemos que

$$\int i f d\mu = \int (-v + iu) d\mu = \int (-v) d\mu + i \int u d\mu = -\int v d\mu + i \int u d\mu = i \int f d\mu.$$

Finalmente, si $\alpha = a + ib$ para $a, b \in \mathbb{R}$, podemos utilizar lo anterior y el primer inciso para obtener

$$\begin{aligned}\int \alpha f d\mu &= \int (af + ibf) d\mu = \int a f d\mu + \int i b f d\mu \\ &= a \int f d\mu + ib \int f d\mu = \alpha \int f d\mu.\end{aligned}$$

□

Corolario 2.30. *Si $f \in L^1(X, \Sigma, \mu)$ entonces $\left|\int f d\mu\right| \leq \int |f| d\mu$.*

Demostración. Si f es una función real entonces

$$\left|\int f d\mu\right| \leq \left|\int f^{+} d\mu\right| + \left|\int f^{-} d\mu\right| \leq \int f^{+} d\mu + \int f^{-} d\mu = \int |f| d\mu,$$

pues $|f| = f^{+} + f^{-}$. En el caso que f sea compleja definamos $\theta = \arg \int f d\mu$ y $\alpha = e^{-i\theta}$. Tenemos que $\left|\int f d\mu\right| = \alpha \int f d\mu = \int \alpha f d\mu$. Como todas las cantidades son reales positivas y $\int \alpha f d\mu = \int \Re(\alpha f) d\mu + i \int \Im(\alpha f) d\mu$, entonces

$$\left|\int f d\mu\right| = \int \Re(\alpha f) d\mu \leq \int |\alpha f| d\mu = \int |f| d\mu. \qquad \square$$

Con el desarrollo que hemos hecho de la integral hasta ahora, podremos generalizar el resultado del inciso (4) de la proposición 2.22.

Proposición 2.31. *Sean (X, Σ, μ) un espacio de medida y $f \in L^{+}(X, \Sigma)$. Entonces la función $\nu : \Sigma \to [0, \infty]$ dada por $\nu(A) := \int_A f d\mu$ es una medida. Además si $g : X \to \mathbb{C}$ es medible, $g \in L^1(X, \Sigma, \nu)$ si y sólo si $gf \in L^1(X, \Sigma, \mu)$ y en este caso*

$$\int g d\nu = \int g f d\mu.$$

La igualdad anterior también es válida para $g \in L^{+}$.

Demostración. Para ver que ν es efectivamente una medida notemos que basta probar la σ-aditividad pues es claro que $\nu(\emptyset) = 0$ y $\nu(A) \geq 0$. Así, sea $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ una sucesión de conjuntos disjuntos y definamos $A = \bigcup_{n\in\mathbb{N}} A_n$, entonces por el corolario 2.28 tendremos

$$\begin{aligned}\nu(A) &= \int_A f d\mu = \int f \mathbb{1}_A d\mu = \int f \left(\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}\right) d\mu = \int \sum_{n\in\mathbb{N}} f \mathbb{1}_{A_n} d\mu \\ &= \sum_{n\in\mathbb{N}} \int f \mathbb{1}_{A_n} d\mu = \sum_{n\in\mathbb{N}} \int_{A_n} f d\mu = \sum_{n\in\mathbb{N}} \nu(A_n),\end{aligned}$$

estableciendo la primera parte de la proposición.

Para la segunda parte notemos que si $g = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} \in L^{+}(X, \Sigma)$ es simple, entonces

$$\int g d\nu = \sum_{i=1}^{n} \alpha_i \nu(A_i) = \sum_{i=1}^{n} \alpha_i \int_{A_i} f d\mu = \sum_{i=1}^{n} \alpha_i \int f \mathbb{1}_{A_i} d\mu = \int \left(\sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}\right) f d\mu.$$

Para $g \in L^{+}(X, \Sigma)$ podemos tomar una sucesión no-decreciente de funciones simples $\{g_n\}_{n\in\mathbb{N}}$ que converjan a g y por el teorema de convergencia monótona,

$$\int g d\nu = \lim_{n\to\infty} \int g_n d\nu = \lim_{n\to\infty} \int g_n f d\mu = \int g f d\mu.$$

En general, ya sea que $g \in L^1(X, \Sigma, \nu)$ ó $gf \in L^1(X, \Sigma, \mu)$ descomponiendo a g en $\Re g^{\pm}$ y $\Im g^{\pm}$ y a gf en $\Re(gf)^{\pm} = \Re(g)^{\pm} f$ y $\Im(gf)^{\pm} = \Im(g)^{\pm} f$, de la igualdad anterior podemos concluir la segunda parte del enunciado. $\square$

Ahora veremos un resultado que nos permitirá relajar condiciones en los teoremas de convergencia, pidiendo que alguna propiedad se cumpla casi dondequiera en lugar de para todo el espacio.

Proposición 2.32. *Sean (X, Σ, μ) un espacio de medida y $f \in L^+(X, \Sigma)$, entonces $\int f d\mu = 0$ si y sólo si $f = 0$ c.d.s.*

Demostración. Definamos $E_n = \{x \in X : f(x) \geq n^{-1}\}$ y $E := \{x \in X : f(x) > 0\}$. Notemos que $E_n \subset E_{n+1}$ y $E = \bigcup_{n \in \mathbb{N}} E_n$.
Si $\int f d\mu = 0$, dado que $n^{-1} \mathbb{1}_{E_n} \leq f \mathbb{1}_{E_n} \leq f$ para toda $n \in \mathbb{N}$, por monotonía de la integral

$$n^{-1} \mu(E_n) \leq \int f \mathbb{1}_{E_n} d\mu \leq \int f d\mu = 0.$$

Por lo tanto $\mu(E_n) = 0$ para toda $n \in \mathbb{N}$ y se sigue que $\mu(E) = 0$; es decir $f = 0$ c.d.s.

Para una función simple $\varphi = \sum_{j=1}^{n} \alpha_j \mathbb{1}_{E_j}$, $\varphi = 0$ c.d.s. si y sólo si $\alpha_j = 0$ o $\mu(E_j) = 0$ para $j \in \{1, \ldots, n\}$. En este caso está claro que $\int \varphi d\mu = 0$. Si $f = 0$ c.d.s para cualquier función simple $\varphi \in L^+$ tal que $0 \leq \varphi \leq f$ se sigue que $\varphi = 0$ c.d.s. Dado que la integral de cualquier función simple menor que f es cero, tomando el supremo de estos valores vemos que $\int f d\mu = 0$. □

Corolario 2.33. *Sea (X, Σ, μ) un espacio de medida.*

1) Si $f \leq g$ c.d.s. con $f, g \in L^+$ entonces $\int f d\mu \leq \int g d\mu$.

2) Si $f \leq g$ c.d.s. con f, g integrables entonces $\int f d\mu \leq \int g d\mu$.

3) Si $f = g$ c.d.s. con $f, g \in L^+$ entonces $\int f d\mu = \int g d\mu$.

4) Si $f = g$ c.d.s. con $f, g \in L^1$ entonces $\int f d\mu = \int g d\mu$.

5) Si $\{f_n\}_{n \in \mathbb{N}} \subset L^+$ con $f_n \leq f_{n+1}$ c.d.s. y f_n converge a $f \in L^+$ c.d.s entonces $\int f d\mu = \lim_{n \to \infty} \int f_n d\mu$.

6) Si $\{f_n\}_{n \in \mathbb{N}} \subset L^+$ y f_n converge a $f \in L^+$ c.d.s. entonces $\int f d\mu \leq \liminf_n \int f_n d\mu$.

Demostración. 1) Sea $N \in \Sigma$ un conjunto nulo tal que $f \leq g$ en N^c. Notemos que $f \mathbb{1}_N = g \mathbb{1}_N = 0$ c.d.s. y $f \mathbb{1}_{N^c} \leq g \mathbb{1}_{N^c}$. Por la proposición 2.27 se cumple

$$\int f d\mu = \int_N f d\mu + \int_{N^c} f d\mu \leq \int_{N^c} g d\mu = \int_N g d\mu + \int_{N^c} g d\mu = \int g d\mu.$$

2) La hipótesis implica que $f^+ \leq g^+$ c.d.s. y $g^- \leq f^-$ c.d. Usando el inciso anterior tendremos que $\int f^+ d\mu \leq \int g^+ d\mu$ y $\int g^- d\mu \leq \int f^- d\mu$. Dado que estas cantidades son finitas podemos llegar al resultado.

3) $f = g$ c.d.s si y sólo si $f \leq g$ c.d.s. y $g \leq f$ c.d.s. Basta usar el inciso (1) dos veces.

4) Basta usar el inciso (2).

5) Sea E_n un conjunto nulo tal que $f_n \leq f_{n+1}$ en E_n^c y F otro conjunto nulo tal que $\{f_n\}_{n \in \mathbb{N}}$ converge a f en F^c. Es claro que $N = F \cup \bigcup_{n \in \mathbb{N}} E_n$ es nulo. Definimos $g_n := f_n \mathbb{1}_{N^c}$ y $g := f \mathbb{1}_{N^c}$. Estas funciones cumplen las hipótesis del teorema de convergencia monótona además de que $g_n = f_n$ c.d.s. y $f = g$ c.d.s. Por lo tanto

$$\int f d\mu = \int g d\mu = \lim_{n \to \infty} \int g_n d\mu = \lim_{n \to \infty} \int f_n d\mu.$$

6) Sea N el conjunto nulo en el cual $\{f_n\}_{n\in\mathbb{N}}$ no converge a f. Definiendo $g_n := f_n\mathbb{1}_{N^c}$ y $g := f\mathbb{1}_{N^c}$ tendremos que $\lim_{n\to\infty} g_n = \liminf_n g_n = g$. Por el lema de Fatou y el inciso (3) se sigue que

$$\int f d\mu = \int g d\mu \leq \liminf_{n\to\infty} \int g_n d\mu = \liminf_{n\to\infty} \int f_n d\mu. \qquad \square$$

El siguiente teorema también resulta ser de vital importancia pues nos permite intercambiar el límite y la integral para funciones que ya no sean necesariamente no-negativas y se debe a Henri Lebesgue.

Teorema 2.34 (De convergencia dominada de Lebesgue)**.** *Sean (X,Σ,μ) un espacio de medida y $\{f_n\}_{n\in\mathbb{N}} \subset L^1$ tal que converja c.d.s. a f medible. Si existe $g \in L^1$ no-negativa tal que $|f_n| \leq g$ c.d.s. para toda $n \in \mathbb{N}$. Entonces $f \in L^1$ y*

$$\lim_{n\to\infty} \int |f - f_n| d\mu = 0 \quad y \quad \lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

Demostración. Podemos encontrar un conjunto nulo $N \in \Sigma$ tal que en N^c, $\{f_n\}_{n\in\mathbb{N}}$ converge a f y $|f_n| \leq g$ para todo $n \in \mathbb{N}$. Se sigue que en N^c, $|f| \leq g$; es decir, $|f| \leq g$ c.d.s. Por el inciso (1) del corolario 2.33 se sigue que $|f|$ es integrable y esto equivale a que $f \in L^1$.

Ahora, notemos que $|f - f_n| \leq 2g$ c.d.s. por lo que $0 \leq 2g - |f - f_n|$ c.d.s. y además converge a $2g$ c.d.s. Usando el inciso (6) del corolario 2.33 tendremos que

$$\int 2g d\mu \leq \liminf_{n\to\infty} \int (2g - |f - f_n|) d\mu = \int 2g d\mu + \liminf_{n\to\infty} \left(- \int |f - f_n| d\mu \right).$$

En consecuencia $\limsup_n \int |f - f_n| d\mu = -\liminf_n \left(- \int |f - f_n| d\mu \right) \leq 0$ y como $\int |f - f_n| d\mu \geq 0$ para todo $n \in \mathbb{N}$, se sigue que $\lim_{n\to\infty} \int |f - f_n| d\mu = 0$. Finalmente, notamos que

$$\left| \int f_n d\mu - \int f d\mu \right| = \left| \int (f_n - f) d\mu \right| \leq \int |f - f_n| d\mu,$$

de donde se sigue fácilmente que $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$. $\square$

Siempre hay que tener en cuenta la importancia de las hipótesis del teorema de convergencia dominada. Incluso si la sucesión de funciones fuera uniformemente acotada, sus integrales no necesariamente van a converger. Considérese el espacio de medida $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ y la sucesión dada por $f_n = \mathbb{1}_{(n,n+1]}$. Tenemos $\lim_{n\to\infty} f_n = 0$, sin embargo,

$$\lim_{n\to\infty} \int f_n d\lambda = \lim_{n\to\infty} \lambda\big((n, n+1]\big) = 1.$$

Claramente no existe ninguna función en $L^1(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ que acote a esta sucesión.

Corolario 2.35. *Sea $\{f_n\}_{n\in\mathbb{N}} \subset L^1$ tal que $\sum_{n=1}^{\infty} \int |f_n| d\mu < \infty$. Entonces $\sum_{n=1}^{\infty} f_n$ converge c.d.s. a una función $f \in L^1$ y*

$$\int f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Demostración. Por el corolario 2.28 $\int \big(\sum_{n=1}^{\infty} |f_n|\big) d\mu = \sum_{n=1}^{\infty} \int |f_n| d\mu < \infty$. Se sigue que el conjunto N, donde $\sum_{n=1}^{\infty} |f_n| = \infty$, es nulo. Por lo tanto para todo $x \in N^c$, $\sum_{n=1}^{\infty} f_n(x)$ existe. Si definimos $f = \sum_{n=1}^{\infty} f_n \mathbb{1}_{N^c}$, entonces, la sucesión dada por $g_n = \sum_{k=1}^{n} f_k$, converge c.d.s a f y $|g_n| \leq \sum_{n=1}^{\infty} |f_n| \in L^1$. Por el teorema de convergencia dominada se sigue que $f \in L^1$ y

$$\int f d\mu = \lim_{n\to\infty} \int g_n d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int f_k d\mu = \sum_{n=1}^{\infty} \int f_n d\mu. \qquad \square$$

La integral con respecto a la medida de Lebesgue es llamada la *integral de Lebesgue*. Hasta ahora no sabemos cual es la relación entre la integral de Riemann y la integral de Lebesgue. El siguiente resultado nos muestra, como se había anticipado, que la integración en medida es mucho más general:

Teorema 2.36. *Sea F una función no-decreciente y continua por la derecha. Si g es una función acotada en el intervalo $[a,b]$ y cuya integral de Riemann-Stieltjes con respecto F existe, entonces, g es $(\mathcal{M}_F, \mathcal{B}(\mathbb{R}))$-medible y*

$$\int_{(a,b]} g \, d\mu_F = \int_a^b g \, dF.$$

Demostración. Por hipótesis existe una sucesión anidada de particiones del intervalo $[a,b]$, $P_k = \{x_0^k, \ldots, x_{n_k}^k\}$, tales que las sucesiones dadas por

$$s_k = \sum_{i=1}^{n_k} m_i^k \big(F(x_i^k) - F(x_{i-1}^k)\big) \quad y \quad S_k = \sum_{i=1}^{n_k} M_i^k \big(F(x_i^k) - F(x_{i-1}^k)\big),$$

donde $m_i^k = \inf_{y \in [x_{i-1}^k, x_i^k]} g(y)$ y $M_i^k = \sup_{y \in [x_{i-1}^k, x_i^k]} g(y)$, convergen a $\int_a^b g \, dF$. Si definimos

$$h_k = \sum_{i=1}^{n_k} m_i^k \mathbb{1}_{(x_{i-1}^k, x_i^k]} \quad y \quad H_k = \sum_{i=1}^{n_k} M_i^k \mathbb{1}_{(x_{i-1}^k, x_i^k]},$$

entonces $s_k = \int_{(a,b]} h_k d\mu_F$ y $S_k = \int_{(a,b]} H_k d\mu_F$. Además conforman sucesiones monótonas por lo que existen $h = \lim_{k\to\infty} h_k$ y $H = \lim_{k\to\infty} H_k$. Por el teorema de convergencia dominada,

$$\int_{(a,b]} h \, d\mu_F = \lim_{k\to\infty} s_k = \int_a^b g \, dF = \lim_{k\to\infty} S_k = \int_{(a,b]} H d\mu_F.$$

Entonces tendremos que $H = h$ μ_F-c.d.s. Por otro lado se puede ver fácilmente que $h \leq g \leq H$, por lo que $H = g$ μ_F-c.d.s y como H es medible y μ_F es completa en $\mathcal{M}_F$ se sigue que g es medible. Entonces,

$$\int_{(a,b]} h \, d\mu_F \leq \int_{(a,b]} g \, d\mu_F \leq \int_{(a,b]} H d\mu_F,$$

de donde podemos concluir que $\int_{(a,b]} g \, d\mu_F = \int_a^b g \, dF$. $\square$

Como caso particular de este resultado tenemos que toda función acotada en el intervalo $[a,b]$ y Riemann integrable, f, es Lebesgue-medible y

$$\int_{[a,b]} f d\lambda = \int_a^b f dx.$$

Concluimos esta sección con un teorema importante que identificaremos como el *teorema de cambio de variable* en medida. Para esto necesitamos introducir a la *medida inducida por una función medible* también conocida como la *medida imagen* ó *medida pushforward*.

Proposición 2.37. *Sean (X, Σ_X) y (Y, Σ_Y) espacios medibles y $\varphi : X \to Y$ una función (Σ_X, Σ_Y)-medible. Si μ es una medida en Σ_X, entonces φ induce una medida en Σ_Y a través de la identidad*

$$\varphi_*\mu(E) \equiv \mu \circ \varphi^{-1}(E) := \mu\big(\varphi^{-1}(E)\big),$$

para todo $E \in \Sigma_Y$.

Demostración. Claramente $\varphi_*\mu(\emptyset) = \mu(\emptyset) = 0$. Ahora si $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma_Y$ son disjuntos, entonces $\{\varphi^{-1}(E_n)\}_{n\in\mathbb{N}} \subset \Sigma_X$ también son disjuntos y $\varphi^{-1}\big(\bigcup_{n\in\mathbb{N}} E_n\big) = \bigcup_{n\in\mathbb{N}} \varphi^{-1}(E_n)$. Por lo tanto,

$$\varphi_*\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} \varphi^{-1}(E_n)\right) = \sum_{n=1}^{\infty} \mu(\varphi^{-1}(E_n)) = \sum_{n=1}^{\infty} \varphi_*\mu(E_n). \qquad \square$$

Teorema 2.38. *(Cambio de variable) Sea (X, Σ_X, μ) un espacio de medida, (Y, Σ_Y) un espacio medible, $\varphi : X \to Y$ una función (Σ_X, Σ_Y)-medible y $f : Y \to \mathbb{R}$ medible. Sea $\varphi_*\mu$ la medida inducida por φ. Entonces $f \circ \varphi \in L^1(X, \Sigma_X, \mu)$ si y sólo si $f \in L^1(Y, \Sigma_Y, \varphi_*\mu)$ y en este caso*

$$\int f \circ \varphi d\mu = \int f d\varphi_*\mu.$$

La igualdad anterior también se cumple para $f \in L^+(X, \Sigma)$.

Demostración. Primero suponemos que f es simple y no-negativa; es decir, $f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$. Notando que para todo $A \in \Sigma_Y$, $\mathbb{1}_A \circ \varphi = \mathbb{1}_{\varphi^{-1}(A_i)}$, se sigue que

$$\int f \circ \varphi d\mu = \sum_{i=1}^{n} a_i \int \mathbb{1}_{\varphi^{-1}(A_i)} d\mu = \sum_{i=1}^{n} a_i \mu\big(\varphi^{-1}(A_i)\big) = \sum_{i=1}^{n} a_i \varphi_*\mu(A_i) = \int f d\varphi_*\mu.$$

Ahora si $f \geq 0$, sabemos que existe una sucesión de funciones simples, $\{f_n\}_{n\in\mathbb{N}}$, tales que $0 \leq f_n \leq f$ para todo $n \in \mathbb{N}$ y $\lim_{n\to\infty} f_n = f$. Por convergencia monótona y el caso para funciones simples, se tiene

$$\int f \circ \varphi d\mu = \lim_{n\to\infty} \int f_n \circ \varphi d\mu = \lim_{n\to\infty} \int f_n d\varphi_*\mu = \int f d\varphi_*\mu.$$

Si $f \circ \varphi \in L^1(X, \Sigma_X, \mu)$ ó $f \in L^1(Y, \Sigma_Y, \varphi_*\mu)$, descomponiendo a f en $\Re f^{\pm}$ y $\Im f^{\pm}$, podemos concluir el resultado de la igualdad anterior. $\square$

2.3. Medidas producto y teorema de Tonelli-Fubini

En el primer capítulo se habló de la σ-álgebra producto cuando (X, Σ_X) y (Y, Σ_Y) son espacios medibles. Si ahora (X, Σ_X, μ) y (Y, Σ_Y, ν) fuesen espacios de medida sería razonable construir una medida sobre $\Sigma_X \otimes \Sigma_Y$ que sea el producto de μ y ν en cierto sentido.

Para introducir este concepto primero definiremos los *rectángulos medibles* como los elementos de la forma $A \times B \subset X \times Y$ donde $A \in \Sigma_X$ y $B \in \Sigma_Y$. Por la proposición 1.6 la σ-álgebra generada por $\mathcal{C} := \{A \times B : A \in \Sigma_X, B \in \Sigma_Y\}$ resulta ser $\Sigma_X \otimes \Sigma_Y$. Además es fácil ver que $\mathcal{C}$ es una clase elemental y por tanto un π-sistema pues:

1) $\emptyset \in \mathcal{C}$

2) Si $A, E \in \Sigma_X$ y $B, F \in \Sigma_Y$, $(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F) \in \mathcal{C}$

3) Si $A \in \Sigma_X$ y $B \in \Sigma_Y$, $(A \times B)^c = (A^c \times B^c) \cup (A \times B^c) \cup (A^c \times B)$ y todas las componentes son disjuntas por pares y están en $\mathcal{C}$.

Ahora, podemos definir de manera natural $\rho : \mathcal{C} \to [0, \infty]$ a través de la identidad $\rho(A \times B) = \mu(A)\nu(B)$ para $A \times B \in \mathcal{C}$. En virtud del teorema 1.23, sujeto a que ρ sea σ-aditiva y σ-finita, podremos extenderla a una única medida en $\Sigma_X \otimes \Sigma_Y$.

Si μ y ν son σ-finitas, entonces existen sucesiones crecientes $\{X_n\}_{n\in\mathbb{N}} \subset \Sigma_X$ y $\{Y_n\}_{n\in\mathbb{N}} \subset \Sigma_Y$ tales que $X = \bigcup_{n\in\mathbb{N}} X_n$, $Y = \bigcup_{n\in\mathbb{N}} Y_n$ y para todo $n \in \mathbb{N}$, $\mu(X_n) < \infty$ y $\nu(Y_n) < \infty$. Entonces, $X \times Y = \bigcup_{n\in\mathbb{N}} X_n \times Y_n$ y $\rho(X_n \times Y_n) = \mu(X_n)\nu(Y_n) < \infty$; es decir, ρ es σ-finita.

Veamos que en general ρ es σ-aditiva. Tomemos un rectángulo $A \times B \in \mathcal{C}$ que sea la unión disjunta de rectángulos medibles $\{A_n \times B_n\}_{n\in\mathbb{N}} \subset \mathcal{C}$. Notemos que

$$\mathbb{1}_A(x)\mathbb{1}_B(y) = \mathbb{1}_{A\times B}(x, y) = \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n\times B_n}(x, y) = \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(x)\mathbb{1}_{B_n}(y).$$

Si fijamos la variable y, de ambos lados de la ecuación tenemos funciones Σ_X-medibles. Integrando con respecto a μ obtenemos

$$\mu(A)\mathbb{1}_B(y) = \int \mathbb{1}_A(x)\mathbb{1}_B(y)d\mu(x) = \int \Big(\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(x)\mathbb{1}_{B_n}(y)\Big)d\mu(x) = \sum_{n\in\mathbb{N}} \mu(A_n)\mathbb{1}_{B_n}(y),$$

donde ahora ambos lados definen funciones Σ_Y-medibles. Si integramos con respecto a ν se sigue que

$$\mu(A)\nu(B) = \sum_{n\in\mathbb{N}} \mu(A_n)\nu(B_n).$$

En conclusión, si μ y ν son σ-finitas, existe una única medida en $\Sigma_X \otimes \Sigma_Y$ (también σ-finita), que llamaremos la *medida producto* de μ y ν y denotaremos por $\mu \times \nu$, tal que para todo $A \in \Sigma_X$ y $B \in \Sigma_Y$,

$$\mu \times \nu(A \times B) = \mu(A)\nu(B).$$

Más aún, para todo $E \in \Sigma_X \otimes \Sigma_Y$,

$$\mu \times \nu(E) = \text{ínf}\left\{\sum_{n=1}^{\infty} \mu(A_n)\nu(B_n) : A_n \in \Sigma_X, B_n \in \Sigma_Y, E \subset \bigcup_{n\in\mathbb{N}}(A_n \times B_n)\right\}.$$

Ahora introducimos algunos conceptos nuevos con los que trabajaremos a lo largo de esta sección:

Definición 2.39. Sean $E \subset X \times Y$ y $f : X \times Y \to Z$ dados. Definimos la **x-sección** de E y la **y-sección** de E por

$$E_x = \{y \in Y : (x, y) \in E\} \quad \text{y} \quad E_y = \{x \in X : (x, y) \in E\}$$

respectivamente. De manera similar, definimos la **x-sección** de f, $f_x : Y \to Z$, y la **y-sección** de f, $f^y : X \to Z$, por

$$f_x(y) = f^y(x) = f(x, y).$$

En la siguiente proposición veremos que las secciones preservan medibilidad.

Proposición 2.40. *Sean* (X, Σ_X) *y* (Y, Σ_Y) *espacios medibles.*

1) Si $E \in \Sigma_X \otimes \Sigma_Y$ *entonces* $E_x \in \Sigma_Y$ *para todo* $x \in X$ *y* $E_y \in \Sigma_X$ *para todo* $y \in Y$.

2) Si f *es una función* $\Sigma_X \otimes \Sigma_Y$*-medible entonces* f_x *es* Σ_Y*-medible para toda* $x \in X$ *y* f^y *es* Σ_X*-medible para toda* $y \in Y$.

Demostración. 1) Definamos

$$\mathcal{K} = \{E \subset X \times Y : E_x \in \Sigma_Y \textit{ para toda } x \in X, E_y \in \Sigma_X \textit{ para toda } y \in Y\}$$

y notemos que $\mathcal{C} \subset \mathcal{K}$ ya que para todo rectángulo medible $A \times B \in \mathcal{C}$,

$$(A \times B)_x = \begin{cases} B & \text{si } x \in A, \\ \emptyset & \text{si } x \notin A, \end{cases} \quad \text{y} \quad (A \times B)_y = \begin{cases} A & \text{si } y \in B, \\ \emptyset & \text{si } y \notin B. \end{cases}$$

Si probamos que $\mathcal{K}$ es λ-sistema habremos acabado pues por el teorema de clases monótonas tendremos que $\Sigma_X \otimes \Sigma_Y = \sigma(\mathcal{C}) \subset \mathcal{K}$. Es claro que $X \times Y \in \mathcal{K}$. Por otra parte, si $E, F \in \mathcal{K}$ son tales que $E \subset F$, notemos que $(F - E)_x = F_x - E_x \in \Sigma_Y$ y $(F - E)_y = F_y - E_y \in \Sigma_X$, de donde $F - E \in \mathcal{K}$. Finalmente consideremos $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{K}$ tal que $E_n \subset E_{n+1}$ para toda n. Es fácil ver que $\left(\bigcup_{n \in \mathbb{N}} E_n\right)_x = \bigcup_{n \in \mathbb{N}} E_x \in \Sigma_Y$ y análogamente $\left(\bigcup_{n \in \mathbb{N}} E_n\right)_y = \bigcup_{n \in \mathbb{N}} E_y \in \Sigma_X$, concluyendo que $\mathcal{K}$ es un λ-sistema.

2) Notemos que

$$\begin{aligned} (f_x)^{-1}(B) &= \{y \in Y : f_x(y) \in B\} \\ &= \{y \in Y : f(x, y) \in B\} = \{y \in Y : (x, y) \in f^{-1}(B)\} = (f^{-1}(B))_x. \end{aligned}$$

y análogamente $(f_y)^{-1}(B) = (f^{-1}(B))_y$. Se sigue el resultado por el inciso anterior. $\square$

Hasta este punto, todos los argumentos que se han expuesto en esta sección se pueden generalizar fácilmente al producto de los espacios $(X_1, \Sigma_1, \mu_1), \ldots, (X_n, \Sigma_n, \mu_n)$. En el caso general, si toda μ_i es σ-finita entonces la extensión que se obtiene, $\mu_1 \times \cdots \times \mu_n$, también será σ-finita y además será única como extensión. Esto nos permite obtener propiedades de asociatividad con repecto al producto de medidas. Asimismo, los resultados que aparecerán tienen su análogo para productos finitos. Por simplicidad trabajaremos en el caso $n = 2$.

El objetivo ahora será relacionar la integral en $X \times Y$ con respecto $\mu \times \nu$ y las integrales en X y Y con respecto μ y ν. Un primer paso para lograr esto está dado por la siguiente proposición.

Proposición 2.41. *Sean* (X, Σ_X, μ) *y* (Y, Σ_Y, ν) *espacios de medida* σ*-finitos. Sea* $E \in \Sigma_X \otimes \Sigma_Y$ *dado y definamos* $\varphi_E(x) = \nu(E_x)$ *y* $\psi_E(y) = \mu(E_y)$. *Entonces* φ_E *es* Σ_X*-medible,* ψ_E *es* Σ_Y*-medible y*

$$\mu \times \nu(E) = \int \varphi_E d\mu = \int \psi_E d\nu.$$

Demostración. Denotemos por $\mathcal{L}$ a la clase de conjuntos $E \in \Sigma_X \otimes \Sigma_Y$ para los cuales las conclusiones de la proposición son ciertas y sea $\mathcal{C} = \{A \times B : A \in \Sigma_X, B \in \Sigma_Y\}$. Notemos que si $E = A \times B \in \mathcal{C}$ entonces $\varphi_{A\times B} = \nu(B)\mathbb{1}_A$ y $\psi_{A\times B} = \mu(A)\mathbb{1}_B$, de donde es clara la medibilidad de las funciones y

$$\int \varphi_{A\times B} d\mu = \mu(A)\nu(B) = \int \psi_{A\times B} d\nu.$$

Por lo tanto $\mathcal{C} \subset \mathcal{L}$ y por ser π-sistema, basta probar que $\mathcal{L}$ es λ-sistema por el teorema de clases monótonas. La demostración se dividirá en el caso finito y σ-finito.

Primero supongamos que μ y ν son finitas. Como $\mathcal{C} \subset \mathcal{L}$ se sigue que $X \times Y \in \mathcal{L}$. Si $E, F \in \mathcal{L}$ son tales que $E \subset F$, como ν es finita

$$\varphi_{F-E}(x) = \nu((F-E)_x) = \nu(F_x - E_x) = \nu(F_x) - \nu(E_x) = \varphi_F(x) - \varphi_E(x).$$

Por lo tanto $\varphi_{F-E} = \varphi_F - \varphi_E$ es medibles ya que por hipótesis φ_F y φ_E lo son. Análogamente $\psi_{F-E} = \psi_F - \psi_E$ también es medible. Al ser μ y ν finitas, $(\mu\times\nu)(F-E) = (\mu\times\nu)(F) - (\mu\times\nu)(E)$ y

$$\begin{aligned}\int \varphi_{F-E} d\mu &= \int \varphi_F d\mu - \int \varphi_E d\mu \\ &= (\mu\times\nu)(F) - (\mu\times\nu)(E) = \int \psi_F d\nu - \int \psi_E d\nu = \int \psi_{F-E} d\nu,\end{aligned}$$

estableciendo que $F - E \in \mathcal{L}$. Falta ver que $\mathcal{L}$ es cerrada bajo la unión de sucesiones de conjuntos crecientes; para esto consideremos $\{E_n\}_{n\in\mathbb{N}} \subset \mathcal{L}$ tal que $E_n \subset E_{n+1}$ para todo $n \in \mathbb{N}$. Por propiedades de las medidas tendremos que $0 \leq \varphi_{E_n} \leq \varphi_{E_{n+1}}$, $0 \leq \psi_{E_n} \leq \psi_{E_{n+1}}$, $\varphi_{E_n} \to \varphi_E$ y $\psi_{E_n} \to \psi_E$, donde $E := \bigcup_{n\in\mathbb{N}} E_n$. También, $(\mu\times\nu)(E) = \lim_{n\to\infty}(\mu\times\nu)(E_n)$. Por el teorema de convergencia monótona se sigue que

$$\int \varphi_E d\mu = \lim_{n\to\infty} \int \varphi_{E_n} d\mu = \lim_{n\to\infty} (\mu\times\nu)(E_n) = \lim_{n\to\infty} \int \psi_{E_n} d\nu = \int \psi_E d\nu,$$

estableciendo que $E \in \mathcal{L}$. Así $\mathcal{L}$ es un λ-sistema y, en este caso, $\mathcal{L} = \Sigma_X \otimes \Sigma_Y$.

En el caso que μ y ν sean σ-finitas, existen sucesiones de rectángulos medibles $E_n := X_n \times Y_n \in \mathcal{C}$ para cada $n \in \mathbb{N}$, tales que $E_n \subset E_{n+1}$, $(\mu\times\nu)(E_n) < \infty$ y $X \times Y = \bigcup_{n\in\mathbb{N}} E_n$. Luego, si $F \in \Sigma_X \otimes \Sigma_Y$ es claro que

$$(F \cap E_n)_x = \begin{cases} F_x \cap Y_n & \text{si } x \in X_n, \\ \emptyset & \text{si } x \notin X_n, \end{cases} \quad \text{y} \quad (F \cap E_n)_y = \begin{cases} F_y \cap X_n & \text{si } y \in Y_n, \\ \emptyset & \text{si } y \notin Y_n, \end{cases}$$

de donde $\varphi_{F\cap E_n}(x) = \mathbb{1}_{X_n}(x)\nu(F_x \cap Y_n)$ y $\psi_{F\cap E_n}(y) = \mu(F_y \cap X_n)\mathbb{1}_{Y_n}(y)$. Sean μ_{X_n} y ν_{Y_n} las restricciones de ν y μ a $\Sigma_X \cap X_n$ y $\Sigma_Y \cap Y_n$, respectivamente. Notando que $(\Sigma_X \otimes \Sigma_Y) \cap E_n = (\Sigma_X \cap X_n) \otimes (\Sigma_Y \cap Y_n)$ y que $(\mu\times\nu)_{E_n} = \mu_{X_n} \times \nu_{Y_n}$, del caso finito se sigue que

$$\begin{aligned}\int \mathbb{1}_{X_n}(x)\nu(F_x \cap Y_n) d\mu(x) &= \int \nu_{Y_n}\big((F\cap E_n)_x\big) d\mu_{X_n}(x) \\ &= (\mu_{X_n} \times \nu_{Y_n})(F \cap E_n) = (\mu\times\nu)(F\cap E_n).\end{aligned}$$

Análogamente $(\mu\times\nu)(F\cap E_n) = \int \mu(F_y \cap X_n)\mathbb{1}_{Y_n}(y) d\nu(y)$ y podemos concluir al tomar el límite cuando $n \to \infty$ y aplicar el teorema de convergencia monótona. □

Finalmente introducimos uno de los teoremas clave en la teoría de la medida. En realidad es la combinación de dos teoremas, uno introducido por Fubini en 1907 y otro introducido por Tonelli en 1909. El teorema de Fubini nos permite reducir la integral con respecto a la medida producto a dos integrales iteradas; esto para cualquier función integrable y medible con respecto a la σ-álgebra producto. Por otra parte, el teorema de Tonelli es una variación para funciones no-negativas.

Teorema 2.42 (Tonelli–Fubini)**.** *Sean* (X, Σ_X, μ) *y* (Y, Σ_Y, ν) *espacios de medida* σ*-finitos.*

1) (Tonelli) Si $f \in L^+(X \times Y, \Sigma_X \otimes \Sigma_Y)$ *entonces las funciones* $\varphi(x) = \int f_x d\nu$ *y* $\psi(y) = \int f_y d\mu$ *pertenecen a* $L^+(X, \Sigma_X)$ *y* $L^+(Y, \Sigma_Y)$*, respectivamente, y además*

$$\int f d(\mu \times \nu) = \int \varphi d\mu = \int \psi d\nu. \tag{2.1}$$

2) (Fubini) Si $f \in L^1(X \times Y, \Sigma_X \otimes \Sigma_Y, \mu \times \nu)$ *entonces* $f_x \in L^1(Y, \Sigma_Y, \nu)$ *para casi toda* $x \in X$ *y* $f_y \in L^1(X, \Sigma_X, \mu)$ *para casi toda* $y \in Y$*. Entonces las funciones* $\varphi(x) = \int f_x d\nu$ *y* $\psi(y) = \int f_y d\mu$ *definidas c.d.s. están en* $L^1(X, \Sigma_X, \mu)$ *y* $L^1(Y, \Sigma_Y, \nu)$ *respectivamente y (2.1) se cumple.*

Lo que el teorema anterior plantea es que bajo ciertas condiciones podremos integrar de forma iterada. Otras formas de escribir (2.1) equivalentes son

$$\int f d(\mu \times \nu) = \int \left(\int f_x(y) d\nu(y) \right) d\mu(x) = \int \left(\int f_y(x) d\mu(x) \right) d\nu(y)$$

además de

$$\int f d(\mu \times \nu) = \int \left(\int f(x, y) d\nu(y) \right) d\mu(x) = \int \left(\int f(x, y) d\mu(x) \right) d\nu(y).$$

Demostración. 1) Por la proposición anterior, el resultado se sigue para funciones indicadoras $\mathbb{1}_E$ con $E \in \Sigma_X \otimes \Sigma_Y$. Por linealidad de la integral y de las secciones es claro que el resultado es válido si $f \in L^+(X \times Y, \Sigma_X \otimes \Sigma_Y)$ es $\Sigma_X \otimes \Sigma_Y$-simple. En general, si $f \in L^+(X \times Y, \Sigma_X \otimes \Sigma_Y)$, podemos construir una sucesión creciente de funciones simples $\{g_n\}_{n \in \mathbb{N}} \subset L^+(X \times Y, \Sigma_X \otimes \Sigma_Y)$ tal que $g_n \to f$. Por la validez del resultado para funciones simples positivas y convergencia monótona

$$\varphi(x) = \int f_x d\nu = \lim_{n \to \infty} \int g_n(x, y) d\nu(y),$$

estableciendo la medibilidad de φ. También por convergencia monótona,

$$\begin{aligned} \int f d(\mu \times \nu) &= \lim_{n \to \infty} \int g_n d(\mu \times \nu) = \lim_{n \to \infty} \int \left(\int g_n(x, y) d\nu(y) \right) d\mu(x) \\ &= \int \left(\lim_{n \to \infty} \int g_n(x, y) d\nu(y) \right) d\mu(x) = \int \left(\int f_x(y) d\nu(y) \right) d\mu(x). \end{aligned}$$

Análogamente obtenemos que ψ es medible y

$$\int f d(\mu \times \nu) = \int \left(\int f_y(x) d\mu(x) \right) d\nu(y),$$

probando la parte 1) del teorema.

2) Por definición de L^1, basta probar el teorema para funciones reales, pues el caso complejo se sigue inmediatamente. Si $f \in L^1(X \times Y, \Sigma_X \otimes \Sigma_Y, \mu \times \nu)$ es real, consideremos f^+ y f^-. Dado que $f^+, f^- \leq |f|$, se sigue que $f^+, f^- \in L^1(X \times Y, \Sigma_X \otimes \Sigma_Y, \mu \times \nu) \cap L^+(X \times Y, \Sigma_X \otimes \Sigma_Y)$. Por el inciso anterior obtenemos lo siguiente:

$$\int \left(\int f^+(x,y) d\nu(y) \right) d\mu(x) = \int \left(\int f^+(x,y) d\mu(x) \right) d\nu(y) = \int f^+ d(\mu \times \nu) < \infty,$$
$$\int \left(\int f^-(x,y) d\nu(y) \right) d\mu(x) = \int \left(\int f^-(x,y) d\mu(x) \right) d\nu(y) = \int f^- d(\mu \times \nu) < \infty.$$

Así obtenemos que $\int f_x^+ d\nu, \int f_x^- d\nu \in L^1(X, \Sigma_X, \mu)$ y $\int f_y^+ d\mu, \int f_y^- d\mu \in L^1(Y, \Sigma_Y, \nu)$. Esto último implica que $f_x^+, f_x^- \in L^1(Y, \Sigma_Y, \nu)$ μ-c.d.s y $f_y^+, f_y^- \in L^1(X, \Sigma_X, \mu)$ ν-c.d.s. Notando que $f_x = f_x^+ - f_x^-$ y $f_y = f_y^+ - f_y^-$ se siguen todas las conclusiones del teorema. $\square$

De ahora en adelante se omitirán los paréntesis en las integrales iteradas, conveniendo que primero se realiza la integral interior.

A continuación veremos que las distintas hipótesis del teorema de Tonelli–Fubini son importantes y no pueden ser dispensados, incluso en casos que parecerían ser triviales.

1) Consideremos $X = Y = [0,1]$, $\Sigma_X = \Sigma_Y = \mathcal{B}([0,1])$, μ la medida de Lebesgue y ν la medida de conteo. En este caso (Y, Σ_Y, ν) no es un espacio σ-finito. Sea $D = \{(x,x) : x \in [0,1]\}$ la diagonal, que es medible por lo que $\mathbb{1}_D \in L^+(X \times Y, \Sigma_X \otimes \Sigma_Y)$. Es fácil ver que $\int \mathbb{1}_D(x,y) d(\mu \times \nu)(x,y) = \infty$, $\int \int \mathbb{1}_D(x,y) d\mu(x) d\nu(y) = 0$ y $\int \int \mathbb{1}_D(x,y) d\nu(y) d\mu(x) = 1$, por lo que las tres integrales existen pero son distintas.

2) Ahora consideremos $X, Y = [0,1]$, $\Sigma_X = \Sigma_Y = \mathcal{B}([0,1])$ y tanto μ como ν la medida de Lebesgue. Definamos $\{\delta_n\}_{n \in \mathbb{N}} \subset [0,1)$ tales que $\delta_1 = 0$, $\delta_n < \delta_{n+1}$ y $\lim_{n \to \infty} \delta_n = 1$. Podemos construir una sucesión de funciones continuas $\{g_n\}_{n \in \mathbb{N}}$ tales que $g_n \geq 0$ tenga soporte en (δ_n, δ_{n+1}), esto es que $g_n(x) = 0$ para toda $x \notin (\delta_n, \delta_{n+1})$, y $\int g_n(x) dx = 1$ para toda n. Al definir la función medible

$$f(x,y) = \sum_{n=1}^{\infty} [g_n(x) - g_{n+1}(x)] g_n(y),$$

se puede comprobar fácilmente que

$$|f(x,y)| = \sum_{n=1}^{\infty} [g_n(x) + g_{n+1}(x)] g_n(y),$$

de donde es inmediato que $\int \int |f(x,y)| dy dx = \infty$, por lo que f no es integrable y además

$$\int \int f(x,y) dx dy = 0 \neq 1 = \int \int f(x,y) dy dx.$$

3) Consideremos ω_1, el conjunto de ordinales contables con su orden usual. Entonces ω_1 no es contable y para todo $x \in \omega_1$, el conjunto $\{y \in \omega_1 : y < x\}$ es contable. Equipamos a ω_1 con la σ-álgebra de conjuntos contables o cocontables, $\Sigma = \{E \subset \omega_1 : E \textit{ es contable o } E^c \textit{ es contable}\}$,

y la medida $\mu : \Sigma \to [0,\infty]$ tal que $\mu(E) = 0$ si E es contable y $\mu(E) = 1$ si E no es contable. Sea $Q = \{(x,y) \in \omega_1 \times \omega_1 : y < x\}$; claramente para todo $x, y \in \omega_1$, Q_x y Q_y^c son contables y por lo tanto son medibles. Además tendremos que las funciones $f(x) = \int \mathbb{1}_Q(x,y)d\mu(y) = 0$ y $g(y) = \int \mathbb{1}_Q(x,y)d\mu(x) = 1$ son Σ-medibles. Sin embargo $\int f d\mu = 0 \neq 1 = \int g d\mu$. Podemos concluir que $Q \notin \Sigma \otimes \Sigma$.

2.4. Medida e Integral de Lebesgue en $\mathbb{R}^n$

Definimos a la medida de Lebesgue en $\mathbb{R}^n$ como la completación de la única medida, μ, en $\mathcal{B}(\mathbb{R}^n)$ para la cual

$$\mu\left(\prod_{i=1}^{n}(a_i, b_i]\right) = \prod_{i=1}^{n}(b_i - a_i)$$

para todo rectángulo semiabierto $\prod_{i=1}^{n}(a_i, b_i]$. La medida de Lebesgue en $\mathbb{R}^n$ será denotada por λ^n (ó simplemente λ si no se genera alguna confusión) y podemos llegar a ella al completar la medida producto $\prod_{i=1}^{n}\lambda$ (el n-producto de la medida Lebesgue en $\mathcal{B}(\mathbb{R})$) definida en $\bigotimes_{i=1}^{n}\mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^n)$. Naturalmente, denotaremos por $\mathcal{L}(\mathbb{R}^n)$ a la σ-álgebra donde λ^n es completa. Cabe mencionar que la definición de λ^n como medida exterior se extiende a todo $\mathcal{L}(\mathbb{R}^n)$. Es decir, para todo $E \in \mathcal{L}(\mathbb{R}^n)$,

$$\lambda^n(E) = \text{ínf}\left\{\sum_{k=1}^{\infty}\left(\prod_{i=1}^{n}\lambda(A_k^i)\right) : A_k^i \in \mathcal{B}(\mathbb{R}),\ E \subset \bigcup_{n\in\mathbb{N}}\left(\prod_{i=1}^{n}A_k^i\right)\right\}.$$

Otra observación importante es que para cualquier función Borel-medible positiva ó integrable, $\int f d\lambda^n = \int f d\big(\prod_{i=1}^{n}\lambda\big)$.
Un resultado que no probaremos ya que su demostración es análoga a la del caso unidimensional, es que si f es una función acotada y Riemann-integrable en $R = \prod_{i=1}^{n}[a_i, b_i]$, entonces f es Lebesgue medible y

$$\int_R f d\lambda^n = \int_R f dx.$$

También, como en el caso unidimensional, tendremos que la medida de Lebesgue en $\mathbb{R}^n$ es regular. Para probar regularidad necesitaremos el siguiente lema:

Lema 2.43. *Dados $A_1, \ldots A_n \in \mathcal{B}(\mathbb{R})$ y $\epsilon > 0$ existe un abierto $U \subset \mathbb{R}^n$ tal que $\prod_{i=1}^{n}A_i \subset U$ y*

$$\lambda^n(U) \leq \lambda^n\left(\prod_{i=1}^{n}A_i\right) + \epsilon$$

Demostración. Primero supongamos que $\lambda(A_i) < \infty$ para $1 \leq i \leq n$. La demostración se hará por inducción sobre n. Para el caso $n = 1$ es consecuencia del teorema 1.26. Supongamos que es valido para $n-1$ y probemos el resultado para n. Sean $\delta_1 = \text{máx}\{1, \lambda(A_n)\}$ y $\delta_2 = \text{máx}\left\{1, \prod_{i=1}^{n-1}\lambda(A_i)\right\}$. Luego tomemos $\epsilon_1 = \frac{1}{3}\text{mín}\{\epsilon/\delta_1, \sqrt{\epsilon}\}$ y $\epsilon_2 = \frac{1}{3}\text{mín}\{\epsilon/\delta_2, \sqrt{\epsilon}\}$. Por la hipótesis inductiva existe un abierto $U \subset \mathbb{R}^{n-1}$ tal que

$$\prod_{i=1}^{n-1}A_i \subset U \quad y \quad \lambda^{n-1}(U) \leq \lambda^{n-1}\left(\prod_{i=1}^{n-1}A_i\right) + \epsilon_1$$

(sin mayor perdida de generalidad podemos suponer que $U = \prod_{i=1}^{n-1} U_i$). Luego, sea $V \subset \mathbb{R}$ un abierto tal que $A_n \subset V$ y $\lambda(V) \leq \lambda(A_n) + \epsilon_2$. Entonces,

$$\begin{aligned}\lambda^n (U \times V) &= \lambda^{n-1}(U)\,\lambda(V) \\ &\leq \left(\lambda^{n-1}\left(\prod_{i=1}^{n-1} A_i\right) + \epsilon_1\right)(\lambda(A_n) + \epsilon_2) \\ &= \lambda^n\left(\prod_{i=1}^{n} A_i\right) + \epsilon_2\lambda^{n-1}\left(\prod_{i=1}^{n-1} A_i\right) + \epsilon_1\lambda(A_n) + \epsilon_1\epsilon_2 \leq \lambda^n\left(\prod_{i=1}^{n} A_i\right) + \epsilon.\end{aligned}$$

Esto prueba el enunciado cuando $\lambda(A_i) < \infty$ para $1 \leq i \leq n$. Para el caso general tomemos, $\{R_k\}_{k\in\mathbb{N}}$, una enumeración de los rectángulos de la forma $\prod_{i=1}^n (m_i, m_i + 1]$ para $m_1, \ldots, m_n \in \mathbb{Z}$. Entonces para cada $k \in \mathbb{N}$, $(\prod_{i=1}^n A_i) \cap R_k$ cumple las hipótesis del caso que ya probamos. Sea $U_k \subset \mathbb{R}^n$ un abierto tal que $(\prod_{i=1}^n A_i) \cap R_k \subset U_k$ y $\lambda^n(U_k) \leq \lambda^n((\prod_{i=1}^n A_i) \cap R_k) + \epsilon/2^k$. Se sigue que

$$\lambda^n\left(\bigcup_{k\in\mathbb{N}} U_k\right) \leq \sum_{k=1}^{\infty} \lambda^n(U_k) \leq \sum_{k=1}^{\infty} \lambda^n\left(\left(\prod_{i=1}^{n} A_i\right) \cap R_k\right) + \epsilon/2^k = \lambda^n\left(\prod_{i=1}^{n} A_i\right) + \epsilon. \qquad \square$$

Teorema 2.44. *Para todo $E \in \mathcal{L}(\mathbb{R}^n)$,*

$$\begin{aligned}\lambda^n(E) &= \inf\left\{\lambda^n\big(U\big) : E \subset U \text{ y } U \textit{ es abierto}\right\} \\ &= \sup\left\{\lambda^n\big(K\big) : K \subset E \text{ y } K \textit{ es compacto}\right\}.\end{aligned}$$

Demostración. Probemos la primera igualdad. Dado $\epsilon > 0$, existen conjuntos $A_i^k \in \mathcal{B}(\mathbb{R})$ para $k \in \mathbb{N}$ y $1 \leq i \leq n$ tales que

$$E \subset \bigcup_{k\in\mathbb{N}} \prod_{i=1}^{n} A_i^k \quad y \quad \sum_{k=1}^{\infty} \lambda^n\left(\prod_{i=1}^{n} A_i^k\right) \leq \lambda^n(E) + \epsilon.$$

Luego, por el lema 2.43, para cada $k \in \mathbb{N}$ existe un abierto $U_k \subset \mathbb{R}^n$ tal que $\prod_{i=1}^n A_i^k \subset U_k$ y $\lambda^n(U_k) \leq \lambda^n\left(\prod_{i=1}^n A_i^k\right) + \epsilon/2^k$. Se sigue que

$$\lambda^n\left(\bigcup_{k\in\mathbb{N}} U_k\right) \leq \sum_{k=1}^{\infty} \lambda^n(U_k) \leq \sum_{k=1}^{\infty} \lambda^n\left(\prod_{i=1}^{n} A_i^k\right) + \epsilon/2^k \leq \lambda^n(E) + 2\epsilon.$$

Como ϵ es arbitrario podemos concluir que $\lambda^n(E) \geq \inf\left\{\lambda^n\big(U\big) : E \subset U \text{ y } U \textit{ es abierto}\right\}$. La otra desigualdad es trivial y podemos concluir. Una vez que tenemos la primera igualdad, probar la segunda es análogo al caso unidimensional. $\square$

Las propiedades de escalamiento e invarianza ante traslación también se extienden fácilmente del caso unidimensional:

Proposición 2.45. *Para todo $a \in \mathbb{R}^n$ y $r \in \mathbb{R} - \{0\}$ definamos $\tau_a, T_r : \mathbb{R}^n \to \mathbb{R}^n$ por medio de $\tau_a(x) = x + a$ y $T_r(x) = rx$. Entonces,*

1) para todo $E \in \mathcal{L}(\mathbb{R}^n)$, $E + a, rE \in \mathcal{L}(\mathbb{R}^n)$ y

$$\lambda^n(E + a) = \lambda^n(E), \qquad \lambda^n(rE) = |r|^n\lambda^n(E).$$

2) τ_a *y* T_r *son Lebesgue-medibles. Además, si* $f \in L^1(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \lambda^n)$ *(ó* $L^+(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n))$ *), también lo están* $f \circ \tau_a$ *y* $f \circ T_r$ *y se tiene*

$$\int f(x+a)d\lambda^n(x) = \int f(x)d\lambda^n(x) \quad \text{y} \quad \int f(rx)d\lambda^n(x) = |r|^{-n} \int f(x)d\lambda^n(x).$$

Demostración. 1) Probaremos la parte del enunciado correspondiente a escalamientos (la parte correspondiente a traslaciones es análoga). Nótese que las transformaciones T_r son continuas y por ende Borel-medibles, además $T_r^{-1} = T_{r^{-1}}$. Así que si $E \in \mathcal{B}(\mathbb{R}^n)$,

$$rE = T_r(E) = (T_{r^{-1}})^{-1}(E) \in \mathcal{B}(\mathbb{R}^n).$$

Por otro lado, si $E \subset \bigcup_{k\in\mathbb{N}} \prod_{i=1}^n A_i^k$ para conjuntos $A_i^k \in \mathcal{B}(\mathbb{R})$, entonces $rE \subset \bigcup_{k\in\mathbb{N}} r\left(\prod_{i=1}^n A_i^k\right)$ $= \bigcup_{k\in\mathbb{N}} \prod_{i=1}^n rA_i^k$ y por la propiedad de escalamiento para el caso unidimensional tenemos que

$$\lambda^n(rE) \leq \sum_{k=1}^{\infty} \lambda^n \left(\prod_{i=1}^n rA_i \right) = \sum_{k=1}^{\infty} \prod_{i=1}^n \lambda^n (rA_i) = |r|^n \sum_{k=1}^{\infty} \prod_{i=1}^n \lambda^n (A_i).$$

En consecuencia $\lambda^n(rE) \leq |r|^n \lambda^n(E)$ y siguiendo el mismo razonamiento para r^{-1}, se tiene que $\lambda^n(E) = \lambda(r^{-1}rE) \leq |r^{-1}|^n \lambda^n(rE)$. Por lo tanto $\lambda^n(rE) = |r|^n \lambda^n(E)$. Ahora, si $B \subset A \in \mathcal{B}(\mathbb{R}^n)$ con A nulo, entonces $rB \subset rA$ y $\lambda^n(rA) = |r|^n \lambda(A) = 0$. En consecuencia, para todo $E \in \mathcal{B}(\mathbb{R}^n)$,

$$r(E \cup B) = rE \cup rB \in \mathcal{L}(\mathbb{R}^n) = \big\{E \cup B : E \in \mathcal{B}(\mathbb{R}^n),\ B \subset A \in \mathcal{N}\big\},$$

además de que $\lambda^n\big(r(E \cup B)\big) = \lambda^n(rE) = |r|^n \lambda^n(E) = |r|^n \lambda^n(E \cup B)$. Es decir, el resultado se extiende a todo $\mathcal{L}(\mathbb{R}^n)$.
2) Dado que $T_r^{-1} = T_{r^{-1}}$, del inciso 1) se sigue fácilmente que T_r es Lebesgue-medible y además para todo $E \in \mathcal{L}(\mathbb{R}^n)$,

$$T_{r*}\lambda^n(E) = \lambda^n\big(T_r^{-1}(E)\big) = \lambda^n\big(T_{r^{-1}}(E)\big) = \lambda^n\big(r^{-1}E\big) = |r|^{-n}\lambda^n(E).$$

Por el teorema 2.38, $f \circ T_r \in L^1(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \lambda^n)$ (ó $L^+(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n))$) y

$$\int f(rx)d\lambda^n(x) = \int f \circ T_r d\lambda^n = \int f dT_{r*}\lambda^n = |r|^{-n} \int f(x)d\lambda^n(x). \qquad \square$$

Las igualdades integrales en la proposición 2.45 son casos particulares del teorema de cambio de variable de cálculo integral. Nuestra siguiente tarea será demostrar la versión general del teorema de cambio de variable de cálculo y veremos, como en la demostración de la proposición 2.45, que no es más que un caso particular del teorema respectivo de medida.

A todo operador lineal $T : \mathbb{R}^n \to \mathbb{R}^n$ lo podemos identificar con la matriz de entradas $T_{i,j} = e_i \cdot Te_j$ y así definimos $\det(T) := \det\big((T_{i,j})\big)$. Consideremos el espacio de operadores lineales invertibles en $\mathbb{R}^n$ denotado por $GL(n, \mathbb{R})$. De cursos elementales de álgebra lineal sabemos que:

1) Para operadores lineales S y T, $\det(S \circ T) = \det(S)\det(T)$.

2) $T \in GL(n, \mathbb{R})$ si y sólo si $\det(T) \neq 0$.

3) Si $T \in GL(n, \mathbb{R})$ entonces $\det(T^{-1}) = \big(\det(T)\big)^{-1}$.

Además todo operador en $GL(n, \mathbb{R})$ es la composición de una cantidad finita de operadores del siguiente tipo:

I) $T_1(x_1, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, cx_j, \ldots, x_n)$ para $c \neq 0$ y $\det(T_1) = c$;

II) $T_2(x_1, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, x_j + cx_k, \ldots, x_n)$ para $c \neq 0$, $k \neq j$ y $\det(T_2) = 1$;

III) $T_3(x_1, \ldots, x_k, \ldots, x_j, \ldots, x_n) = (x_1, \ldots, x_j, \ldots, x_k, \ldots, x_n)$ para $k \neq j$ y $\det(T_3) = -1$.

Teorema 2.46. *Sea $T \in GL(n, \mathbb{R})$. Entonces,*

1) Si $E \in \mathcal{L}(\mathbb{R}^n)$, entonces $T(E) \in \mathcal{L}(\mathbb{R}^n)$ y $\lambda^n\big(T(E)\big) = |\det(T)|\lambda^n(E)$.

2) T es Lebesgue-medible y si $f \in L^1(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n), \lambda^n)$ (ó $L^+(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n))$), también $f \circ T$ y

$$\int f d\lambda^n = |\det(T)| \int f \circ T d\lambda^n.$$

Demostración. Probemos 2) cuando $f \geq 0$ y Borel-medible. En este caso $f \circ T$ también es Borel-medible. Si T es de la forma (I), por el teorema de Tonelli-Fubini y la proposición 2.45,

$$\begin{aligned}
\int f \circ T d\lambda^n &= \int \left(\int f(x_1, \ldots, cx_j, \ldots, x_n) d\lambda(x_j) \right) d\big(\Pi_{i \neq j} \lambda(x_i)\big) \\
&= \int |c|^{-1} \left(\int f(x_1, \ldots, x_j, \ldots, x_n) d\lambda(x_j) \right) d\big(\Pi_{i \neq j} \lambda(x_i)\big) = |c|^{-1} \int f d\lambda^n.
\end{aligned}$$

En consecuencia $\int f d\lambda^n = |\det(T)| \int f \circ T d\lambda^n$. Para operadores de tipo (II) es análogo al caso anterior ya que en la integral iterada tendremos una traslación. Para operadores de tipo (III) se sigue de que la medida producto $\prod_{i=1}^n \lambda$ es simétrica. Ahora, si el resultado es válido para $S, T \in GL(n, \mathbb{R})$, tenemos

$$\begin{aligned}
|\det(T \circ S)| \int f \circ (T \circ S) d\lambda^n &= |\det(T)||\det(S)| \int (f \circ T) \circ S d\lambda^n \\
&= |\det(T)| \int f \circ T d\lambda^n = \int f d\lambda^n.
\end{aligned}$$

Dado que todo operador $T \in GL(n, \mathbb{R})$ es la composición de una cantidad finita de operadores de tipo (I), (II) y (III), utilizando el argumento anterior de manera inductiva, podemos obtener que

$$|\det(T)| \int f \circ T d\lambda^n = \int f d\lambda^n.$$

Claramente para todo $E \in \mathcal{B}(\mathbb{R}^n)$, $T(E) \in \mathcal{B}(\mathbb{R}^n)$ y aplicando lo que probamos previamente a T^{-1} y $f = \mathbb{1}_E$, se sigue que

$$\lambda^n\big(T(E)\big) = \int \mathbb{1}_{T(E)} d\lambda^n = |\det(T)||\det(T^{-1})| \int \mathbb{1}_E \circ T^{-1} d\lambda^n = |\det(T)|\lambda^n(E).$$

Esto se extiende para todo $E \in \mathcal{L}(\mathbb{R}^n)$ por el mismo argumento de la demostración de la proposición 2.45 y podemos concluir 1).
2) De la primera parte es evidente que T es Lebesgue-medible y para todo $E \in \mathcal{L}(\mathbb{R}^n)$,

$$T_*\lambda^n(E) = \lambda^n\big(T^{-1}(E)\big) = |\det(T^{-1})|\lambda^n(E) = |\det(T)|^{-1}\lambda^n(E).$$

Por el teorema 2.38 podemos concluir. □

Ya estamos en condiciones de probar el teorema de cambio de variable de cursos elementales de cálculo pero antes requerimos el siguiente resultado:

Lema 2.47. *Todo abierto $U \subset \mathbb{R}^n$ es la unión numerable de cubos de interiores disjuntos.*

Demostración. Para cada $k \in \mathbb{N}$ denotemos por $\mathcal{Q}_k$ al conjunto de cubos de longitud 2^{-k} y vértices diádicios, $\prod_{i=1}^n \left[\frac{m_i}{2^k}, \frac{m_i+1}{2^k}\right]$, donde $m_1, \dots, m_n \in \mathbb{Z}$. Sea

$$\mathcal{I}(U,k) = \bigcup \{Q \in \mathcal{Q}_k : Q \subset U\}.$$

Claramente $\mathcal{I}(U,k) \subset \mathcal{I}(U,k+1)$. Ahora, dado $x \in U$ sea $\delta = \operatorname{dist}(x, U^c) > 0$. Tómese $k \in \mathbb{N}$ tal que $2^{-k}\sqrt{n} < \delta$. Dado que para algún $Q \in \mathcal{Q}_k$, $x \in Q$ y $\operatorname{diam}(Q) = 2^{-k}\sqrt{n}$, se sigue que $Q \subset U$. Por lo tanto existe un primer entero k para el cual $x \in \mathcal{I}(U,k)$ y tenemos $U = \mathcal{I}(U,1) \cup \bigcup_{k\in\mathbb{N}} \big(\mathcal{I}(U,k+1) - \mathcal{I}(U,k)\big)$. Podemos concluir ya que para cada $k \in \mathbb{N}$, la cerradura de $\mathcal{I}(U,k+1) - \mathcal{I}(U,k)$ es la unión numerable de cubos en $\mathcal{Q}_k$. □

Dado un abierto $\Omega \subset \mathbb{R}^n$ y $G : \Omega \to \mathbb{R}^n$, decimos que G es un difeomorfismo de clase C^1 si G es inyectivo y D_xG es continuo (en cada entrada) e invertible para todo $x \in \Omega$. Por el teorema de la función inversa $G^{-1} : G(\Omega) \to \Omega$ también es un difeomorfismo y $D_y\big(G^{-1}\big) = \big(D_{G^{-1}(y)}G\big)^{-1}$ para todo $y \in G(\Omega)$.

Teorema 2.48. *Sea $\Omega \subset \mathbb{R}^n$ abierto y $G : \Omega \to \mathbb{R}^n$ un difeomorfismo de clase C^1. Entonces,*

1) *Si $E \in \mathcal{L}(\mathbb{R}^n) \cap \Omega$, entonces $G(E) \in \mathcal{L}(\mathbb{R}^n)$ y $\lambda^n\big(G(E)\big) = \int_E |\det(D_xG)| d\lambda^n(x)$.*

2) *Si f es Lebesgue-medible en $G(\Omega)$ e integrable (ó positiva), entonces,*

$$x \mapsto f \circ G(x) |\det(D_xG)|$$

es integrable (ó positiva) en Ω y $\int_{G(\Omega)} f d\lambda^n = \int_\Omega f \circ G(x)\, |\det(D_xG)| d\lambda^n(x)$.

Demostración. 1) Dado $T \in GL(n,\mathbb{R})$ definamos $\|T\| := \max_{1\le i\le n} \sum_{j=1}^n |T_{i,j}|$. Consideremos un cubo Q de diámetro $2h$ centrado en $a \in \Omega$. Para $1 \le j \le n$ y todo $x \in Q$ existe un y en el segmento $[a,x]$ tal que $G_j(x) - G_j(a) = \nabla G_j(y) \cdot (x-a)$. En consecuencia,

$$\max_{1\le j\le n} |G_j(x) - G_j(a)| \le \sup_{y\in Q} \max_{1\le j\le n} |\nabla G_j(y)||x-a| \le \sup_{y\in Q} \|D_yG\| h.$$

Podemos concluir que $G(Q)$ está contenido en el cubo de diámetro $\sup_{y\in Q} \|D_yG\| 2h$ y centrado en $G(a)$. Por las propiedades de reescalamiento e minvarianza bajo traslación de λ, se sigue que

$$\lambda\big(G(Q)\big) \le \Big(\sup_{y\in Q} \|D_yG\|\Big)^n \lambda(Q).$$

Para $x \in \Omega$ fijo consideremos $(D_x G)^{-1}$ como operador lineal en $\mathbb{R}^n$. Aplicando el razonamiento previo al difeomorfismo $(D_x G)^{-1} \circ G$ y la proposición 2.46 vemos que

$$\lambda\big(G(Q)\big) = |\det(D_x G)|\lambda\big((D_x G)^{-1} \circ G(Q)\big) \leq |\det(D_x G)|\Big(\sup_{y \in Q} \|(D_x G)^{-1} D_y G\|\Big)^n \lambda(Q).$$

Nótese que lo anterior se cumple para todo $x \in \Omega$ y todo cubo $Q \subset \Omega$. Ahora, como G es C^1, el mapeo $(x, y) \mapsto (D_x G)^{-1} D_y G$ es continuo en cada entrada y en consecuencia también con respecto a la norma $\| \cdot \|$. Sea $\epsilon > 0$, dado que Q^2 es compacto, existe $\delta > 0$ tal que si $(x, y), (x', y') \in Q^2$ y $|(x, y) - (x', y')| < \delta$, entonces $\|(D_x G)^{-1} D_y G - (D_{x'} G)^{-1} D_{y'} G\| < \epsilon$. En particular, si $x, y \in Q$ cumplen $|x - y| < \delta$, tenemos $|(x, y) - (y, y)| < \delta$ y se sigue que $\|(D_x G)^{-1} D_y G - Id\| < \epsilon$. Esto a su vez implica que

$$\|(D_x G)^{-1} D_y G\| < 1 + \epsilon.$$

Podemos dividir a Q en una cantidad finita de subcubos, $Q_1, \ldots, Q_k$ diámetro menor a δ y centrados en $x_1, \ldots, x_k$, respectivamente, para obtener

$$\begin{aligned}\lambda\big(G(Q)\big) &\leq \sum_{j=1}^{k} \lambda\big(G(Q_j)\big) \leq \sum_{j=1}^{k} |\det(D_{x_j} G)|\Big(\sup_{y \in Q_j} \|(D_{x_j} G)^{-1} D_y G\|\Big)^n \lambda(Q_j) \\ &\leq (1 + \epsilon)^n \sum_{j=1}^{k} |\det(D_{x_j} G)|\lambda(Q_j).\end{aligned}$$

Notamos que el último término coincide con la integral de $\sum_{j=1}^{k} |\det(D_{x_j} G)|\, \mathbb{1}_{Q_j}$. Si hacemos que el diámetro de los cubos Q_j tienda a cero obtenemos una sucesión de funciones uniformemente acotadas por $2 \sup_{y \in Q} |\det(D_y G)|$ y que convergen c.d.s. a $x \mapsto |\det(D_x G)|$ en Q. Por el teorema de convergencia dominada y dado que ϵ es arbitrario, llegamos a la desigualdad

$$\lambda\big(G(Q)\big) \leq \int_Q |\det(D_x G)| d\lambda(x).$$

Ahora, sea $U \subset \Omega$ abierto. Por el lema 2.47 $U = \bigcup_{k \in \mathbb{N}} Q_k$, donde los conjuntos Q_k son cubos de interiores disjuntos. Dado que la frontera de los cubos tiene medida de Lebesgue cero, se sigue que

$$\lambda\big(G(U)\big) \leq \sum_{k=1}^{\infty} \lambda\big(G(Q_k)\big) \leq \sum_{k=1}^{\infty} \int_{Q_k} |\det(D_x G)| d\lambda(x) = \int_U |\det(D_x G)| d\lambda(x).$$

Sea $E \in \mathcal{B}(\mathbb{R}^n) \cap \Omega$. Podemos particionar E en una cantidad numerable de conjuntos de medida finita $E_m \in \mathcal{B}(\mathbb{R}^n) \cap \Omega$. Luego, por la regularidad de la medida de Lebesgue para cada $m, k \in \mathbb{N}$ existen abiertos U_m^k tales que $E_m \subset U_m^k$ y $\lambda(U_m^k - E_m) \leq \frac{1}{k}$. Entonces $E_m \subset \bigcap_{k \in \mathbb{N}} U_m^k$, $\lambda\big(\bigcap_{k \in \mathbb{N}} U_m^k - E_m\big) \leq \lim_{k \to \infty} \lambda(U_m^k - E_m) = 0$ y

$$\begin{aligned}\lambda\big(G(E)\big) &\leq \sum_{m=1}^{\infty} \lambda\bigg(G\Big(\bigcap_{k \in \mathbb{N}} U_m^k\Big)\bigg) = \sum_{m=1}^{\infty} \lim_{k \to \infty} \lambda\bigg(G\Big(\bigcap_{j=1}^{k} U_m^j\Big)\bigg) \\ &\leq \sum_{m=1}^{\infty} \lim_{k \to \infty} \int_{\bigcap_{j=1}^{k} U_m^j} |\det(D_x G)| d\lambda(x) = \sum_{m=1}^{\infty} \int_{\bigcap_{k \in \mathbb{N}} U_m^k} |\det(D_x G)| d\lambda(x) \\ &= \sum_{m=1}^{\infty} \int_{E_m} |\det(D_x G)| d\lambda(x) = \int_E |\det(D_x G)| d\lambda(x).\end{aligned}$$

Para obtener la otra desigualdad consideremos la medida definida en $\mathcal{B}(\mathbb{R}^n) \cap G(E)$ y dada por $\nu(A) = \int_A |\det(D_y G^{-1})| d\lambda(y)$. Aplicando la primera parte de la demostración a G^{-1}, obtenemos que para todo $A \in \mathcal{B}(\mathbb{R}^n) \cap G(E)$

$$G_*\lambda(A) = \lambda\big(G^{-1}(A)\big) \leq \int_A |\det(D_y G^{-1})| d\lambda(y) = \nu(A).$$

Luego, al integrar la función positiva $y \mapsto |\det(D_{G^{-1}(y)} G)|$ con respecto a estas medidas y aplicar el teorema 2.38 se sigue que

$$\begin{aligned}\int_E |\det(D_x G)| d\lambda(x) &= \int_{G(E)} |\det(D_{G^{-1}(y)} G)| d\big(\lambda_* G(y)\big) \\ &\leq \int_{G(E)} |\det(D_{G^{-1}(y)} G)| d\nu(y) \\ &= \int_{G(E)} |\det(D_{G^{-1}(y)} G)| |\det(D_y G^{-1})| d\lambda(y) = \lambda\big(G(E)\big).\end{aligned}$$

Por lo tanto, para todo $E \in \mathcal{B}(\mathbb{R}^n) \cap \Omega$ tenemos $\lambda\big(G(E)\big) = \int_E |\det(D_x G)| d\lambda(x)$. Podemos extender esta identidad a todo $\mathcal{L}(\mathbb{R}^n) \cap \Omega$ como en la proposición 2.45.

2) Por la primera parte tenemos que para todo $E \in \mathcal{L}(\mathbb{R}^n) \cap G(\Omega)$,

$$G_*^{-1}\lambda(E) = \lambda\big(G(E)\big) = \int_E |\det(D_x G)| d\lambda(x).$$

Una conjunción del el teorema 2.38 y la proposición 2.31 aplicadas a $f \circ G$, nos permite concluir que

$$\int_{G(\Omega)} f d\lambda = \int_{G(\Omega)} f \circ G \circ G^{-1} d\lambda = \int_\Omega f \circ G \, d\big(G_*^{-1}\lambda\big) = \int_\Omega f \circ G(x) |\det(D_x G)| d\lambda(x). \quad \square$$

2.5. Convolución

Una de las aplicaciones esenciales de la medida producto es es el poder definir la operación de *convolución* entre medidas. Para definir esta operación consideremos que X es un conjunto distinto del vacío y $\cdot : X \times X \to X$ es una operación binaria, que escribiremos de forma multiplicativa, sobre éste tal que $(X, \cdot)$ sea un grupo. Si $\Sigma \subset \mathcal{P}(X)$ es una σ-álgebra de conjuntos tales que el producto sea $(\Sigma \otimes \Sigma, \Sigma)$-medible y la operación de inversión sea (Σ, Σ)-medible[1] entonces podremos definir la convolución de dos medidas como sigue.

Definición 2.49. Sean $\mu, \nu : \Sigma \to [0, \infty]$ dos medidas sobre (X, Σ). Definimos la convolución de μ y ν inducida por $\cdot$ como la medida $\mu * \nu : \Sigma \to [0, \infty]$ definida por

$$(\mu * \nu)(B) := \int \mathbb{1}_B(xy) d(\mu \times \nu)(x, y).$$

[1]Éste es el caso de los *grupos topológicos*.

En el caso particular que μ y ν sean medidas σ-finitas entonces tendremos las siguientes igualdades, por la proposión 2.41 y una propiedad de los grupos, tendremos que

$$(\mu * \nu)(B) = \int\int \mathbb{1}_B(xy)d\nu(y)d\mu(x) = \int\int \mathbb{1}_{x^{-1}B}(y)d\nu(y)d\mu(x) = \int \nu(x^{-1}B)d\mu(x).$$

Análogamente se cumple que

$$(\mu * \nu)(B) = \int\int \mathbb{1}_B(xy)d\mu(x)d\nu(y) = \int\int \mathbb{1}_{By^{-1}}(x)d\mu(x)d\nu(y) = \int \mu(By^{-1})d\nu(y).$$

Esto nos permite ver que, en general, $\mu * \nu \neq \nu * \mu$. La igualdad se da, por ejemplo, cuando $(X, \cdot)$ es un grupo abeliano y ambas medidas son σ-finitas.

Si además consideramos $\mu_1, \mu_2, \dots, \mu_n$ medidas σ-finitas sobre (X, Σ) entonces podremos definir de forma inductiva

$$\mu_1 * \mu_2 * \cdots * \mu_n = (\mu_1 * \cdots * \mu_{n-1}) * \mu_n.$$

En el caso de que $(X, \cdot) = (\mathbb{R}^n, +)$ entonces recuperaremos la convolución más familiar

$$(\mu * \nu)(B) = \int \mathbb{1}_B(x+y)d(\mu \times \nu)(x, y) = \int \mu(B-y)d\nu(y) = \int \nu(B-x)d\mu(x),$$

la cual será interpretada más adelante en términos probabilísticos. En este mismo caso tendremos los siguientes resultados.

Proposición 2.50. *Si $f \in L^+(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ con μ y ν medidas σ-finitas entonces*

$$\int f(x+y)d(\mu \times \nu)(x, y) = \int f(z)d(\mu * \nu)(z).$$

Similarmente, si $f \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu * \nu)$ entonces se cumple la ecuación anterior.

Demostración. La primera parte se cumple por definición para las funciones indicadoras y por linealidad de la integral, se sigue que también se cumple para funciones simples $\varphi \in L^+(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Para $f \in L^+(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ general, sea $\{\varphi_n\} \subset L^+(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ una sucesión creciente de funciones simples tales que $f = \lim_{n\to\infty} \varphi_n$, por el teorema de convergencia monótona tendremos que

$$\begin{aligned}\int f(z)d(\mu * \nu)(z) &= \lim_{n\to\infty} \int \varphi(z)d(\mu * \nu)(z) = \lim_{n\to\infty} \int \varphi(x+y)d(\mu \times \nu)(x, y)\\ &= \int f(x+y)d(\mu \times \nu)(x, y).\end{aligned}$$

Para $f \in L^1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu * \nu)$ basta pensar en el caso real pues el caso complejo se obtendrá como consecuencia. Así

$$\begin{aligned}\int f(z)d(\mu * \nu)(z) &= \int f^+(z)d(\mu * \nu)(z) - \int f^-(z)d(\mu * \nu)(z)\\ &= \int f^+(x+y)d(\mu \times \nu)(x, y) - \int f^-(x+y)d(\mu \times \nu)(x, y)\\ &= \int f(x+y)d(\mu \times \nu)(x, y).\end{aligned}$$

□

Proposición 2.51. *Si μ y ν son medidas σ-finitas sobre $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ con λ^n-densidades f y g entonces $\mu * \nu$ tendrá λ^n-densidad dada por*

$$(f * g)(t) := \int f(s)g(t-s)d\lambda^n(s) = \int f(t-s)g(s)d\lambda^n(s).$$

Demostración. Notemos que por definición tendremos que

$$\begin{aligned}(\mu * \nu)(B) &= \int \int \mathbb{1}_B(x+y)d\nu(y)d\mu(x) = \int \int f(x)g(y)\mathbb{1}_B(x+y)d\lambda^n(y)d\lambda^n(x)\\ &= \int f(x)g(y)\mathbb{1}_B(x+y)d(\lambda^n \times \lambda^n)(x,y).\end{aligned}$$

Por el teorema 2.48 bajo la transformación $(x,y) \mapsto (x, x+y) =: (s,t)$ tendremos que

$$\int f(x)g(y)\mathbb{1}_B(x+y)d(\lambda^n \times \lambda^n)(x,y) = \int f(s)g(t-s)\mathbb{1}_B(t)d(\lambda^n \times \lambda^n)(s,t),$$

de donde se sigue, mediante una aplicación del teorema de Fubini, que

$$(\mu * \nu)(B) = \int \mathbb{1}_B(t)\left(\int f(s)g(t-s)d\lambda^n(s)\right)d\lambda^n(t) = \int \mathbb{1}_B(t)(f*g)(t)d\lambda^n(t).$$

La otra igualdad se prueba con un argumento análogo. $\square$

Capítulo 3

Espacios L^p y Medidas con Signo

3.1. Espacios L^p

Introducimos a los espacios L^p que son de mucha importancia en distintas áreas de las matemáticas como el Análisis Funcional y las Ecuaciones Diferenciales Parciales. Como veremos en esta sección, el principal atributo de estos espacios es que son espacios de Banach.

Consideremos un espacio de medida (X, Σ, μ) y definamos

$$L^p(X, \Sigma, \mu) := \Big\{ f : X \to \mathbb{C} : f \textit{ es medible y } \int |f|^p d\mu < \infty \Big\},$$

para $0 < p < \infty$. A menos que sea necesario especificar el espacio abreviamos $L^p(X, \Sigma, \mu)$ simplemente por L^p. Previamente ya hemos definido este espacio para el caso $p = 1$. Podemos ver fácilmente que en general L^p es un espacio vectorial ya que si $f, g \in L^p$,

$$|f + g|^p \leq \big(2 \,\mathrm{máx}(|f|, |g|)\big)^p = 2^p \,\mathrm{máx}(|f|^p, |g|^p) \leq 2^p\big(|f|^p + |g|^p\big)$$

y por monotonía de la integral se sigue que $f + g \in L^p$. Para todo $f \in L^p$ definimos

$$\|f\|_p := \left(\int |f|^p d\mu \right)^{1/p}.$$

Claramente para todo $\alpha \in \mathbb{C}$, $\|\alpha f\|_p = |\alpha| \|f\|_p$. Ahora, por la proposición 2.32 sabemos que $\|f\|_p = 0$ implica $f = 0$ c.d.s. Si convenimos en que todo $f \in L^p$ es una clase de equivalencia bajo la razón de equivalencia $f \sim g$ si y sólo si $f = g$ c.d.s., entonces $\|f\|_p = 0$ si y sólo si $f = 0$ (la clase de equivalencia representada por el cero). Bajo esta convención, sólo nos queda probar la desigualdad del triangulo para tener que $\|\cdot\|_p$ es una norma en L^p. Este es el caso cuando $p \geq 1$ pero para probarlo requerimos un resultado previo que será de mucha importancia y es conocido como la *desigualdad de Hölder*.

Lema 3.1. *Sean* $a, b \geq 0$ *y* $0 < \lambda < 1$. *Entonces,*

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1 - \lambda) b$$

y la igualdad se cumple si y sólo si $a = b$.

Demostración. El resultado es trivial cuando $a = 0$ ó $b = 0$. Suponiendo $a, b > 0$ consideremos a la función $f : \mathbb{R}^+ \to \mathbb{R}$ dada por $f(t) = t^\lambda - \lambda t$. Su derivada, $f'(t) = \lambda t^{\lambda - 1} - \lambda$ es positiva para $t < 1$ y negativa para $t > 1$. Se sigue que f toma su máximo en $t = 1$ por lo que

$$(a/b)^\lambda - \lambda(a/b) \leq 1 - \lambda$$

y la igualdad se cumple si y sólo si $a/b = 1$. Al agregar $\lambda(a/b)$ y luego multiplicar por b de ambos lados de la desigualdad, llegamos al resultado que buscábamos. □

Dados $p, q \in (1, \infty)$ decimos que éstos son **exponentes conjugados** si y sólo si $\frac{1}{p} + \frac{1}{q} = 1$. De ser el caso tenemos que $q = p/(p-1)$.

Proposición 3.2. *(Desigualdad de Hölder) Sean $p, q \in (1, \infty)$ exponentes conjugados. Si $f \in L^p$ y $g \in L^q$ entonces $fg \in L^1$ y*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

La igualdad se cumple si y sólo si $a|f|^p = b|g|^q$ c.d.s para constantes $a, b \geq 0$ tales que $ab \neq 0$.

Demostración. El caso en que $f = 0$ ó $g = 0$ c.d.s. es trivial así que lo podemos descartar. Por el lema 3.1,

$$\frac{|fg|}{\|f\|_p \|g\|_q} = \left(\frac{|f|^p}{\|f\|_p^p}\right)^{1/p} \left(\frac{|g|^q}{\|g\|_q^q}\right)^{1-1/p} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g|^q}{\|g\|_q^q}. \tag{3.1}$$

En consecuencia

$$\frac{1}{\|f\|_p \|g\|_q} \int |fg| d\mu \leq \frac{1}{p\|f\|_p^p} \int |f|^p d\mu + \frac{1}{q\|g\|_q^q} \int |g|^q d\mu = \frac{1}{p} + \frac{1}{q} = 1,$$

de donde se sigue la desigualdad que buscamos. Para finalizar nótese que la igualdad se cumple si y sólo si (3.1) es igualdad c.d.s. También por el lema 3.1 esto es cierto si y sólo si $\|g\|_q^q |f|^p = \|f\|_p^p |g|^q$ c.d.s. $\square$

Proposición 3.3. *(Desigualdad de Minkowski) Si $1 \leq p < \infty$ y $f, g \in L^p$, entonces,*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Demostración. Los casos $p = 1$ y $f + g = 0$ c.d.s. son triviales así que los descartamos. Sea q el exponente conjugado de p. Dado que $|f + g| \in L^p$ tenemos $(|f + g|)^{p-1} \in L^q$ puesto que $(p-1)q = p$. Ahora, del hecho de que $|f + g|^p \leq (|f| + |g|)|f + g|^{p-1}$ y la desigualdad de Hölder podemos ver que

$$\begin{aligned} \|f + g\|_p^p &= \int |f + g|^p d\mu \\ &\leq \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu \\ &\leq \big(\|f\|_p + \|g\|_p\big) \left(\int |f + g|^{(p-1)q} d\mu\right)^{1/q} = \big(\|f\|_p + \|g\|_p\big) \|f + g\|_p^{p/q}. \end{aligned}$$

Multiplicando ambos lados de la última desigualdad por $\|f + g\|_p^{-p/q}$ llegamos al resultado que buscamos. $\square$

Con la desigualdad de Minkowski concluimos que $\|\cdot\|_p$ es una norma en L^p para $1 \leq p < \infty$. Si $f_n, f \in L^p$ decimos que la sucesión $\{f_n\}_{n\in\mathbb{N}}$ converge a f en L^p si y sólo si $\lim_{n\to\infty} \|f - f_n\| = 0$ y lo denotamos por $f_n \to f$ en L^p ó $f_n \to^{L^p} f$. Veamos ahora que bajo esta norma L^p es completo.

Teorema 3.4. *Para $1 \leq p < \infty$ el espacio normando $(L^p, \|\cdot\|_p)$ es completo. Más aún, si $f_n \to^{L^p} f$, entonces existe una subsucesión $\{f_{n_k}\}_{k\in\mathbb{N}}$ que converge a f c.d.s.*

Demostración. Sea $\{f_n\}_{n\in\mathbb{N}} \subset L^p$ una sucesión de Cauchy. Podemos encontrar una subsucesión $\{f_{n_k}\}_{k\in\mathbb{N}}$ tal que $\|f_{n_{k+1}} - f_{n_k}\|_p < 2^{-k}$ para todo $k \in \mathbb{N}$. Consideremos a la función medible (y posiblemente extendida) $g := \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}|$. Por el lema de Fatou y la desigualdad de Minkowski

$$\int g^p d\mu \leq \liminf_k \int \left(\sum_{j=1}^{k} |f_{n_{j+1}} - f_{n_j}| \right)^p d\mu \leq \liminf_k \left(\sum_{j=1}^{k} \|f_{n_{j+1}} - f_{n_j}\|_p \right)^p \leq 1.$$

Podemos concluir que $g < \infty$ c.d.s. que a su vez implica que

$$f(x) = \lim_{k\to\infty} f_{n_k}(x) = f_{n_1}(x) + \lim_{k\to\infty} \sum_{j=1}^{k} \big(f_{n_{j+1}}(x) - f_{n_j}(x)\big) \ \textit{ existe c.d.s.}$$

ya que la serie del lado derecho es absolutamente convergente c.d.s. Para x en un conjunto nulo donde posiblemente este límite no exista definimos $f(x) = 0$. Veamos que $f_n \to f$ en L^p. Dado $\epsilon > 0$ sea N tal que para todo $n, m \geq N$, $\|f_n - f_m\|_p < \epsilon$. Nuevamente por Fatou, sabemos que para todo $n \geq N$,

$$\int |f_n - f|^p d\mu \leq \liminf_k \int |f_n - f_{n_k}|^p d\mu < \epsilon^p.$$

Podemos concluir que $f = (f - f_n) + f_n \in L^p$ y que la sucesión $\{f_n\}_{n\in\mathbb{N}}$ converge a f en L^p. Para la segunda parte del enunciado, notamos que si $f_n \to^{L^p} f$, entonces $\{f_n\}_{n\in\mathbb{N}}$ es una sucesión es Cauchy y por la primera parte de la demostración, existe una subsucesión $\{f_{n_k}\}_{k\in\mathbb{N}}$ que converge a una función $\tilde{f} \in L^p$ c.d.s. y en L^p. Dado que el límite en cualquier espacio normado es único, tenemos que $f = \tilde{f}$ c.d.s. y podemos concluir. $\square$

Ahora introducimos el espacio L^p para el caso peculiar $p = \infty$. Para cualquier función medible (real o compleja) definimos

$$\|f\|_\infty := \inf \Big\{ M \geq 0 : \mu\big(\{x \in X : |f(x)| > M\}\big) = 0 \Big\},$$

donde convenimos que $\inf \emptyset = \infty$. El valor $\|f\|_\infty$ es llamado el **supremo esencial** de f y esto se debe a que $|f| \leq \|f\|_\infty$ c.d.s. Para ver esto tomemos una sucesión decreciente M_k tal que $M_k \downarrow \|f\|_\infty$ y $\mu\big(\{x \in X : |f(x)| > M_k\}\big) = 0$. Dado que

$$\{x \in X : |f(x)| > \|f\|_\infty\} = \bigcup_{k\in\mathbb{N}} \{x \in X : |f(x)| > M_k\},$$

podemos concluir por propiedades de continuidad de las medidas. Así, definimos

$$L^\infty \equiv L^\infty(X, \Sigma, \mu) := \big\{ f : X \to \mathbb{C} : f \textit{ es medible y } \|f\|_\infty < \infty \big\}.$$

Veamos que las propiedades que hemos probado para $1 \leq p < \infty$ se extienden al caso $p = \infty$.

Teorema 3.5.

1) *Si $f \in L^1$ y $g \in L^\infty$, entonces $fg \in L^1$ y $\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$. La igualdad se cumple si y sólo si $|g| = \|g\|_\infty$ c.d.s. en $\{x : f(x) \neq 0\}$.*

2) *$\|\cdot\|_\infty$ es una norma en L^∞.*

3) L^∞ *es un espacio de Banach.*

4) $f_n \to^{L^\infty} f$ *si y sólo si existe un conjunto* E *tal que* $\mu(E^c) = 0$ *y en* E $f_n \to f$ *c.d.s.*

Demostración. 1) Sabemos que $|fg| \leq |f| \|g\|_\infty$ c.d.s. por lo que

$$\|fg\|_1 = \int |fg| d\mu \leq \int |f| \|g\|_\infty d\mu = \|f\|_1 \|g\|_\infty.$$

Además la igualdad se da si y sólo si

$$\big\{x : |f(x)| \|g\|_\infty > |f(x)g(x)|\big\} = \big\{x : \|g\|_\infty > |g(x)|\big\} \cap \big\{x : f(x) \neq 0\big\}$$

es un conjunto nulo.
2) Claramente $\|f\|_\infty = 0$ si y sólo si $f = 0$ c.d.s. Luego, dado $\alpha \in \mathbb{C} - \{0\}$,

$$\begin{aligned}\|\alpha f\|_\infty &= \operatorname{ínf}\Big\{M \geq 0 : \mu\big(\{x \in X : |\alpha||f(x)| > M\}\big) = 0\Big\} \\ &= \operatorname{ínf}\Big\{|\alpha| N \geq 0 : \mu\big(\{x \in X : |\alpha||f(x)| > |\alpha| N\}\big) = 0\Big\} \\ &= |\alpha| \operatorname{ínf}\Big\{N \geq 0 : \mu\big(\{x \in X : |f(x)| > N\}\big) = 0\Big\} = |\alpha| \|f\|_\infty.\end{aligned}$$

Finalmente, dado que $|f + g| \leq |f| + |g| \leq \|f\|_\infty + \|g\|_\infty$ c.d.s es inmediato que

$$\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty.$$

3) Si $\{f_n\}_{n \in \mathbb{N}} \subset L^\infty$ es una sucesión de Cauchy, entonces para todo $k \in \mathbb{N}$ existe $N_k \in \mathbb{N}$ tal que para $n, m \geq N_k$, $\|f_n - f_m\|_\infty < 1/k$; es decir, $|f_n - f_m| < 1/k$ c.d.s. Luego, por numerabilidad de los pares (n, m) se sigue que existe un conjunto nulo E_k tal que si $n, m \geq N_k$ y $x \in E_k^c$, entonces $|f_n(x) - f_m(x)| < 1/k$. Si tomamos $E = \bigcup_{k \in \mathbb{N}} E_k$, dado que $E^c = \bigcap_{k \in \mathbb{N}} E_k^c$, podemos ver que en E^c la sucesión $\{f_n\}_{n \in \mathbb{N}}$ es uniformemente Cauchy. De lo anterior tenemos que para todo $x \in E^c$, $f(x) = \lim_{n \to \infty} f_n(x)$ existe. Sólo queda probar que $f_n \to f$ uniformemente en E^c. Pero como $\{f_n\}_{n \in \mathbb{N}}$ es uniformemente Cauchy en E^c, dado $\epsilon > 0$ existe N tal que $|f_n(x) - f_m(x)| < \epsilon/2$ para todo $n, m \geq N$ y $x \in E^c$. Así, para todo $n \geq N$ y $x \in E^c$ tenemos

$$|f(x) - f_n(x)| \leq |f(x) - f_m(x)| + |f_m(x) - f_n(x)| < \epsilon/2 + \epsilon/2 = \epsilon,$$

tomando m suficientemente grande.
4) Se sigue de la demostración del inciso anterior. □

Para concluir esta sección probaremos algunas contenciones y desigualdades importantes.

Proposición 3.6. *Si* $0 < r < p < s \leq \infty$*, entonces* $L^r \cap L^s \subset L^p$ *y para todo* $f \in L^r \cap L^s$*,*

$$\|f\|_p \leq \|f\|_r^\lambda \|f\|_s^{1-\lambda},$$

donde $\frac{1}{p} = \lambda \frac{1}{r} + (1 - \lambda) \frac{1}{s}$ *y* $\lambda = \frac{p^{-1} - s^{-1}}{r^{-1} - s^{-1}}$.

Demostración. En el caso $s = \infty$ tenemos $|f|^p \leq \|f\|_\infty^{p-r}|f|^r$ c.d.s. y en este caso convenimos en que $\lambda = r/p$. Integrando obtenemos que $\|f\|_p^p \leq \|f\|_r^r\|f\|_\infty^{p-r}$ de donde se sigue fácilmente la desigualdad que buscamos.
Si $s < \infty$ simplemente notamos que $r/\lambda p$ y $s/(1-\lambda)p$ son exponentes conjugados y por la desigualdad de Hölder,

$$\begin{aligned}\int |f|^p d\mu &= \int |f|^{\lambda p}|f|^{(1-\lambda)p} d\mu \\ &\leq \left(\int |f|^{\lambda p \frac{r}{\lambda p}} d\mu\right)^{\frac{\lambda p}{r}} \left(\int |f|^{(1-\lambda)p\frac{s}{(1-\lambda)p}} d\mu\right)^{\frac{(1-\lambda)p}{s}} = \|f\|_r^{\lambda p}\|f\|_s^{(1-\lambda)p}.\end{aligned}$$ □

Proposición 3.7. *Si $\mu(X) < \infty$ y $0 < p < r \leq \infty$, entonces $L^r \subset L^p$ y para todo $f \in L^r$,*

$$\|f\|_p \leq \|f\|_r \, \mu(X)^{\frac{1}{p}-\frac{1}{r}}.$$

Demostración. En el caso que $r = \infty$,

$$\int |f|^p d\mu \leq \int \|f\|_\infty^p d\mu = \|f\|_\infty^p \int 1 d\mu = \|f\|_\infty^p \, \mu(X).$$

Si $r < \infty$, consideramos a los exponentes conjugados r/p y $r/(r-p)$ y utilizamos Hölder para ver que

$$\int |f|^p d\mu = \left(\int |f|^{p\frac{r}{p}} d\mu\right)^{p/r} \left(\int 1 d\mu\right)^{(r-p)/r} = \|f\|_r^p \, \mu(X)^{1-\frac{p}{r}}.$$ □

Veamos un ejemplo para mostrar que en la última proposición, la hipótesis $\mu(X) < \infty$ es necesaria. Consideremos el espacio de medida $([1,\infty), \mathcal{B}([1,\infty)), \lambda)$ donde λ es la medida de Lebesgue. Si tomamos $f(x) = 1/x$ podemos ver que

$$\int f^p d\lambda = \lim_{n\to\infty} \int_{[1,n]} f^p d\lambda = \lim_{n\to\infty} \int_1^n x^{-p} dx = \begin{cases} \lim_{n\to\infty} \frac{1}{1-p} x^{1-p}\Big|_1^n & \textit{si } p \neq 1, \\ \lim_{n\to\infty} \log x\Big|_1^n & \textit{si } p = 1. \end{cases}$$

De manera que $f \in L^p$ para $p > 1$ pero no para $0 < p \leq 1$.

3.2. Medidas con Signo

Definición 3.8. Sea (X, Σ) un espacio medible. Una medida con signo es una función $\nu : \Sigma \to [-\infty, \infty]$ tal que

1) $\nu(\emptyset) = 0$.

2) ν toma a lo más uno de los valores $\pm\infty$.

3) Si $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ son disjuntos, entonces $\nu(\bigcup_{n\in\mathbb{N}} E_n) = \sum_{n=1}^\infty \nu(E_n)$ y si $\nu(\bigcup_{n\in\mathbb{N}} E_n)$ es finito la serie es absolutamente convergente.

Si una medida sólo toma valores no-negativos diremos que es una **medida positiva**.

Proposición 3.9. *Sea ν es una medida con signo en (X, Σ).*

1) Si $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ es una sucesión creciente, entonces $\nu\big(\bigcup_{n\in\mathbb{N}} E_n\big) = \lim_{n\to\infty} \nu(E_n)$.

2) Si $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ es una sucesión decreciente y $\nu(E_1)$ es finito, entonces $\nu\big(\bigcap_{n\in\mathbb{N}} E_n\big) = \lim_{n\to\infty} \nu(E_n)$.

La demostración de este resultado es la misma que para el caso de una medida positiva (ver proposición 1.13)

Como ejemplo de una medida con signo consideremos dos medidas positivas μ_1 y μ_2, donde alguna es finita. Entonces $\nu = \mu_1 - \mu_2$ es una medida con signo: Las propiedades 1) y 2) son inmediatas. Mientras que

$$\nu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \mu_1\left(\bigcup_{n\in\mathbb{N}} E_n\right) - \mu_2\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \mu_1(E_n) - \sum_{n\in\mathbb{N}} \mu_2(E_n).$$

Si alguna de las series diverge, como la otra serie converge, entonces también $\sum_{n\in\mathbb{N}} \nu(E_n) = \sum_{n\in\mathbb{N}} \big(\mu_1(E_n) - \mu_2(E_n)\big)$ diverge y tenemos

$$\nu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \nu(E_n).$$

Si ambas series convergen claramente se cumple al igualdad anterior y la convergencia es absoluta.

Ahora supongamos que μ es una medida positiva. Otro ejemplo de medida con signo se obtiene al considerar una función $\big(\Sigma, \mathcal{B}(\overline{\mathbb{R}})\big)$-medible, f, tal que alguno de los valores $\int f^+ d\mu$ ó $\int f^- d\mu$ son finitos. Entonces $\nu(E) = \int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$ define una medida con signo. De hecho éste es un caso particular del ejemplo anterior. Sin embargo, si denotamos $P = \{x \in X : f(x) \geq 0\}$ y $N = \{x \in X : f(x) < 0\}$, tendremos que para todo subconjunto medible $E \subset P$, $\nu(E) \geq 0$ y para todo subconjunto medible $E \subset N$, $\nu(E) \leq 0$. Además de que $X = P \cup N$. Como veremos a continuación, este comportamiento es característico de las medidas con signo.

Definición 3.10. Sea ν es una medida con signo en (X, Σ). Decimos que $E \in \Sigma$ es un conjunto **positivo** (respectivamente **negativo** ó **nulo**) de ν si y sólo si $\nu(F) \geq 0$ (respectivamente ≤ 0 ó $= 0$) para todo subconjunto medible $F \subset E$.

Equivalentemente E es un conjunto positivo de ν si la restricción de ν en $\Sigma \cap E$, denotada por ν_E, es una medida positiva.

Lema 3.11. *Todo subconjunto medible de un conjunto positivo es positivo y la unión numerable de conjuntos positivos es un conjunto positivo.*

Demostración. La primera afirmación es trivial. Ahora, sea $\{P_n\}_{n\in\mathbb{N}} \subset \Sigma$ una sucesión de conjuntos positivos. Para cada $n \in \mathbb{N}$, sea $Q_n = P_n - \bigcup_{k=1}^{n-1} P_k$. Dado que $Q_n \subset P_n$, tenemos que Q_n es un conjunto positivo. Por lo tanto, para todo $E \in \Sigma$ tenemos

$$\nu\left(E \cap \bigcup_{n\in\mathbb{N}} P_n\right) = \sum_{n=1}^{\infty} \nu(E \cap Q_n) \geq 0.$$

Podemos concluir que $\bigcup_{n\in\mathbb{N}} P_n$ es un conjunto positivo de ν. $\square$

Teorema 3.12. *(Descomposición de Hahn) Si ν es una medida con signo en (X, Σ), entonces existen un conjunto positivo P y un conjunto negativo N, ambos para ν, tales que $X = P \cup N$ y $P \cap N = \emptyset$. Si P' y N' conforman otra descomposición que cumple lo anterior, tendremos que $P \Delta P' = N \Delta N'$ es un conjunto nulo de ν.*

Demostración. Sin mayor perdida de generalidad supongamos que ν no toma el valor $-\infty$. Sea

$$\alpha = \sup \left\{ \nu(P) : P \textit{ es un conjunto positivo de } \nu \right\}.$$

Nótese que α está bien definido ya que el conjunto vacío es un conjunto positivo para cualquier medida. Entonces existe una sucesión $\{P_n\}_{n\in\mathbb{N}}$ de conjuntos positivos tales que $\nu(P_n) \to \alpha$. Por el lema anterior sabemos que $P = \bigcup_{n\in\mathbb{N}} P_n$ es un conjunto positivo y por lo tanto $\nu(P) \leq \alpha$. Además, para todo $n \in \mathbb{N}$, $\nu(P) = \nu(P - P_n) + \nu(P_n) \geq \nu(P_n)$ y podemos concluir que $\nu(P) = \alpha$.
Sea $N = X - P$ y notemos que N no puede tener un subconjunto positivo E tal que $\nu(E) > 0$. En efecto, de ser así tendríamos que $P \cup E$ es positivo y $\nu(P \cup E) = \nu(P) + \nu(E) > \alpha$.
Veamos que N tiene que ser un conjunto negativo. De no ser así existe un conjunto medible $E \subset N$ tal que $\nu(E) > 0$. Como E no puede ser positivo podemos tomar el mínimo natural, n_1, para el cual existe un conjunto medible $E_1 \subset E$ tal que $\nu(E_1) < -1/n_1$. Luego tendremos que $\nu(E - E_1) > 0$ y $E - E_1$ tampoco puede ser un conjunto positivo. Entonces tomamos n_2, el mínimo natural para el cual existe $E_2 \subset E - E_1$ tal que $\nu(E_2) < -1/n_2$. De manera inductiva podemos tomar una sucesión de enteros $n_1, n_2, \cdots$ tal que n_k es el mínimo entero para el cual existe $E_k \subset E - \bigcup_{j=1}^{k-1} E_j$ con $\nu(E_k) < -1/n_k$. Como estamos suponiendo que ν no toma el valor $-\infty$ y dado que los E_k son disjuntos,

$$-\infty < \nu\left(\bigcup_{k\in\mathbb{N}} E_k\right) = \sum_{k=1}^{\infty} \nu(E_k) = -\sum_{k=1}^{\infty} \frac{1}{n_k}.$$

Podemos inferir que $n_k \to \infty$. Pero esto implica que $E - \bigcup_{k\in\mathbb{N}} E_k$ es un conjunto positivo ya que para todo subconjunto medible $A \subset E - \bigcup_{k\in\mathbb{N}} E_k \subset E - \bigcup_{j=1}^{k-1} E_j$ tenemos que $\nu(A) \geq -1/(n_k - 1)$ a partir de alguna k ya que la otra desigualdad contradice la elección de n_k. Esto es una contradicción puesto que además $\nu\big(E - \bigcup_{k\in\mathbb{N}} E_k\big) > 0$. Por lo tanto N es negativo.
Finalmente, supongamos P' y N' también satisfacen el enunciado. Entonces $P - P' = P \cap N'$ por lo que es un conjunto positivo y a la vez negativo; es decir, $P - P'$ es nulo. Análogamente podemos ver que $P' - P$ es nulo y concluimos que $P \Delta P'$ es nulo. □

Si P y N satisfacen el enunciado del teorema anterior decimos que éstos conforman una **descomposición de Hahn** para ν. Aunque la descomposición de Hahn usualmente no es única nos va a permitir dar una descomposición canónica de ν por lo cual introducimos la siguiente definición:

Definición 3.13. Sean μ y ν medidas en (X, Σ). Decimos que μ y ν son (mutuamente) **singulares** si existen conjuntos disjuntos $A, B \in \Sigma$ tales que $X = A \cup B$, A es un conjunto nulo de ν y B es un conjunto nulo de μ. Esto lo denotamos por $\nu \perp \mu$.

Teorema 3.14. *(Descomposición de Jordan) Si ν es una medida con signo existe una única dupla de medidas positivas ν^+ y ν^- tales que $\nu = \nu^+ - \nu^-$ y $\nu^+ \perp \nu^-$.*

Demostración. Sea $X = P \cup N$ una descomposición de Hahn. Entonces $\nu^+(E) = \nu(E \cap P)$ y $\nu^-(E) = -\nu(E \cap N)$ son positivas, singulares y claramente $\nu = \nu^+ - \nu^-$. Supongamos que $\nu = \mu^+ - \mu^-$ con $\mu^+ \perp \mu^-$ y positivas. Entonces $X = P' \cup N'$ con N' un conjunto nulo de μ^+ y P'

un conjunto nulo de ν^-. Se sigue que P' y N' conforman una descomposición de Hahn para ν y por lo tanto $P\Delta P'$ es un conjunto nulo. Así, para todo $E \in \Sigma$,

$$\begin{aligned}\mu^+(E) = \mu^+(E\cap P') = \nu(E\cap P') &= \nu(E\cap P'\cap P) + \nu\big(E\cap (P'-P)\big)\\ &= \nu(E\cap P'\cap P) + \nu\big(E\cap (P-P')\big) = \nu(E\cap P) = \nu^+(E).\end{aligned}$$

Es decir, $\mu^+ = \nu^+$ y análogamente $\mu^- = \nu^-$. □

La descomposición $\nu = \nu^+ - \nu^-$ es llamada la **descomposición de Jordan** de ν. Definimos a la **variación total** de ν como

$$|\nu| := \nu^+ + \nu^-.$$

Es fácil ver que $E \in \Sigma$ es un conjunto nulo de ν si y sólo si $\nu^+(E) = \nu^-(E) = 0$ si y sólo si $|\nu|(E) = 0$. Se sigue que $\nu \perp \mu$ si y sólo si $\nu^+ \perp \mu$ y $\nu^- \perp \mu$ si y sólo si $|\nu| \perp \mu$. Para hacer evidente esta última afirmación veamos que, de manera general, si $\eta \perp \mu$ y $\nu \perp \mu$ entonces $\eta + \nu \perp \mu$:
Sea $X = A \cup B = \tilde{A} \cup \tilde{B}$, donde $A \cap B = \tilde{A} \cap \tilde{B} = \emptyset$, A y $\tilde{A}$ son conjuntos nulos de μ, B es un conjunto nulo de η y $\tilde{B}$ es un conjunto nulo de ν. En consecuencia $X = (A \cup \tilde{A}) \cup (B \cap \tilde{B})$, donde $A \cup \tilde{A}$ es un conjunto nulo de μ y $B \cap \tilde{B}$ es un conjunto nulo de η y ν y por ende de $\eta + \nu$.

Para integrar con respecto ν definimos $L^1(\nu) = L^1(\nu^+) \cap L^1(\nu^-)$ y para $f \in L^1(\nu)$,

$$\int f d\nu := \int f d\nu^+ - \int f d\nu^-.$$

Nótese que por definición también tenemos $L^1(\nu) = L^1(|\nu|)$. Finalmente, notamos que $E \in \Sigma$ es un conjunto de medida finita para ν si y sólo si $|\nu|(E) < \infty$. De esta manera tenemos que ν es σ-finita (ó finita) si y sólo si $|\nu|$ es σ-finita (ó finita) si y sólo si ν^+ y ν^- son σ-finitas (ó finitas).

3.3. Teorema de Lebesgue-Radon-Nikodym.

Definición 3.15. Sean ν una medida con signo y μ una medida positiva en (X, Σ). Decimos que ν es **absolutamente continua** con respecto μ si y sólo si para todo $E \in \Sigma$ tal que $\mu(E) = 0$ tenemos que E es un conjunto nulo de ν. Esto lo denotamos por $\nu \ll \mu$.

Dado que $E \in \Sigma$ es un conjunto nulo de ν si y sólo si $\nu^+(E) = \nu^-(E) = 0$ si y sólo si $|\nu|(E) = 0$, se sigue que $\nu \ll \mu$ si y sólo si $\nu^+ \ll \mu$ y $\nu^- \ll \mu$ si y sólo $|\nu| \ll \mu$.
Por otro lado, si $\nu \ll \mu$ y $\nu \perp \mu$, entonces $\nu = 0$. Para ver esto sea $X = A \cup B$, donde $|\nu|(A) = \mu(B) = 0$. Se sigue que $|\nu|(X) = 0$ y obtenemos el resultado.

El siguiente resultado hará evidente de donde proviene el termino de continuidad absoluta:

Proposición 3.16. *Sean ν una medida con signo finita y μ una medida positiva en (X, Σ). Entonces $\nu \ll \mu$ si y sólo si para todo $\epsilon > 0$ existe $\delta > 0$ tal que si $E \in \Sigma$ y $\mu(E) < \delta$ entonces $|\nu(E)| < \epsilon$.*

Demostración. Dado que $\nu \ll \mu$ si y sólo si $|\nu| \ll \mu$ y $|\nu(E)| \leq |\nu|(E)$, podemos suponer sin pérdida de generalidad que ν es positiva. Supongamos que se cumple la propiedad de ϵ-δ del enunciado. Por hipótesis, para todo $k \in \mathbb{N}$ existe $\delta_k > 0$ tal que si $\mu(E) < \delta_k$ entonces $\nu(E) < 1/k$. Si $\mu(E) = 0$ se

sigue que $\nu(E) < 1/k$ para todo $k \in \mathbb{N}$; es decir, $\nu(E) = 0$.
Ahora supongamos que $\nu \ll \mu$. Si no se cumpliera la propiedad de ϵ-δ enunciada, entonces existiría $\epsilon > 0$ tal que para todo $k \in \mathbb{N}$ existe $E_k \in \Sigma$ con $\mu(E_k) < 2^{-k}$ pero $\nu(E_k) \geq \epsilon$. Dado que $\sum_{k\in\mathbb{N}} \mu(E_k) < \infty$, por el lema de Borel-Cantelli tenemos que $\mu(\limsup_k E_k) = 0$. Sin embargo para todo $k \in \mathbb{N}$, $\nu\big(\bigcup_{m\geq k} E_m\big) \geq \epsilon$ y como ν es finita, $\nu(\limsup_k E_k) = \lim_{k\to\infty} \nu\big(\bigcup_{m\geq k} E_m\big) \geq \epsilon$. Esto contradice $\nu \ll \mu$ y podemos concluir. □

Si μ es una media positiva en (X, Σ) y $f \in L^1(X, \Sigma, \mu)$, la medida con signo $\nu(E) = \int_E f d\mu$ claramente es absolutamente continua con respecto μ y además es finita. Por lo tanto tenemos el siguiente corolario:

Corolario 3.17. *Si $f \in L^1(X, \Sigma, \mu)$, dado $\epsilon > 0$ existe $\delta > 0$ tal que si $\mu(E) < \delta$, entonces $|\int_E f d\mu| < \epsilon$.*

De ahora en adelante, para denotar a que $\nu(E) = \int_E f d\mu$ utilizaremos la expresión $d\nu = f d\mu$.

Lema 3.18. *Sean ν y μ medidas positivas y finitas en (X, Σ). Si ν y μ no son singulares, entonces existen $\epsilon > 0$ y $E \in \Sigma$ tales que $\mu(E) > 0$ y $\nu \geq \epsilon\mu$ en E; es decir, E es un conjunto positivo de la medida $\nu - \epsilon\mu$.*

Demostración. Sea $X = P_n \cup N_n$ una descomposición de Hahn para la medida con signo $\nu - n^{-1}\mu$. Luego definimos $P = \bigcup_{n\in\mathbb{N}} P_n$ y $N = \bigcap_{n\in\mathbb{N}} N_n$. Claramente N es un conjunto negativo de $\nu - n^{-1}\mu$ para toda $n \in \mathbb{N}$. En consecuencia tenemos que $\nu(N) \leq n^{-1}\mu(N)$ para todo $n \in \mathbb{N}$; es decir $\nu(N) = 0$. En consecuencia no podemos tener $\mu(P) = 0$ pues por definición tendríamos $\nu \perp \mu$. Entonces $\mu(P) > 0$ lo que a su vez implica $\mu(P_n) > 0$ para algún $n \in \mathbb{N}$. Por lo tanto $E = P_n$ y $\epsilon = n^{-1}$ satisfacen el enunciado. □

Teorema 3.19. *(Lebesgue-Radon-Nikodym) Sean ν una medida con signo y μ una medida positiva en (X, Σ), ambas σ-finitas. Entonces existen una única medida σ-finita η en (X, Σ) y una función medible $f : X \to \mathbb{R}$ tales $\eta \perp \mu$, $f d\mu$ es σ-finita y $d\nu = d\eta + f d\mu$. Si $\tilde{f}$ es otra función que cumple esto, se tendrá que $f = \tilde{f}$ μ-c.d.s.*

Demostración. Primero probaremos el caso en el que ν y μ son finitas y positivas. Sea

$$\mathcal{F} = \left\{ f : X \to [0, \infty] : \int_E f d\mu \leq \nu(E) \text{ para todo } E \in \Sigma \right\}.$$

$\mathcal{F}$ no es vacío ya que $0 \in \mathcal{F}$. Por otro lado notemos que si $f, g \in \mathcal{F}$ entonces $h = \max(f, g) \in \mathcal{F}$. Para ver esto consideremos $A = \{x : f(x) > g(x)\} \in \Sigma$. Entonces para todo $E \in \Sigma$,

$$\int_E h\, d\mu = \int_{E\cap A} f d\mu + \int_{E\cap A^c} g\, d\mu \leq \nu(E \cap A) + \nu(E \cap A^c) = \nu(E).$$

Sea $\alpha := \sup\big\{ \int f d\mu : f \in \mathcal{F} \big\}$ y nótese que $\alpha \leq \nu(X) < \infty$. Tomemos $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ tal que $\int f_n d\mu \to \alpha$. Podemos suponer que f_n es una sucesión no decreciente de funciones ya que de no ser así simplemente reemplazamos f_n por $g_n = \max\{f_1, \ldots, f_n\} \in \mathcal{F}$ y como $\int g_n d\mu \geq \int f_n d\mu$ se sigue que $\int g_n d\mu \to \alpha$. De esta manera $f = \lim_{n\to\infty} f_n$ existe y por convergencia monótona tenemos que para todo $E \in \Sigma$,

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu \leq \nu(E).$$

Por lo tanto $f \in \mathcal{F}$ y tomando $E = X$ en la última desigualdad vemos que $\int f d\mu = \alpha < \infty$. Por otro lado, dado que f es integrable, $f < \infty$ μ-c.d.s. y podemos asumir que no toma el valor ∞. Definamos $d\eta = d\nu - f d\mu$. Claramente η es finita y además positiva puesto que $f \in \mathcal{F}$. Si η y ν no fueran singulares, por el lema 3.18 existiría $\epsilon > 0$ y $E \in \Sigma$ tal que $\mu(E) > 0$ y $\eta \geq \epsilon\mu$ en E. Esto implicaría que $\epsilon \mathbb{1}_E \, d\mu \leq d\eta$. Por definición de η y dado que todas las medidas son finitas tendríamos que $(f + \epsilon \mathbb{1}_E) \, d\mu \leq d\nu$. En consecuencia $f + \epsilon \mathbb{1}_E \in \mathcal{F}$ y esto es una contradicción ya que $\int (f + \epsilon \mathbb{1}_E) \, d\mu > \alpha$. Por lo tanto $\eta \perp \mu$ y por construcción $d\nu = d\eta + f d\mu$.

Supongamos que $\tilde{\eta}$ y $\tilde{f}$ también satisfacen el enunciado. Tenemos que $d\tilde{\eta} - d\eta = (f - \tilde{f})d\mu$. Dado que $\eta \perp \mu$ y $\tilde{\eta} \perp \mu$ tenemos que $\tilde{\eta} - \eta \perp \mu$. Como también tenemos $\tilde{\eta} - \eta \ll \mu$ podemos concluir que $\eta = \tilde{\eta}$ ($\tilde{\eta} - \eta = 0$). Esto a su vez implica que $(f - \tilde{f})d\mu = 0$ de donde se sigue fácilmente que $f = \tilde{f}$ μ-c.d.s.

Ahora supongamos que ν y μ son positivas y σ-finitas por lo que existen sucesiones de conjuntos disjuntos y de medida finita respectivas para cada medida. Tomando todas las posibles intersecciones podemos llegar a una sucesión de conjuntos disjuntos $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ tales que $\nu(E_n)$ y $\mu(E_n)$ son finitos para cada $n \in \mathbb{N}$ y $X = \bigcup_{n\in\mathbb{N}} E_n$. Luego, las restricciones $\nu_{E_n} = \nu|_{\Sigma\cap E_n}$ y $\mu_{E_n} = \mu|_{\Sigma\cap E_n}$ son medidas finitas y positivas. Por el caso que ya probamos existen medidas finitas η_n en $(E_n, \Sigma \cap E_n)$ y funciones medibles e integrables $f_n : E_n \to [0, \infty)$ tales que $d\nu_{E_n} = \eta_n + f_n d\mu_{E_n}$ y $\eta_n \perp \mu_{E_n}$. Si para todo $E \in \Sigma$ definimos $\eta(E) = \sum_{n\in\mathbb{N}} \eta_n(E \cap E_n)$ y para $x \in E_n$ $f(x) = f_n(x)$, se sigue que η es una medida σ-finita tal que $\eta \perp \mu$, f es una función medible tal que $f d\mu$ es σ-finita y $d\nu = d\eta + f d\mu$. Para la parte de unicidad supongamos que $d\nu = d\tilde{\eta} + \tilde{f} d\mu$ es una descomposición que satisface el enunciado. Al restringirnos a E_n tenemos $d\nu_{E_n} = d\tilde{\eta}_{E_n} + \tilde{f}|_{E_n} d\mu_{E_n}$ y como ya probamos unicidad para el caso finito, concluimos que $\tilde{\eta}_{E_n} = \eta_n$ y $\tilde{f}|_{E_n} = f_n$ μ-c.d.s. De esto se sigue fácilmente que $\tilde{\eta} = \eta$ y $\tilde{f} = f$ μ-c.d.s.

Finalmente, si ν es una medida con signo y σ-finita, sea $\nu = \nu^+ - \nu^-$ la descomposición de Jordan. Por el caso anterior podemos encontrar descomposiciones para ν^+ y ν^-, respectivamente, y sustrayendo estas descomposiciones obtenemos una descomposición para ν que satisface el enunciado. Mientras que la unicidad se prueba de manera análoga al caso anterior. $\square$

La representación (única) $d\nu = d\eta + f d\mu$ con $\eta \perp \mu$ es llamada la **descomposición de Lebesgue** de ν con respecto μ. Por otro lado, si tuviéramos $\nu \ll \mu$, también tendríamos $\eta \ll \mu$. De manera que $\eta = 0$ y $d\nu = f d\mu$. Bajo este supuesto el teorema 3.19 es llamado el **Teorema de Radon-Nikodym** y la función f es llamada la **derivada de Radon-Nikodym** de ν con respecto μ. Esta última suele denotarse por $d\nu/d\mu$ y es importante recalcar que es única módulo una clase de equivalencia de funciones que coinciden μ-c.d.s.

Ahora, por unicidad de la descomposición de Jordan tenemos que $d\nu^+ = (d\nu/d\mu)^+ d\mu$ y $d\nu^- = (d\nu/d\mu)^- d\mu$. Por la proposición 2.31 podemos ver que para toda $g \in L^1(\nu)$,

$$\int g d\nu = \int g d\nu^+ - \int g d\nu^- = \int g \frac{d\nu}{d\mu}^+ d\mu - \int g \frac{d\nu}{d\mu}^- d\mu = \int g \frac{d\nu}{d\mu} d\mu.$$

Con los supuestos anteriores, si ρ es una medida positiva tal que $\mu \ll \rho$ entonces tendremos $\nu \ll \rho$.

Por lo tanto $d\nu/d\rho$ existe y la notación sugiere que

$$\frac{d\nu}{d\rho} = \frac{d\nu}{d\mu}\frac{d\mu}{d\rho}.$$

Para ver esto consideremos $E \in \Sigma$. Nuevamente por la proposición 2.31,

$$\begin{aligned}\nu(E) = \int_E \frac{d\nu}{d\mu}\, d\mu &= \int \mathbb{1}_E \frac{d\nu}{d\mu}^{+} d\mu - \int \mathbb{1}_E \frac{d\nu}{d\mu}^{-} d\mu \\ &= \int \mathbb{1}_E \frac{d\nu}{d\mu}^{+} \frac{d\mu}{d\rho}\, d\rho - \int \mathbb{1}_E \frac{d\nu}{d\mu}^{-} \frac{d\mu}{d\rho}\, d\rho = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\rho}\, d\rho.\end{aligned}$$

Como un corolario inmediato tenemos que si ν y μ son medidas positivas no triviales tales $\nu \ll \mu$ y $\mu \ll \nu$, entonces $(d\nu/d\mu)(d\mu/d\nu) = 1$; es decir,

$$\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1} \quad \mu\text{-}c.d.s..$$

3.4. Medidas Complejas

Definición 3.20. Una medida compleja en (X, Σ) es una función $\nu : \Sigma \to \mathbb{C}$ tal que

1) $\nu(\emptyset) = 0$.
2) Si $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ son disjuntos entonces $\nu\big(\bigcup_{n\in\mathbb{N}} E_n\big) = \sum_{n\in\mathbb{N}} \nu(E_n)$, donde la convergencia es absoluta.

Nótese que por definición las medidas complejas son finitas. Denotemos por ν_{Re} y ν_{Im} a las partes real e imaginaria de ν, respectivamente. Es casi inmediato que ν_{Re} y ν_{Im} son medidas con signo finitas. De esta manera resultará muy sencillo generalizar los conceptos y resultados que ya hemos obtenido para medidas con signo.

El primer resultado que podemos generalizar es la proposición 3.9; es decir, las propiedades de continuidad de medidas. Por otro lado, definimos $L^1(\nu) = L^1(\nu_{Re}) \cap L^1(\nu_{Im})$ y para $f \in L^1(\nu)$,

$$\int f d\nu := \int f d\nu_{Re} + i \int f d\nu_{Im}.$$

Ahora, diremos que $E \in \Sigma$ es un conjunto nulo de ν si y sólo si para todos subconjunto medible $A \subset E$, $\nu(A) = 0$. Evidentemente E es un conjunto nulo de ν si y sólo si E es un conjunto nulo de ν_{Re} y de ν_{Im}. Si extendemos las definiciones de singularidad y continuidad absoluta a medidas complejas, de lo anterior es fácil ver que si μ es una medida positiva en (X, Σ), entonces:

1) $\nu \perp \mu$ si y sólo si $\nu_{Re} \perp \mu$ y $\nu_{Im} \perp \mu$.
2) $\nu \ll \mu$ si y sólo si $\nu_{Re} \ll \mu$ y $\nu_{Im} \ll \mu$.

Por lo tanto, si aplicamos el teorema de Lebesgue-Radon-Nikodym a la parte real e imaginaria de ν podemos llegar fácilmente a la siguinete generalización:

Teorema 3.21. *Sean ν una medida medida compleja y μ una medida σ-finita y positiva en (X, Σ). Entonces existen una única medida compleja η en (X, Σ) y $f \in L^1(\mu)$ tales que $\eta \perp \mu$ y $d\nu = d\eta + f d\mu$. Si $\tilde{f}$ es otra función que cumple esto, se tendrá que $f = \tilde{f}$ μ-c.d.s.*

En el caso que $\nu \ll \mu$, al igual que en el caso de medidas con signo, tendremos $\eta = 0$. La clase de equivalencia de f, sigue siendo llamada la derivada de Radon-Nikodym de ν con respecto μ y también la denotamos por $d\nu/d\mu$. Por unicidad podemos comprobar que

$$\frac{d\nu}{d\mu} = \frac{d\nu_{Re}}{d\mu} + i\,\frac{d\nu_{Im}}{d\mu}.$$

Así, por los resultados que ya probamos para medidas con signo se sigue que:

1) Para todo $g \in L^1(\nu)$, $\int g d\nu = \int g \frac{d\nu}{d\mu} d\mu$.

2) Si $\mu \ll \rho$ entonces $d\nu/d\rho = (d\nu/d\mu)(d\mu/d\rho)$.

Para definir la variación total de ν consideremos cualquier medida positiva μ tal que $\nu \ll \mu$, por ejemplo $\mu = |\nu_{Re}| + |\nu_{Im}|$. Definimos $d|\nu| = |d\nu/d\mu|\, d\mu$. Veamos que $|\nu|$ está bien definida: Supongamos que μ_1 y μ_2 son medidas positivas tales que $\nu \ll \mu_1$ y $\nu \ll \mu_2$. Sea $\rho = \mu_1 + \mu_2$. Entonces,

$$\frac{d\nu}{d\mu_1}\frac{d\mu_1}{d\rho} = \frac{d\nu}{d\rho} = \frac{d\nu}{d\mu_2}\frac{d\mu_2}{d\rho}.$$

Como $d\mu_1/d\rho$ y $d\mu_2/d\rho$ deben ser positivas ρ-c.d.s., $|d\nu/d\mu_1|(d\mu_1/d\rho) = |d\nu/d\mu_2|(d\mu_2/d\rho)$ ρ-c.d.s. y se sigue que

$$\left|\frac{d\nu}{d\mu_1}\right| d\mu_1 = \left|\frac{d\nu}{d\mu_1}\right| \frac{d\mu_1}{d\rho} d\rho = \left|\frac{d\nu}{d\mu_2}\right| \frac{d\mu_2}{d\rho} d\rho = \left|\frac{d\nu}{d\mu_2}\right| d\mu_2;$$

es decir, $|\nu|$ está bien definida. Nótese que esta definición es consistente con la de medidas con signo ya que si ν fuera real entonces $\nu \ll \nu^+ + \nu^-$. Si $X = P \cup N$ es una descomposición de Hahn tenemos que $d\nu = (\mathbb{1}_P - \mathbb{1}_N)d(\nu^+ + \nu^-)$. Por lo tanto $d|\nu| = |\mathbb{1}_P - \mathbb{1}_N| d(\nu^+ + \nu^-) = d\nu^+ + d\nu^-$ que es precisamente como habíamos definido $|\nu|$.

Proposición 3.22. *Si ν y μ son medidas complejas en (X, Σ), entonces $|\nu + \mu| \leq |\nu| + |\mu|$.*

Demostración. Sea η una medida positiva tal que $\nu \ll \eta$ y $\mu \ll \eta$, por ejemplo $\eta = |\nu_{Re}| + |\nu_{Im}| + |\mu_{Re}| + |\mu_{Im}|$. Entonces $\nu + \mu \ll \eta$ y $d(\nu + \mu)/d\eta = d\nu/d\eta + d\mu/d\eta$. Por lo tanto,

$$d|\nu + \mu| = |d(\nu + \mu)/d\eta|\, d\eta \leq |d\nu/d\eta|\, d\eta + |d\mu/d\eta|\, d\eta = d|\mu| + d|\nu|.$$ □

Proposición 3.23. *Sea ν una medida compleja en (X, Σ). Entonces*

1) $|\nu(E)| \leq |\nu|(E)$ para todo $E \in \Sigma$.

2) $\nu \ll |\nu|$ y $\big|d\nu/d|\nu|\big| = 1$ $|\nu|$-c.d.s.

3) $|\nu_{Re}|, |\nu_{Im}| \leq |\nu| \leq |\nu_{Re}| + |\nu_{Im}|$, $L^1(\nu) = L^1(|\nu|)$ y si $f \in L^1(\nu)$, entonces

$$\left|\int f d\nu\right| \leq \int |f| d|\nu|.$$

Demostración. 1) Sea $\mu = |\nu_{Re}| + |\nu_{Im}|$. Dado que $d|\nu| = |d\nu/d\mu|\, d\mu$, tenemos que para todo $E \in \Sigma$,

$$|\nu(E)| = \left| \int_E \frac{d\nu}{d\mu} d\mu \right| \leq \int_E \left| \frac{d\nu}{d\mu} \right| d\mu = |\nu|(E).$$

2) $\nu \ll |\nu|$ se sigue del inciso anterior. Ahora, sabemos que $d\nu/d\mu = (d\nu/d|\nu|)(d|\nu|/d\mu)$. Pero por definición de $|\nu|$ esto es $d\nu/d\mu = (d\nu/d|\nu|)\,|d\nu/d\mu|$. Por lo tanto $\big|d\nu/d|\nu|\big| = 1$ en $A = \{x \in X : d\nu/d\mu(x) \neq 0\}$. Claramente A^c es un conjunto nulo de ν y por lo tanto de $\mu = |\nu_{Re}| + |\nu_{Im}|$. Como $|\nu| \ll \mu$ podemos concluir.

3) La desigualdad $|\nu| \leq |\nu_{Re}| + |\nu_{Im}|$ es consecuencia inmediata de la proposición 3.22. Para la otra desigualdad, dado que

$$\frac{d\nu}{d|\nu|} = \frac{d\nu_{Re}}{d|\nu|} + i\,\frac{d\nu_{Im}}{d|\nu|},$$

del inciso anterior tenemos $|d\nu_{Re}/d|\nu| \leq |d\nu/d|\nu|| = 1$ $|\nu|$-c.d.s. Por lo tanto,

$$|\nu_{Re}|(E) = \int_E \left| \frac{d\nu_{Re}}{d|\nu|} \right| d|\nu| \leq |\nu|(E).$$

Análogamente tenemos $|\nu_{Im}| \leq |\nu|$. De las desigualdades obtenidas se sigue fácilmente que $L^1(|\nu|) = L^1(\nu)$ y si $f \in L^1(\nu)$,

$$\left| \int f d\nu \right| = \left| \int f\, \frac{d\nu}{d|\nu|}\, d|\nu| \right| \leq \int |f| \left| \frac{d\nu}{d|\nu|} \right| d|\nu| = \int |f| d|\nu|. \qquad \square$$

Para finalizar esta sección definamos $M(X,\Sigma)$ como el espacio de medidas complejas en (X,Σ). Claramente éste es un espacio vectorial. Además por la proposición 3.22 podemos ver que $\|\nu\|_{TV} := |\nu|(X)$ define una norma en $M(X,\Sigma)$. Más aún,

$$\sup_{A \in \Sigma} |\nu(A)| \leq \|\nu\|_{TV},$$

de donde se puede deducir que $M(X,\Sigma)$ resulta ser un espacio de Banach. Esta afirmación se deja como ejercicio al lector.

3.5. El dual de L^p

Un funcional acotado en L^p con $p \geq 1$ es una función lineal $\phi : L^p \to \mathbb{C}$ para la cual existe $c > 0$ tal que

$$|\phi(f)| \leq c\|f\|_p,$$

para todo $f \in L^p$. Es inmediato que ϕ es continua con respecto a la norma $\|\cdot\|_p$. De hecho de manera general, un operador lineal sobre espacios normandos es acotado si y sólo si éste es continuo. Adoptando la notación usual, denotaremos al espacio de funcionales acotados en L^p por $(L^p)^*$. Para todo $\phi \in (L^p)^*$ también podemos definir su norma de la manera usual:

$$\|\phi\| = \sup\Big\{ \big|\phi(f)\big| : \|f\|_p = 1 \Big\}.$$

Dado que $|\phi(f)| \leq c$ para todo $f \in L^p$ con $\|f\|_p = 1$, se sigue que $\|\phi\| \leq c$. De esto resulta evidente que $\|\phi\|$ es la mínima cota para la cual se cumple

$$|\phi(f)| \leq \|\phi\| \|f\|_p,$$

para todo $f \in L^p$. Ahora, sea q el exponente conjugado de p. Si $g \in L^q$, para todo $f \in L^p$, $fg \in L^1$ y por lo tanto $\phi_g : L^p \to \mathbb{C}$, dado por

$$\phi_g(f) = \int fg \, d\mu,$$

está bien definido. Además, claramente es lineal y acotado pues de la desigualdad de Hölder tenemos

$$\left|\phi_g(f)\right| \leq \int |fg| \, d\mu \leq \|g\|_q \|f\|_p.$$

En particular tenemos que $\|\phi_g\| \leq \|g\|_q$. El siguiente resultado nos muestra que, bajo ciertas hipótesis, todos los funcionales acotados en L^p de hecho tienen una representación de la forma ϕ_g:

Teorema 3.24. *Sean p y q exponentes conjugados y μ una medida σ-finita en (X, Σ). Si $1 \leq p < \infty$, para todo $\phi \in (L^p)^*$ existe una única $g \in L^q$ tal que $\phi(f) = \int fg \, d\mu$ para todo $f \in L^p$ y $\|\phi\| = \|g\|_q$. Por lo tanto $g \mapsto \phi_g$ define una isometría de L^q a $(L^p)^*$.*

Demostración. Comencemos por ver La parte de unicidad. Si g y $\tilde{g}$ cumplen el enunciado, entonces $\int \mathbb{1}_E(g - \tilde{g}) d\mu = 0$ para todo conjunto E de medida finita. Como μ es σ-finita podemos concluir que $g = \tilde{g}$ μ-c.d.s. de manera que pertenecen a la misma clase de equivalencia en L^q.

Para la parte de existencia primero suponemos que μ es finita. Definimos $\nu : \Sigma \to \mathbb{C}$ a través de la identidad $\nu(E) = \phi(\mathbb{1}_E)$. Veamos que ν es una medida. Dado que ϕ es lineal $\nu(\emptyset) = \phi(\mathbb{1}_\emptyset) = \phi(0) = 0$. Sean $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ disjuntos y $E = \bigcup_{n\in\mathbb{N}} E_n$. Como ϕ es lineal y acotada,

$$\begin{aligned}
\left|\nu(E) - \sum_{i=1}^n \nu(E_i)\right| &= \left|\phi\Big(\mathbb{1}_E - \sum_{i=1}^n \mathbb{1}_{E_i}\Big)\right| \\
&= \left|\phi\Big(\mathbb{1}_{\bigcup_{m>n} E_n}\Big)\right| \leq \|\phi\| \|\mathbb{1}_{\bigcup_{m>n} E_n}\|_p = \|\phi\| \, \mu\Big(\bigcup_{m>n} E_n\Big)^{1/p}.
\end{aligned}$$

Dado que μ es finita, $\mu\big(\bigcup_{m>n} E_n\big) \to 0$ cuando $n \to \infty$ puesto que $\mu\big(\bigcup_{m>n} E_n\big) = \sum_{n>m} \mu(E_n)$ es la cola de una serie convergente. Entonces $\nu(E) = \sum_{i=1}^\infty \nu(E_i)$ y la convergencia no depende del orden de la sucesión $\{E_n\}_{n\in\mathbb{N}}$. Por lo tanto ν es medida.
Por otro lado si $\mu(E) = 0$, entonces $\mathbb{1}_E = 0$ μ-c.d.s. y por lo tanto $\nu(E) = \phi(\mathbb{1}_E) = 0$; es decir, $\nu \ll \mu$. Por el teorema de Radon-Nikodym existe $g \in L^1$ tal que $d\nu = g d\mu$.
Ahora, por linealidad podemos ver que para toda función simple $\phi(f) = \int f d\nu = \int fg d\mu$. Si $f \in L^\infty$ podemos tomar una sucesión de funciones simples tales que $f_n \to f$ y $|f_n| \leq |f|$. Por convergencia dominada tendremos que $f_n \to^{L^p} f$ y $f_n g \to^{L^1} fg$. Luego, por continuidad de ϕ se sigue que

$$\phi(f) = \lim_{n\to\infty} \phi(f_n) = \lim_{n\to\infty} \int f_n g d\mu = \int fg d\mu.$$

Utilicemos esto para probar que $g \in L^q$. Primero consideramos el caso $p > 1$:
Consideremos $E_n = \{x : |g(x)| \leq n\}$ y la función medible $h : \mathbb{C} \to \mathbb{C}$ dada por

$$h(z) = \begin{cases} \overline{z}/|z| & z \neq 0, \\ 0 & z = 0. \end{cases}$$

Nótese que para esta función se cumple $|h(z)| = 1$ si $z \neq 0$ y $h(z)z = |z|$. Luego, es claro que $f_n = \mathbb{1}_{E_n}|g|^{q-1}h \circ g$ está en L^∞ y en consecuencia

$$\int_{E_n} |g|^q d\mu = \int f_n g d\mu = \phi(f_n) \leq \|\phi\| \left(\int |f_n|^p d\mu \right)^{1/p} = \|\phi\| \left(\int_{E_n} |g|^q d\mu \right)^{1/p}.$$

Por lo tanto $\int_{E_n} |g|^q d\mu \leq \|\phi\|^q$ para toda $n \in \mathbb{N}$ y por convergencia monótona concluimos que $g \in L^q$ y $\|g\|_q \leq \|\phi\|$.

Para el caso $p = 1$:
Dado $\epsilon > 0$ sea $E = \{x : |g(x)| \geq \|\phi\| + \epsilon\}$. Dado que $f = \mathbb{1}_E \, h \circ g$ está en L^∞,

$$\int_E |g| d\mu = \int f g d\mu = \phi(f) \leq \|\phi\| \int |f| d\mu = \|\phi\| \mu(E).$$

Sin embargo es fácil ver que $(\|\phi\| + \epsilon)\mu(E) \leq \int_E |g| d\mu$ de donde se sigue que $\mu(E) = 0$. Es decir, $|g| \leq \|\phi\| + \epsilon$ μ-c.d.s y como ϵ era arbitrario tenemos que $\|g\|_\infty \leq \|\phi\|$.

Ahora que sabemos que $g \in L^q$, con un argumento análogo al que usamos para funciones en L^∞, podemos ver que $\phi(f) = \int fg \, d\mu$ para toda $f \in L^p$ ya que $fg \in L^1$. Además, como se mencionó previamente, por la desigualdad de Hölder $\|\phi\| \leq \|g\|_q$. Esto establece el resultado cuando μ es finita.

Para el caso en que μ es σ-finita nótese que para todo conjunto E de medida finita, la restricción $\mu_E = \mu|_{\Sigma \cap E}$ es una medida finita. Luego, para toda $f \in L^p(E)$ podemos tomar una extensión $\tilde{f} \in L^p$ que coincida con f en E y valga cero en E^c. Así, $\phi_E(f) = \phi(\tilde{f})$ define un funcional acotado en $L^p(E)$. Más aún, si $h \in L^p$, para la restricción $h|_E$ tenemos $\phi_E(h|_E) = \phi(\mathbb{1}_E h)$ y claramente $\|\phi_E\| \leq \|\phi\|$. Sea $\{E_n\} \subset \Sigma$ una sucesión creciente de conjuntos de medida finita y tales que $X = \bigcup_{n \in \mathbb{N}} E_n$. Por el caso que ya probamos existe $g_n \in L^q(E_n)$ tal que $\phi_{E_n}(f) = \int f g_n d\mu_{E_n}$ para todo $f \in L^p(E_n)$ y $\|\phi_{E_n}\| = \|g_n\|_q$. Denotamos por $\tilde{g}_n$ a la extensión de g_n a todo X al definirla como 0 en E_n^c. Nótese que si $n > m$ para $f \in L^p(E_m)$,

$$\int f g_n|_{E_m} d\mu_{E_m} = \int_{E_m} \tilde{f} \tilde{g}_n d\mu = \phi_{E_n}(\tilde{f}|_{E_n}) = \phi(\mathbb{1}_{E_n} \tilde{f}) = \phi(\mathbb{1}_{E_m} \tilde{f}) = \phi_{E_m}(f).$$

Por unicidad tenemos que $g_n|_{E_m} = g_m$ c.d.s. y en particular $|\tilde{g}_n| \leq |\tilde{g}_{n+1}|$ c.d.s. Si definimos $g(x) = g_{n+1}(x)$ para $x \in E_{n+1} - E_n$ tendremos $\tilde{g}_n \to g$ c.d.s. y por convergencia monótona,

$$\int |g|^q d\mu = \lim_{n \to \infty} \int |\tilde{g}_n|^q d\mu = \lim_{n \to \infty} \int |g_n|^q d\mu_{E_n} = \lim_{n \to \infty} \|\phi_{E_n}\|^q \leq \|\phi\|^q.$$

Por lo tanto $g \in L^q$ y $\|g\|_q \leq \|\phi\|$. De manera similar, si $f \in L^p$, etnonces $\mathbb{1}_{E_n} f \to^{L^p} f$ y en consecuencia

$$\phi(f) = \lim_{n \to \infty} \phi(\mathbb{1}_{E_n} f) = \lim_{n \to \infty} \int f|_{E_n} g_n \, d\mu_{E_n} = \lim_{n \to \infty} \int_{E_n} f \tilde{g}_n \, d\mu = \int f g \, d\mu.$$

La desigualdad $\|\phi\| \leq \|g\|_q$ se sigue de Hölder y podemos concluir el resultado. $\square$

Para el caso $p = \infty$ claramente para todo $g \in L^1$, $\phi_g \in (L^\infty)^*$. Sin embargo, el reciproco es falso. De hecho, el teorema de representación de Kantorovitch nos dice que $(L^\infty)^*$ es isométrico al espacio vectorial de funciones en Σ que son finitamente aditivas y absolutamente continuas con respecto μ. La demostración de este hecho es muy similar a la del caso $1 \leq p < \infty$.
Otra nota importante es que para $1 < p < \infty$ el teorema 3.24 se puede extender fácilmente al caso en que μ no es σ-finita. Este resultado se deja como ejercicio al lector.

3.6. Convergencia Débil en L^p

En esta sección enunciaremos algunos resultados importantes de análisis funcional que son válidos para espacios L^p. Estos resultados no se probaran pero se pueden consultar en las siguientes referencias:

1) Royden, H. L., & Fitzpatrick, P. (1988). Real analysis (Vol. 32). New York: Macmillan. Capítulo 14.

2) Schechter, M. (2001). Principles of functional analysis (No. 36). American Mathematical Soc. Capítulo 8.

De manera general, dado un espacio normado X podemos definir al espacio dual X^* como los funcionales continuos o acotados en X. Es decir, las funciones lineales continuas ó acotadas que van de X a su respectivo campo, usualmente $\mathbb{R}$ o $\mathbb{C}$. Claramente X^* es un espacio vectorial y para todo $l \in X^*$, el valor finito

$$\|l\| = \sup\left\{|l(x)| : \|x\| = 1\right\}$$

define una norma en X^*. Además si el campo es completo se tiene que X^* equipado con esta norma es un espacio de Banach. En nuestro contexto asumimos que X es un espacio vectorial sobre $\mathbb{C}$.
Nótese que por definición todo elemento en X^* es continuo con respecto a la topología inducida por la norma en X. Sin embargo podemos considerar la mínima topología que hace continuos a los funcionales en X^*. Esta topología es llamada la topología débil y es inmediato que ésta es generada por la subbase de conjuntos $l^{-1}(G)$, donde $l \in X^*$ y G es abierto en $\mathbb{C}$. En particular, tenemos una base de vecindades de $x \in X$ conformada por los conjuntos:

$$U_{\epsilon, l_1, \ldots, l_n}(x) := \left\{y \in X : |l_i(x) - l_i(y)| < \epsilon,\, 1 \leq i \leq n\right\}.$$

Si una sucesión $\{x_n\}_{n\in\mathbb{N}}$ converge a x en la topología débil esto es denotado por $x_n \to^w x$. Por la forma de que tiene una base de vecindades para todo $x \in X$ podemos ver que

$$x_n \to^w x \textit{ si y sólo si } l(x_n) \to l(x) \textit{ para todo } l \in X^*.$$

Claramente la convergencia en norma implica convergencia débil. El reciproco en general es falso. De hecho se puede probar que la topología débil coincide con la topología inducida por la norma si y sólo si X es de dimensión finita.

Teorema 3.25. *Sea X es un espacio vectorial normado. Si $x_n \to^w x$ entonces la sucesión $\{x_n\}_{n\in\mathbb{N}}$ es acotada y*

$$\|x\| \leq \liminf_n \|x_n\|.$$

Dado que X^* es un espacio vectorial normado podemos considerar su dual y lo denotamos por X^{**}. Ahora, dado $x \in X$ podemos definir un elemento $\hat{x} \in X^{**}$ a través de la identidad $\hat{x}(l) = l(x)$. Claramente $\hat{x}$ define una función lineal en X^* y además

$$\|\hat{x}\| = \sup\left\{|\hat{x}(l)| : \|l\| = 1\right\} = \sup\left\{|l(x)| : \|l\| = 1\right\} \leq \|x\|$$

ya que $|l(x)| \leq \|l\|\|x\|$. De hecho se tiene la igualdad $\|\hat{x}\| = \|x\|$.
De esta manera $J : X \to X^{**}$ dado por $J(x) = \hat{x}$ define una isometría de X a su imagen.

Definición 3.26. Decimos que X es reflexivo si y sólo si $J : X \to X^{**}$ es sobreyectiva.

Teorema 3.27. *Si X es un espacio de Banach reflexivo, entonces toda sucesión acotada tiene una subsucesión débilmente convergente.*

En virtud del teorema 3.24 podemos describir de manera explícita al dual de L^p cuando $1 \leq p < \infty$ y μ es σ-finita. Es decir, los funcionales $\phi_g(f) = \int fg\, d\mu$, donde $g \in L^q$ y q es el exponente conjugado de p. De esta manera, dados $f_n, f \in L^p$, tenemos que

$$f_n \to^w f \textit{ si y sólo si } \int f_n g\, d\mu \to \int fg\, d\mu \textit{ para todo } g \in L^q.$$

El teorema 3.25 nos dice que en este caso $\sup_{n\in\mathbb{N}} \|f_n\|_p < \infty$ y $\|f\|_p \leq \liminf_n \|f_n\|_p$.
Ahora veamos que L^p es un espacio reflexivo para $1 < p < \infty$. Si $\Lambda \in (L^p)^{**}$, definimos $\Lambda' : L^q \to \mathbb{C}$ dado por

$$\Lambda'(g) = \Lambda(\phi_g).$$

entonces $\Lambda' \in (L^q)^*$ ya que $|\Lambda'(g)| = |\Lambda(\phi_g)| \leq \|\Lambda\|\|\phi_g\| = \|\Lambda\|\|g\|_q$. Por el teorema 3.24 existe $f \in L^p$ tal que para todo $\phi_g \in (L^p)^*$,

$$\Lambda(\phi_g) = \Lambda'(g) = \int fg d\mu = \phi_g(f) = J(f)(\phi_g).$$

De esta forma el teorema 3.27 nos asegura que si $\{f_n\}_{n\in\mathbb{N}} \subset L^p$ cumple que $\sup_{n\in\mathbb{N}} \|f_n\|_p < \infty$, entonces existe $f \in L^p$ y una subsucesión tal que

$$\int f_{n_k} g\, d\mu \to \int fg\, d\mu \textit{ para todo } g \in L^q.$$

Capítulo 4

Diferenciación

4.1. Teorema de Cubierta de Besicovitch

En esta sección probaremos el Teorema de Cubierta de Besicovitch que será clave para poder estudiar la diferenciación de cierto tipo de medidas en $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Antes probaremos una versión más débil y en seguida probaremos el Teorema de Cubierta de Besicovitch.

Lema 4.1. *Sean $a, b \in \mathbb{R}^n$ tales que $0 < |a| \leq |a-b|$ y $0 < |b| \leq |a-b|$. Entonces,*

$$\big|a/|a| - b/|b|\big| \geq 1.$$

Demostración. El ángulo formado por los vectores a y b claramente es el mismo que forman $a/|a|$ y $b/|b|$. Por la ley del coseno tenemos la igualdad

$$\frac{2 - \big|a/|a| - b/|b|\big|^2}{2} = \frac{|a|^2 + |b|^2 - |a-b|^2}{2|a||b|}.$$

Como $\max(|a|, |b|) \leq |a-b|$ tenemos $|a|^2 + |b|^2 - |a-b|^2 \leq \min(|a|^2, |b|^2)$. En consecuencia

$$\frac{2 - \big|a/|a| - b/|b|\big|^2}{2} \leq \frac{\min(|a|^2, |b|^2)}{2|a||b|} \leq \frac{1}{2},$$

de donde se sigue fácilmente la desigualdad que buscamos. □

Lema 4.2. *Dado $n \in \mathbb{N}$ existe un entero $N(n)$ que cumple lo siguiente:*
Si para $a_1, \ldots, a_k \in \mathbb{R}^n$ y reales positivos $r_1, \ldots, r_k$ se tiene

$$a_i \notin \overline{B}_{r_j}(a_j) \text{ para } i \neq j \text{ y } \bigcap_{i=1}^{k} \overline{B}_{r_i}(a_i) \neq \emptyset,$$

entonces $k \leq N(n)$.

Demostración. Sin perdida de generalidad podemos suponer que $a_i \neq 0$ para $1 \leq i \leq k$ y $0 \in \bigcap_{i=1}^{k} \overline{B}_{r_i}(a_i)$. De no ser así tomamos $b \in \bigcap_{i=1}^{k} \overline{B}_{r_i}(a_i)$ y trasladamos por $-b$. Dado que $a_j \notin \overline{B}_{r_i}(a_i)$ para $j \neq i$ y $0 \in \overline{B}_{r_i}(a_i)$,

$$|a_i| \leq r_i < |a_i - a_j| \text{ para } i \neq j.$$

Por el lema 4.1 tenemos que $\big|a_i/|a_i| - a_j/|a_j|\big| \geq 1$ para $i \neq j$. Ahora, dado que la esfera unitaria S^{n-1} es compacta existe una cantidad finita de bolas abiertas de radio $1/2$ que la cubren. Sea $N(n)$ esta cantidad. Por lo anterior tenemos que dos puntos $a_i/|a_i|$ y $a_j/|a_j|$ no pueden estar en la misma bola abierta de radio $1/2$. Por lo tanto $k \leq N(n)$. □

Teorema 4.3. *(Cubierta de Besicovitch) Dado $n \in \mathbb{N}$ existe un entero $K(n)$ que cumple lo siguiente: Sea A es un conjunto acotado de $\mathbb{R}^n$ y $\mathcal{F}$ una colección de bolas cerradas con radios uniformemente acotados y tales que todo punto en A es el centro de alguna bola en $\mathcal{F}$. entonces existen $\mathcal{F}_1, \ldots, \mathcal{F}_{K(n)} \subset \mathcal{F}$ tales que para $1 \leq i \leq K(n)$, $\mathcal{F}_i$ es una colección a lo más numerable de bolas disjuntas y*

$$A \subset \bigcup_{i=1}^{K(n)} \bigcup \mathcal{F}_i.$$

Demostración. Para cada $x \in A$ tomemos $r(x)$ tal que $\overline{B}_{r(x)}(x) \in \mathcal{F}$. Si $M_1 = \sup_{x \in A} r(x)$, podemos tomar $x_1 \in A$ tal que $r(x_1) \geq M_1/2$. De manera inductiva podemos tomar $x_k \in A - \bigcup_{i=1}^{k-1} \overline{B}_{r(x_i)}(x_i)$ tal que $r(x_k) \geq M_1/2$. Dado que A es acotado existen dos posibilidades. La primera es que bajo este procedimiento obtengamos una cubierta finita de A. La segunda es que en algún paso no exista un punto que cumpla lo que buscamos. En cualquier caso dicho procedimiento termina en una cantidad finita de pasos obteniendo una sucesión finita $x_1, \ldots, x_{k_1}$.
Ahora sea $M_2 = \sup \left\{ r(x) : x \in A - \bigcup_{i=1}^{k_1} \overline{B}_{r(x_i)}(x_i) \right\}$. Como en el caso anterior, tendremos una cantidad finita $x_{k_1+1}, x_{k_1+2}, \cdots, x_{k_2}$ tales que para $k_1 \leq j < k_2$,

$$x_{j+1} \in A - \bigcup_{i=1}^{j} \overline{B}_{r(x_i)}(x_i) \ \textit{con}\ r(x_{j+1}) \geq M_2/2.$$

Podemos seguir este procedimiento y obtener sucesiones (posiblemente finitas) de enteros $0 < k_i < k_{i+1}$, reales positivos $M_i \geq 2M_{i+1}$ y bolas cerradas $\overline{B}_{r_j}(x_j) \in \mathcal{F}$ con las siguientes propiedades:
Si $I_j := \{k_{j-1}+1, \ldots, k_{j+1}\}$ entonces,

1) $M_j/2 \leq r_i \leq M_j$ para $i \in I_j$.
2) $x_{j+1} \in A - \bigcup_{i=1}^{j} \overline{B}_{r_i}(x_i)$.
3) Si $m \neq k$ para todo $i \in I_m$ y $j \in I_k$, $x_j \notin \overline{B}_{r_i}(x_i)$.

Las primeras dos propiedades son inmediatas de la construcción. Para la tercera, si $m < k$ y $j \in I_k$ por 2) tenemos $x_j \notin \bigcup_{i=1}^{j-1} \overline{B}_{r_i}(x_i)$ y en particular $x_j \notin \overline{B}_{r_i}(x_i)$ para todo $i \in I_m$. Si $m > k$, de manera similar, para $i \in I_m$ tenemos $x_i \notin \overline{B}_{r_j}(x_j)$. Como $r_i \leq M_m \leq M_k/2 \leq r_j$ se sigue que $x_j \notin \overline{B}_{r_i}(x_i)$ para todo $i \in I_m$.

Además tendremos $A \subset \bigcup_j \bigcup_{i \in I_j} \overline{B}_{r_i}(x_i)$. Si sólo hay una cantidad finita de bloques I_j no hay nada que demostrar. Por el contrario tendremos que $M_i \to 0$. Por lo que para todo $x \in A$ existe $j \in \mathbb{N}$ tal que $M_j/2 \leq r(x)$. Por construcción se sigue que $x \in \bigcup_{i=1}^{k_j} \overline{B}_{r_i}(x_i)$.

Sea $N(n)$ el valor entero dado por el lema 4.2. Veremos que si

$$x \in \bigcap_{i=1}^{l} \overline{B}_{r_{m_i}}(x_{m_i}),$$

entonces $l \leq P(n) = 16^n N(n)$. Por 3) y el lema 4.2 sabemos que los indices $m_1, \ldots, m_l$ pueden pertenecer a lo más a $N(n)$ bloques I_k distintos. Así que solo hay que probar que para todo k,

$$\left| I_k \cap \{m_1, \ldots, m_l\} \right| \leq 16^n.$$

Supongamos que $I_k \cap \{m_1, \ldots, m_l\} = \{p_1, \ldots, p_q\}$. Por 1) y 2) sabemos que si $p_j < p_i$,

$$|x_{p_i} - x_{p_j}| > r_{p_j} \geq M_k/2.$$

Como $M_k/2 \leq r_{p_i} \leq M_k$, las bolas $\overline{B}_{\frac{1}{4}r_{p_i}}(x_{p_i})$ son disjuntas y $\overline{B}_{\frac{1}{8}M_k}(x_{p_i}) \subset \overline{B}_{\frac{1}{4}r_{p_i}}(x_{p_i}) \subset \overline{B}_{2M_k}(x)$ para todo p_i: si $y \in \overline{B}_{\frac{1}{4}r_{p_i}}(x_{p_i})$, para $j \neq i$,

$$|x_{p_j} - y| \geq |x_{p_j} - x_{p_i}| - |x_{p_i} - y| > M_k/2 - r_{p_i}/4 \geq M_k/2 - M_k/4 = M_k/4$$

y

$$|x - y| \leq |x - x_{p_i}| + |x_{p_i} - y| \leq r_{p_i} + r_{p_i}/4 = 2M_k.$$

Sea λ la medida de Lebesgue en $\mathbb{R}^n$ y denotemos por $v_n = \lambda\big(B_1(0)\big)$. Entonces,

$$q(M_k/8)^n v_n = \sum_{i=1}^{q} \lambda\big(\overline{B}_{\frac{1}{8}M_k}(x_{p_i})\big) \leq \lambda\big(\overline{B}_{2M_k}(x)\big) = (2M_k)^n v_n,$$

de donde se sigue $q \leq 16^n$.

Ahora, sin mayor pérdida de generalidad podemos suponer que $r_{i+1} \leq r_i$. Sea $B_{1,1} = \overline{B}_{r_1}(x_1)$, y de manera inductiva tomemos $B_{1,k+1} = \overline{B}_{r_{i_{k+1}}}(x_{i_{k+1}})$ donde i_{k+1} es el primer entero tal que

$$\overline{B}_{r_{i_{k+1}}}(x_{i_{k+1}}) \cap \bigcup_{i=1}^{k} B_{1,i} = \emptyset.$$

De esta manera obtenemos una familia, $\mathcal{F}_1 = \{B_{1,1}, B_{1,2}, \ldots\}$, de bolas disjuntas y a lo más numerable. Si $\mathcal{F}_1$ no cubre a A, sea $B_{2,1}$ la primer bola en la sucesión que no está en $\mathcal{F}_1$. De manera análoga al primer paso construimos una familia, $\mathcal{F}_2 = \{B_{2,1}, B_{2,2}, \ldots\}$, de bolas disjuntas y a lo más numerable. Siguiendo este procedimiento podemos obtener una sucesión de familias a lo más numerables de bolas disjuntas $\mathcal{F}_1, \mathcal{F}_2, \ldots$

Para concluir probaremos que existe m tal que $A \subset \bigcup_{i=1}^{m} \bigcup \mathcal{F}_i$ y además $m \leq 4^n P(n) + 1$. Sea p arbitrario y supongamos que $x \in A - \bigcup_{i=1}^{p} \bigcup \mathcal{F}_i$. Como la unión de todos los $\mathcal{F}_i$ cubre A, bastará probar que $p \leq 4^n P(n)$. De esta manera tendremos que $A \subset \bigcup_{i=1}^{p+1} \bigcup \mathcal{F}_i$ ó $p + 1 \leq 4^n P(n)$ al repetir el mismo argumento.

Sea i tal que $x \in \overline{B}_{r_i}(x_i)$. Por construcción de $\mathcal{F}_j$, para $1 \leq j \leq p$ debe existir $\overline{B}_{r_{i_j}}(x_{i_j}) \in \mathcal{F}_j$ con $i_j < i$ tal que $\overline{B}_{r_i}(x_i) \cap \overline{B}_{r_{i_j}}(x_{i_j}) \neq \emptyset$ y $r_i \leq r_{i_j}$. Para $1 \leq j \leq p$ sea

$$x_j^* = x_i + \frac{3}{2} r_i \frac{x_{i_j} - x_i}{|x_{i_j} - x_i|}.$$

Entonces $\overline{B}_{r_i/2}(x_j^*) \subset \overline{B}_{2r_i}(x_i) \cap \overline{B}_{r_{i_j}}(x_{i_j})$: sea $y \in \overline{B}_{r_i/2}(x_j^*)$, como $r_i \leq |x_i - x_{i_j}| \leq r_i + r_{i_j}$,

$$\begin{aligned} |y - x_{i_j}| &\leq |y - x_j^*| + |x_j^* - x_{i_j}| \leq r_i/2 + \big|(3r_i/2|x_{i_j} - x_i| - 1)(x_{i_j} - x_i)\big| \\ &= r_i/2 + \big|3r_i/2 - |x_{i_j} - x_i|\big| \leq r_i/2 + r_{i_j} - r_i/2 = r_{i_j}, \end{aligned}$$

y

$$|y - x_i| \leq |y - x_j^*| + |x_j^* - x_i| \leq r_i/2 + 3r_i/2 = 2r_i.$$

Recordemos que todo punto puede estar contenido a lo más en $P(n)$ bolas $\overline{B}_{r_{i_j}}(x_{i_j})$. Como $\overline{B}_{r_i/2}(x_j^*) \subset \overline{B}_{r_{i_j}}(x_{i_j})$,

$$\sum_{j=1}^{p} \mathbb{1}_{\overline{B}_{r_i/2}(x_j^*)} \leq P(n)\, \mathbb{1}_{\bigcup_{j=1}^{p} \overline{B}_{r_i/2}(x_j^*)}.$$

De lo anterior y dado que $\bigcup_{j=1}^{p} \overline{B}_{r_i/2}(x_j^*) \subset B_{2r_i}(x_i)$,

$$\begin{aligned} p\,(r_i/2)^n v_n &= \sum_{j=1}^{p} \lambda\big(\overline{B}_{r_i/2}(x_j^*)\big) = \int \sum_{j=1}^{p} \mathbb{1}_{\overline{B}_{r_i/2}(x_j^*)}\, d\lambda \\ &\leq P(n) \int \mathbb{1}_{\bigcup_{j=1}^{p} \overline{B}_{r_i/2}(x_j^*)}\, d\lambda \leq P(n)\, \lambda\big(\overline{B}_{2r_i}(x_i)\big) = P(n)(2r_i)^n v_n, \end{aligned}$$

de donde se obtiene $p \leq 4^n P(n)$ como buscábamos. Tomando $K(n) = 4^n P(n) + 1$ podemos concluir. $\square$

4.2. Diferenciación de Medidas en $\mathbb{R}^n$

En esta sección estudiaremos límites de la forma

$$\lim_{r \to 0} \frac{\nu\big(\overline{B}_r(x)\big)}{\mu\big(\overline{B}_r(x)\big)},$$

donde μ y ν son medidas en $\mathbb{R}^n$. Veremos que para **medidas de Borel** estos límites existen y coinciden con la noción de derivada que teníamos a través del teorema de Radon-Nikodym.

Definición 4.4. Decimos que una medida positiva μ en $\mathbb{R}^n$ es de Borel si:

1) $\mu(K) < \infty$ para todo K compacto.

2) Para todo $E \in \mathcal{B}(\mathbb{R}^n)$, $\mu(E) = \inf\{\mu(U) : U \textit{ es abierto y } E \subset U\}$.

Si μ es una medida compleja o con signo diremos que es de Borel si y sólo si $|\mu|$ es de Borel. Equivalentemente, μ es de Borel si y sólo si $\mu_{Re}^{\pm}$ y $\mu_{Im}^{\pm}$ son de Borel.

Recordemos que a través de una medida positiva μ definida en $\mathcal{B}(\mathbb{R}^n)$ podemos definir una medida exterior $\mu^* : \mathcal{P}(\mathbb{R}^n) \to [0, \infty]$. Por el teorema de extensión de Carathéodory $\mu(E) = \mu^*(E)$ para todo $E \in \mathcal{B}(\mathbb{R}^n)$. Ahora, por definición de la medida exterior, para todo $A \in \mathcal{P}(\mathbb{R}^n)$ y $\epsilon > 0$, siempre podemos encontrar $E \in \mathcal{B}(\mathbb{R}^n)$ tal que $A \subset E$ y $\mu(E) \leq \mu^*(A) + \epsilon$. Si adicionalmente μ es una medida de Borel entonces podemos suponer que E es abierto.

Teorema 4.5. *Sea μ una medida de Borel en $\mathbb{R}^n$, A un subconjunto acotado de $\mathbb{R}^n$ y $\mathcal{F}$ una colección de bolas cerradas tales que todo punto en A es el centro de una infinidad de bolas en $\mathcal{F}$ y*

$$\inf\big\{r : \overline{B}_r(x) \in \mathcal{F}\big\} = 0 \textit{ para todo } x \in A.$$

Entonces existe una subfamilia $\{B_i\}_{i\in\Lambda} \subset \mathcal{F}$ de bolas disjuntas y a lo más numerable tal que

$$\mu^*\bigg(A - \bigcup_{i\in\Lambda} B_i\bigg) = 0.$$

Demostración. Podemos asumir que $\mu(A) > 0$. Nótese que al ser A acotado y μ una medida de Borel, tenemos $\mu^*(A) \leq \mu(\overline{A}) < \infty$. Sea $K(n)$ la constante dada en el teorema de cubierta de Besicovitch y U un abierto tal que $A \subset U$ y

$$\mu(U) \leq \left(1 + \frac{1}{4K(n)}\right)\mu^*(A).$$

Por el mismo teorema y dado que los radios de las bolas en $\mathcal{F}$ son arbitrariamente pequeños, existen familias $\mathcal{F}_1, \dots, \mathcal{F}_{K(n)} \subset \mathcal{F}$ de bolas disjuntas y a lo más numerables tales que $A \subset \bigcup_{i=1}^{K(n)} \bigcup \mathcal{F}_i \subset U$. Supongamos que de estas familias, $\bigcup \mathcal{F}_j$ es la de mayor medida. Entonces,

$$\mu^*(A) \leq \sum_{i=1}^{K(n)} \mu\Big(\bigcup \mathcal{F}_i\Big) \leq K(n)\,\mu\Big(\bigcup \mathcal{F}_j\Big).$$

Como $\mu\big(\bigcup \mathcal{F}_j\big) \leq \mu(U) < \infty$ y $\bigcup \mathcal{F}_j$ es una unión numerable de bolas disjuntas, por continuidad de las medidas deben existir $B_1, \dots, B_{k_1} \in \mathcal{F}_j$ tales que $\mu\big(\bigcup \mathcal{F}_j\big) \leq 2\mu\big(\bigcup_{i=1}^{k_1} B_i\big)$. De esta manera

$$\frac{1}{2K(n)}\mu^*(A) \leq \mu\Big(\bigcup_{i=1}^{k_1} B_i\Big).$$

Si tomamos $A_1 = A - \bigcup_{i=1}^{k_1} B_i$,

$$\begin{aligned}\mu^*(A_1) &\leq \mu\Big(U - \bigcup_{i=1}^{k_1} B_i\Big) = \mu(U) - \mu\Big(\bigcup_{i=1}^{k_1} B_i\Big) \\ &\leq \left(1 + \frac{1}{4K(n)} - \frac{1}{2K(n)}\right)\mu^*(A) = \theta\mu^*(A),\end{aligned}$$

con $\theta = 1 - \frac{1}{4K(n)}$. Ahora tomemos un abierto U_1 tal que $A_1 \subset U_1 \subset U - \bigcup_{i=1}^{k_1} B_i$ y

$$\mu(U_1) \leq \left(1 + \frac{1}{4K(n)}\right)\mu^*(A_1).$$

Repitiendo el mismo argumento que utilizamos para A, podemos encontrar una cantidad finita de bolas disjuntas $B_{k_1+1}, \dots, B_{k_2} \in \mathcal{F}$ tales que $\bigcup_{i=k_1+1}^{k_2} B_i \subset U_1$ y para $A_2 = A_1 - \bigcup_{i=k_1+1}^{k_2} B_i = A - \bigcup_{i=1}^{k_2} B_i$,

$$\mu^*(A_2) \leq \theta\mu^*(A_1) \leq \theta^2\mu^*(A).$$

Repitiendo este procedimiento de manera inductiva podemos encontrar una familia $\{B_i\}_{i\in\Lambda} \subset \mathcal{F}$ de bolas disjuntas y a lo más numerable tal que para todo $m \in \mathbb{N}$ existe k_m tal que

$$\mu^*\left(A - \bigcup_{i=1}^{k_m} B_i\right) \leq \theta^m \mu^*(A)$$

Por monotonía de la medida exterior y dado que $\theta < 1$ podemos concluir que

$$\mu^*\left(A - \bigcup_{i\in\Lambda} B_i\right) = 0. \qquad \square$$

Definición 4.6. Sean ν y μ medidas positivas de Borel en $\mathbb{R}^n$. Para todo $x \in \mathbb{R}^n$ definimos la **derivada superior e inferior** de ν con respecto μ como

$$\overline{D}_\mu \nu(x) := \begin{cases} \text{lím sup}_{r\to 0} \frac{\nu(\overline{B}_r(x))}{\mu(\overline{B}_r(x))} & \textit{si } \mu(\overline{B}_r(x)) > 0 \textit{ para todo } r > 0, \\ +\infty & \textit{si } \mu(\overline{B}_r(x)) = 0 \textit{ para algún } r > 0, \end{cases}$$

$$\underline{D}_\mu \nu(x) := \begin{cases} \text{lím inf}_{r\to 0} \frac{\nu(\overline{B}_r(x))}{\mu(\overline{B}_r(x))} & \textit{si } \mu(\overline{B}_r(x)) > 0 \textit{ para todo } r > 0, \\ +\infty & \textit{si } \mu(\overline{B}_r(x)) = 0 \textit{ para algún } r > 0. \end{cases}$$

Si $\overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x) < \infty$ decimos que ν es diferenciable con respecto μ en x. En este caso denotamos su derivada como

$$D_\mu \nu(x) := \overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x).$$

Lema 4.7. *Sea* $0 < c < \infty$.

1) Si $A \subset \{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) \geq c\}$ *es acotado entonces* $\nu^*(A) \geq c\, \mu^*(A)$.

2) Si $A \subset \{x \in \mathbb{R}^n : \underline{D}_\mu \nu(x) \leq c\}$ *es acotado entonces* $\nu^*(A) \leq c\, \mu^*(A)$.

Demostración. 1) Dado $c > \epsilon > 0$ sea U abierto tal que $A \subset U$ y $\nu(U) \leq \nu^*(A) + \epsilon$. Ahora, si $x \in A$, por definición de la derivada superior, para todo $R > 0$ existe $r < R$ tal que $\nu(\overline{B}_r(x)) \geq (c-\epsilon)\mu(\overline{B}_r(x))$. De esta manera para todo $x \in A$ podemos obtener una familia de bolas cerradas que cumplan la desigualdad anterior, que estén contenidas en U y cuyos radios sean arbitrariamente pequeños. Por el teorema 4.5 podemos encontrar una subfamilia de bolas disjuntas y a lo más numerable $\{\overline{B}_{r_i}(x_i)\}_{i\in\Lambda}$ tal que

$$\mu^*\left(A - \bigcup_{i\in\Lambda} \overline{B}_{r_i}(x_i)\right) = 0.$$

En consecuencia,

$$\begin{aligned}(c-\epsilon)\mu^*(A) &\leq (c-\epsilon)\mu^*\left(A - \bigcup_{i\in\Lambda} \overline{B}_{r_i}(x_i)\right) + (c-\epsilon)\mu^*\left(\bigcup_{i\in\Lambda} \overline{B}_{r_i}(x_i)\right) \\ &= \sum_{i\in\Lambda} (c-\epsilon)\mu(\overline{B}_{r_i}(x_i)) \leq \sum_{i\in\Lambda} \nu(\overline{B}_{r_i}(x_i)) \leq \nu(U) \leq \nu^*(A) + \epsilon.\end{aligned}$$

Como ϵ es arbitrario podemos concluir que $\nu^*(A) \geq c\, \mu^*(A)$. La prueba de 2) es análoga. $\square$

Proposición 4.8. *Sean* ν *y* μ *medidas positivas de Borel en* $\mathbb{R}^n$. *Entonces,*

1) $D_\mu \nu$ *existe y es finito* μ*-c.d.s.*

2) $D_\mu \nu$ *es* μ^**-medible.*

Demostración. 1) Sea $\{Q_k\}_{k\in\mathbb{N}}$ una partición de $\mathbb{R}^n$ conformada por conjuntos acotados e $I = \{x \in \mathbb{R}^n : \overline{D}_\mu\nu(x) = \infty\}$. Para cada $k \in \mathbb{N}$ y $c > 0$, $I \cap Q_k \subset \{x \in \mathbb{R}^n : \overline{D}_\mu\nu(x) \geq c\}$. Así que por el lema 4.7

$$\mu^*(I \cap Q_k) \leq \frac{1}{c}\nu^*(I \cap Q_k).$$

Como c puede ser arbitrariamente grande se sigue que $\mu^*(I \cap Q_k) = 0$ para todo $k \in \mathbb{N}$. Podemos concluir que $\overline{D}_\mu\nu < \infty$ μ-c.d.s.
De manera similar, para todo $k \in \mathbb{N}$ y $0 < a < b$ sea

$$A^k_{a,b} = \big\{x \in Q_k : \underline{D}_\mu\nu(x) \leq a < b \leq \overline{D}_\mu\nu(x)\big\}.$$

Nuevamente por el lema 4.7, $b\,\mu^*(A^k_{a,b}) \leq \nu^*(A^k_{a,b}) \leq a\,\mu^*(A^k_{a,b})$. Como $a < b$ se sigue que $\mu^*(A^k_{a,b}) = 0$ y como

$$\big\{x \in \mathbb{R}^n : \underline{D}_\mu\nu(x) < \overline{D}_\mu\nu(x)\big\} = \bigcup_{k\in\mathbb{N}} \bigcup_{a,b\in\mathbb{Q}} A^k_{a,b},$$

podemos concluir que $\underline{D}_\mu\nu = \overline{D}_\mu\nu$ μ-c.d.s.

2) Comencemos por ver que para todo $r > 0$ y $x \in \mathbb{R}^n$,

$$\limsup_{y\to x} \mu\big(\overline{B}_r(y)\big) = \lim_{\epsilon\to 0} \sup_{y\in B_\epsilon(x)} \mu\big(\overline{B}_r(y)\big) \leq \mu\big(\overline{B}_r(x)\big).$$

Supongamos que $y_k \to x$ y sean $f_k = \mathbb{1}_{\overline{B}_r(y_k)}$, $f = \mathbb{1}_{\overline{B}_r(x)}$ y $g = \mathbb{1}_{\overline{B}_{2r}(x)}$. Entonces $\limsup_k f_k \leq f$ y a partir de alguna k, $f_k \leq g$. Así $0 \leq g - f \leq \liminf_k (g - f_k)$ y por el lema de Fatou

$$\mu\big(\overline{B}_{2r}(x)\big) - \mu\big(\overline{B}_r(x)\big) \leq \int \liminf_k (g - f_k)d\mu$$
$$\liminf_k \int (g - f_k)d\mu = \mu\big(\overline{B}_{2r}(x)\big) - \limsup_k \mu\big(\overline{B}_r(y_k)\big),$$

de donde se sigue lo que buscábamos. De esta manera $f_r(x) = \mu\big(\overline{B}_r(x)\big)$ define una función semicontinua por arriba; es decir, para todo $\alpha \in \mathbb{R}$, $\{x \in \mathbb{R}^n : f_r(x) < \alpha\}$ es abierto: si $f_r(x) < \alpha$, dado que $\limsup_{y\to x} f(y) \leq f(x)$, existe $\epsilon > 0$ y $\delta > 0$ tal que para todo $y \in B_\epsilon(x)$, $f_r(y) \leq f_r(x) + \delta < \alpha$.

Podemos concluir que f_r es Borel-medible. Análogamente podemos probar que $x \mapsto \nu\big(\overline{B}_r(x)\big)$ es medible y por lo tanto

$$g_r(x) := \begin{cases} \frac{\nu\left(\overline{B}_r(x)\right)}{\mu\left(\overline{B}_r(x)\right)} & \textit{si } \mu\big(\overline{B}_r(x)\big) > 0, \\ +\infty & \textit{si } \mu\big(\overline{B}_r(x)\big) = 0, \end{cases}$$

define una función Borel-Medible. Por 1) sabemos que $D_\mu\nu = \lim_{r\to 0} g_r$ μ-c.d.s. Posiblemente haya que definir $D_\mu\nu$ en un conjunto μ-nulo pero aún así podemos concluir que $D_\mu\nu$ es μ^*-medible. $\square$

Nota: La σ-álgebra de conjuntos μ^*-medibles resulta ser la completación de la σ-álgebra de Borel con respecto μ. Por la proposición 2.20 podemos encontrar una función Borel-medible que coincida con el límite $D_\mu\nu$ μ-c.d.s. Ahora, si $\nu \ll \mu$ se sigue que la completación de $\mathcal{B}(\mathbb{R}^n)$ bajo μ está contenida en la completación bajo ν. Por lo tanto, todo conjunto μ^*-medible es ν^*-medible.

Teorema 4.9. *Sean ν y μ medidas positivas de Borel en $\mathbb{R}^n$. Si $\nu \ll \mu$, entonces para todo conjunto μ^*-medible A,*

$$\nu(A) = \int_A D_\mu \nu \, d\mu.$$

En otras palabras $D_\mu \nu$ es la derivada de Radon-Nikodym de ν con respecto μ.

Demostración. Veamos que $D_\mu \nu \neq 0$ ν-c.d.s. Sea $B = \{x \in \mathbb{R}^n : D_\mu \nu(x) = 0\}$ y $\{Q_k\}_{k\in\mathbb{N}}$ una partición de $\mathbb{R}^n$ conformada por conjuntos acotados. Dado que para todo $k \in \mathbb{N}$ y $c > 0$,

$$B \cap Q_k \subset \{x \in \mathbb{R}^n : \underline{D}_\mu \nu(x) \leq c\},$$

por el lema 4.7 tenemos $\nu^*(B \cap Q_k) \leq c\, \mu^*(B \cap Q_k)$. Como c es arbitrario, $\nu^*(B \cap Q_k) = 0$ para todo $k \in \mathbb{N}$ y podemos concluir que $\nu^*(B) = 0$.

Nótese que basta probar la igualdad que buscamos para conjuntos acotados ya que el caso general se puede extender fácilmente de este caso. Consideremos un conjunto μ^*-medible y acotado A y sea N el conjunto μ-nulo (y por ende ν-nulo) tal que en N^c, $D_\mu \nu$ existe y es finito. Por otro lado, para $1 < t < \infty$ y todo $m \in \mathbb{Z}$ sea

$$A_m := A \cap \left\{x \in \mathbb{R}^n : t^m \leq D_\mu \nu(x) < t^{m+1}\right\}.$$

Entonces $A = \bigcup_{m\in\mathbb{Z}} A_m \cup (A \cap N) \cup (A \cap B)$ y por el lema 4.7,

$$\begin{aligned}\nu(A) &= \sum_{m\in\mathbb{Z}} \nu(A_m) \leq \sum_{m\in\mathbb{Z}} t^{m+1} \mu(A_m)\\ &= t \sum_{m\in\mathbb{Z}} \int t^m \mathbb{1}_{A_m} d\mu \leq t \sum_{m\in\mathbb{Z}} \int_{A_m} D_\mu \nu \, d\mu = t \int_A D_\mu \nu \, d\mu.\end{aligned}$$

De manera análoga,

$$\begin{aligned}\nu(A) &= \sum_{m\in\mathbb{Z}} \nu(A_m) \geq \sum_{m\in\mathbb{Z}} t^m \mu(A_m)\\ &= t^{-1} \sum_{m\in\mathbb{Z}} \int \left(t^{m+1} \mathbb{1}_{A_m}\right) d\mu \geq \sum_{m\in\mathbb{Z}} \int_{A_m} D_\mu \nu \, d\mu = t^{-1} \int_A D_\mu \nu \, d\mu.\end{aligned}$$

Por lo tanto para todo $1 < t < \infty$, $t^{-1} \int_A D_\mu \nu \, d\mu \leq \nu(A) \leq t \int_A D_\mu \nu \, d\mu$. Haciendo que $t \to 1$ obtenemos el resultado que buscamos. □

Definición 4.10. Decimos que una función medible $f : \mathbb{R}^n \to \mathbb{C}$ es **localmente integrable** con respecto μ si para todo compacto K, $\int_K |f| d\mu < \infty$. El espacio de funciones localmente integrables con respecto μ es denotado por $L^1_{loc}(\mu)$.

Proposición 4.11. *Sea μ una medida positiva de Borel. Si $f \in L^1_{loc}(\mu) \cap L^+$ entonces $d\nu = f d\mu$ es una medida de Borel.*

Demostración. Por definición $f \in L^1_{loc}(\mu)$ equivale a que $\nu(K) < \infty$ para todo compacto K. Para la regularidad por fuera consideremos un conjunto acotado E y $\epsilon > 0$. Por otro lado sea $R > 0$ tal que $E \subset B_R(0)$. Como $d\tilde{\nu} = f \mathbb{1}_{B_R(0)} d\mu$ es una medida finita y absolutamente continua con respecto μ, por la proposición 3.16 existe δ tal que si $\mu(A) < \delta$ entonces $\tilde{\nu}(A) < \epsilon$. Como μ es de Borel existe U abierto tal que $E \subset U \subset B_R(0)$ y $\mu(U) \leq \mu(E) + \delta$. Entonces,

$$\nu(U) - \nu(E) = \int_{U-E} f\, d\mu = \int_{U-E} f \mathbb{1}_{B_R(0)}\, d\mu = \tilde{\nu}(U - E) < \epsilon.$$

Si E no es acotado podemos particionarlo en una cantidad numerable de conjuntos acotados E_k y aplicar lo anterior para encontrar abiertos $U_k \supset E_k$ tales que $\nu(U_k) \leq \nu(E_k) + \epsilon/2^k$. Entonces $E \subset \bigcup_{k\in\mathbb{N}} U_k$ y $\nu\big(\bigcup_{k\in\mathbb{N}} U_k\big) \leq \nu(E) + \epsilon$. Podemos concluir que ν es regular por fuera. $\square$

Corolario 4.12. *Si μ es una medida de Borel positiva y $f \in L^1_{loc}(\mu)$ entonces*

$$\lim_{r\to 0} \frac{1}{\mu\big(\overline{B}_r(x)\big)} \int_{\overline{B}_r(x)} f\, d\mu = f(x) \quad \mu\text{-para casi todo } x \in \mathbb{R}^n.$$

Demostración. Considerando por separado las partes real, imaginaria, positiva ó negativa podemos suponer que $f \geq 0$. Por la proposición anterior tenemos que $d\nu = f d\mu$ es una medida de Borel absolutamente continua con respecto μ. Por el teorema 4.9, $d\nu = D_\mu \nu\, d\mu$ y podemos concluir que $D_\mu \nu = f$ μ-c.d.s. que es precisamente el enunciado. $\square$

Podemos dar un resultado aún mas fuerte que el del corolario 4.12. Con este fin introducimos el siguiente concepto:
Para $x \in \mathbb{R}^n$ diremos que una familia $\{E_r(x)\}_{r>0} \subset \mathcal{B}(\mathbb{R}^n)$ se **contrae adecuadamente** a x con respecto μ si:

1) $E_r(x) \subset \overline{B}_r(x)$ para todo $r > 0$,

2) existe $\alpha_x > 0$ tal que para todo $r > 0$, $\mu\big(E_r(x)\big) > \alpha_x\, \mu\big(\overline{B}_r(x)\big)$.

Teorema 4.13. *Sea μ una medida de Borel positiva y $f \in L^1_{loc}(\mu)$. Si para todo $x \in \mathbb{R}^n$, $\{E_r(x)\}_{r>0}$ se contrae adecuadamente a x con respecto μ, entonces*

$$\lim_{r\to 0} \frac{1}{\mu\big(E_r(x)\big)} \int_{E_r(x)} |f(y) - f(x)|\, d\mu(y) = 0 \quad \mu\text{-para casi todo } x \in \mathbb{R}^n.$$

En particular $\lim_{r\to 0} 1/\mu\big(E_r(x)\big) \int_{E_r(x)} f(y)\, d\mu(y) = f(x)$ *μ-c.d.s.*

Demostración. Dado $\beta \in \mathbb{C}$ es fácil ver que $g(x) = |f(x) - \beta|$ está en $L^1_{loc}(\mu)$. Por el corolario 4.12 existe un conjunto μ-nulo N_β tal que para todo $x \in N_\beta^c$,

$$\lim_{r\to 0} \frac{1}{\mu\big(\overline{B}_r(x)\big)} \int_{\overline{B}_r(x)} |f - \beta|\, d\mu = |f(x) - \beta|.$$

Sea D un conjunto denso numerable de $\mathbb{C}$, entonces $N = \bigcup_{\beta\in D} N_\beta$ es μ-nulo. Luego, dado $\epsilon > 0$ y $x \in N^c$ sea $\beta \in D$ tal que $|f(x) - \beta| < \alpha_x \epsilon$. De esta manera $|f(y) - f(x)| \leq |f(y) - \beta| + \alpha_x \epsilon$ y tendremos que

$$\begin{aligned}
\limsup_{r\to 0} & \frac{1}{\mu\big(E_r(x)\big)} \int_{E_r(x)} |f(y) - f(x)|\, d\mu(y) \\
&\leq \limsup_{r\to 0} \frac{1}{\alpha_x \mu\big(\overline{B}_r(x)\big)} \int_{\overline{B}_r(x)} |f(y) - f(x)|\, d\mu(y) \\
&\leq \limsup_{r\to 0} \frac{1}{\alpha_x \mu\big(\overline{B}_r(x)\big)} \int_{\overline{B}_r(x)} \big(|f(y) - \beta| + \alpha_x \epsilon\big)\, d\mu(y) = \frac{1}{\alpha_x} |f(x) - \beta| + \epsilon = 2\epsilon.
\end{aligned}$$

Como ϵ es arbitrario podemos concluir. $\square$

Teorema 4.14. *Sea ν una medida compleja o con signo de Borel y μ una medida positiva de Borel. Si $d\nu = d\eta + fd\mu$ es la descomposición de Lebesgue-Radon-Nikodym de ν con respecto μ y para todo $x \in \mathbb{R}^n$, $\{E_r(x)\}_{r>0}$ es una familia que se contrae adecuadamente a x con respecto μ, entonces*

$$\lim_{r\to 0} \frac{\nu\big(E_r(x)\big)}{\mu\big(E_r(x)\big)} = f(x) \quad \mu\text{-para casi todo } x \in \mathbb{R}^n.$$

Demostración. Es fácil ver que $d\nu_{Re}^{\pm} = d\eta_{Re}^{\pm} + Re(f)^{\pm} d\mu$ y $d\nu_{Im}^{\pm} = d\eta_{Im}^{\pm} + Im(f)^{\pm} d\mu$, así que por linealidad del límite en cuestión podemos suponer que ν y η son medidas positivas al igual que f una función no-negativa. Dado que $\nu(K)$ es finito para todo compacto K, podemos concluir lo mismo para η y $fd\mu$. En particular tenemos $f \in L^1_{loc}(\mu)$. Por otro lado, para todo conjunto acotado $E \in \mathcal{B}(\mathbb{R}^n)$ tenemos $\eta(E) = \nu(E) - \int_E f d\mu$, de donde se sigue fácilmente la regularidad por fuera de η. Por lo tanto η también es una medida de Borel. Ahora, por el teorema anterior sabemos que

$$\lim_{r\to 0} \frac{1}{\mu\big(E_r(x)\big)} \int_{E_r(x)} f\, d\mu = f(x) \quad \mu\text{-para casi todo } x \in \mathbb{R}^n.$$

Así que todo se reduce a probar que $\lim_{r\to 0} \eta\big(E_r(x)\big)/\mu\big(E_r(x)\big) = 0$ μ-c.d.s. Como $\eta \perp \mu$ existen $A, B \in \mathcal{B}(\mathbb{R}^n)$ disjuntos tales que $\mathbb{R}^n = A \cup B$ y $\eta(B) = \mu(A) = 0$. Sea $\{Q_k\}_{k\in\mathbb{N}} \subset \mathcal{B}(\mathbb{R}^n)$ una partición de $\mathbb{R}^n$ conformada por conjuntos acotados y definamos

$$B_k^m = \{x \in B \cap Q_k : D_\mu \eta(x) \geq 1/m\}.$$

Por el lema 4.7 tenemos que $\mu(B_k^m) \leq m\,\eta(B_k^m) \leq m\,\eta(B) = 0$. Dado que

$$\big\{x \in B : D_\mu\eta(x) > 0\big\} = \bigcup_{k,m\in\mathbb{N}} B_k^m,$$

se sigue que $D_\mu\eta = 0$ μ-c.d.s. y como

$$\limsup_{r\to 0} \frac{\eta\big(E_r(x)\big)}{\mu\big(E_r(x)\big)} \leq \limsup_{r\to 0} \frac{\eta\big(\overline{B}_r(x)\big)}{\alpha_x\, \mu\big(\overline{B}_r(x)\big)} = \frac{1}{\alpha_x} D_\mu\eta(x) \quad \mu\text{-c.d.s.}$$

podemos concluir. $\square$

4.3. Funciones de Variación Acotada

En esta sección comenzamos por estudiar algunas propiedades esenciales de las funciones monótonas y posteriormente estudiaremos una generalización de estas, a saber, las funciones de variación acotada. A lo largo de esta sección la expresión c.d.s. corresponde a la medida de Lebesgue.

Teorema 4.15. *Sea $F : \mathbb{R} \to \mathbb{R}$ una función no-decreciente y $G(x) = F(x+) := \lim_{y\to x^+} F(y)$.*

1) F tiene a lo más una cantidad numerable de discontinuidades.

2) F' y G' están definidos c.d.s. y coinciden con una función medible y no-negativa c.d.s.

Demostración. 1) Para cada punto x de discontinuidad de F, $\big(F(x-), F(x+)\big)$ define un intervalo abierto. Por otro lado, si $y > x$ es otro punto de discontinuidad $F(x+) \leq F(y-)$ por lo que

$$\Big\{\big(F(x-), F(x+)\big) : x \text{ es puno de discontinuidad de } F\Big\}$$

conforma una familia de intervalos disjuntos. Como cada intervalo contiene números racionales, de la numerabilidad de éstos podemos inferir que dicha familia es a lo más numerable. Es decir, F tiene a lo más una cantidad numerable de discontinuidades.

2) Recordemos de la sección 1.4 que G induce una medida de Borel μ_G en $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ tal que

$$G(x+h) - G(x) = \begin{cases} \mu_G\big((x, x+h]\big) & \text{si } h > 0, \\ -\mu_G\big((x+h, x]\big) & \text{si } h < 0. \end{cases}$$

Sea λ la medida de Lebesgue y $d\mu_G = d\eta + f\, d\lambda$ la descomposición de Lebesgue-Radon-Nikodym de μ_G con respecto λ. Como las familias $\big\{(x+h, x]\big\}_{h<0}$ y $\big\{(x, x+h]\big\}_{h>0}$ se contraen adecuadamente a x con respecto λ, del teorema 4.14 tenemos que

$$\lim_{h\to 0} \frac{G(x+h) - G(x)}{h} = \lim_{h\to 0^+} \frac{\mu_G\big((x, x+h]\big)}{\lambda\big((x, x+h]\big)} = \lim_{h\to 0^-} \frac{\mu_G\big((x+h, x]\big)}{\lambda\big((x+h, x]\big)} = f(x) \text{ c.d.s.}$$

Es decir, $G' = f \geq 0$ c.d.s.
Para concluir basta probar que $H = G - F$ es diferenciable c.d.s. y $H' = 0$ c.d.s. Nótese que $H \neq 0$ sólo para un subconjunto de puntos de discontinuidad de F. Sea $\{x_i\}_i$ una enumeración de dicho conjunto y definamos $\mu = \sum_i H(x_i)\delta_{x_i}$. Entonces μ es localmente finita:

$$\mu\big((-N, N)\big) = \sum_{|x_i|<N} H(x_i) = \sum_{|x_i|<N} \big(F(x_i+) - F(x_i)\big) \leq F(N) - F(-N) < \infty.$$

De esta forma podemos ver que μ es inducida por la función no-decreciente y continua por la derecha

$$\tilde{H}(x) = \begin{cases} \mu\big((0, x]\big) & \text{si } x > 0, \\ -\mu\big((x, 0]\big) & \text{si } x \leq 0. \end{cases}$$

Por el teorema 1.26 concluimos que μ es de Borel. Más aún, $\mu \perp \lambda$ y otra aplicación del teorema 4.14 nos muestra que

$$\begin{aligned} \limsup_{|h|\to 0} \left| \frac{H(x+h) - H(x)}{h} \right| &\leq \limsup_{|h|\to 0} \frac{H(x+h) + H(x)}{|h|} \\ &\leq 2 \limsup_{h\to 0} \frac{\mu\big([x-h, x+h]\big)}{\lambda\big([x-h, x+h]\big)} = 0 \text{ c.d.s.} \end{aligned}$$

□

Dada una función $F : (a,b) \to \mathbb{C}$ definimos su **función de variación total** $T_F : (a,b) \to [0,\infty]$ como

$$T_F(x) := \sup\left\{ \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| : n \in \mathbb{N},\, a < x_0 < x_1 < \cdots < x_n = x \right\}.$$

Claramente T_F define una función no-decreciente. Si $T_F(b) = \lim_{x\to b} T_F(x) < \infty$ diremos que F es una función de **variación acotada** en (a,b) y lo denotaremos por $F \in BV(a,b)$. Una consecuencia inmediata de que $F \in BV(a,b)$ es que F debe ser acotada ya que para todo $x \in (a,b)$ y x_0 fijo,

$$|F(x)| \leq |F(x_0)| + |F(x_0) - F(x)| \leq |F(x_0)| + T_F(b).$$

Nótese que la definición de la función de variación total también es valida para $a = -\infty$ ó $b = \infty$.

Para un ejemplo de una función acotada que no es de variación acotada consideremos $F(x) = x\cos(\pi/2x)$ en el intervalo $(0,1)$ y para todo $k \in \mathbb{N}$ sea $x_0 = 1/(2k+1), x_1 = 1/2k, \ldots, x_{2k-1} = 1/2, x_{2k} = 1$. Entonces,

$$\begin{aligned}\sum_{i=1}^{2k} |F(x_i) - F(x_{i-1})| &= \sum_{i=0}^{2k-1} \left| \frac{\cos\big((2k-i)\pi/2\big)}{(2k-i)} - \frac{\cos\big((2k+1-i)\pi/2\big)}{(2k+1-i)} \right| \\ &= \frac{1}{2k} + \frac{1}{2k} + \cdots + \frac{1}{4} + \frac{1}{4} + \frac{1}{2} + \frac{1}{2} = \frac{1}{k} + \cdots + \frac{1}{3} + \frac{1}{2} + 1.\end{aligned}$$

Hacienod que $k \to \infty$ podemos ver que $T_F(1) = \infty$.

Si $F : (a,b) \to \mathbb{R}$ es no-decreciente y acotada, entonces $F \in BV(a,b)$. De hecho se ve fácilmente que $T_F(b) = F(b) - F(a)$. El recíproco casi es cierto:

Teorema 4.16. *Dada $F : (a,b) \to \mathbb{R}$ tenemos que $F \in BV(a,b)$ si y sólo si F es la diferencia de dos funciones no-decrecientes y acotadas. En este caso $F^+ = \frac{1}{2}(T_F + F)$ y $F^- = \frac{1}{2}(T_F - F)$ son no-decrecientes y acotadas y podemos tomar la representación $F = F^+ - F^-$.*

Demostración. Si $F = G - H$ con G y H no-decrecientes y acotadas se sigue fácilmente que $T_F(b) \leq \big(G(b) - G(a)\big) + \big(H(b) - H(a)\big)$.
Ahora, si $F \in BV(a,b)$ tanto T_F como F son acotadas. Para concluir probemos que $T_F \pm F$ es no-decreciente. Dados $a < x < y < b$ y $\epsilon > 0$ sean $x_0 < \cdots < x_n = x$ tales que $\sum_{i=1}^n |F(x_i) - F(x_{i-1})| > T_F(x) - \epsilon$. Entonces,

$$\begin{aligned}T_F(y) \pm F(y) &\geq \sum_{i=1}^n |F(x_i) - F(x_{i-1})| + |F(y) - F(x)| \pm \big(F(y) - F(x)\big) \pm F(x) \\ &\geq T_F(x) - \epsilon \pm F(x)\end{aligned}$$

Como ϵ es arbitrario podemos concluir. $\square$

La descomposición $F = F^+ - F^-$ enunciada en el teoram 4.16 es llamada la **descomposición de Jordan** y F^+ y F^- son llamadas la **variación positiva y negativa** de F, respectivamente. El motivo de este nombramiento es el siguiente: supongamos que el límite (existente) $F(a) = \lim_{x\to a} F(x)$ es cero. Entonces

$$\begin{aligned}
&\frac{1}{2}\big(T_F(x) \pm F(x)\big)\\
&\quad = \sup\left\{\frac{1}{2}\left(\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| \pm F(x_n)\right) : a = x_0 < \cdots < x_n = x\right\}\\
&\quad = \sup\left\{\sum_{i=1}^{n} \frac{1}{2}\Big(|F(x_i) - F(x_{i-1})| \pm \big(F(x_i) - F(x_{i-1})\big)\Big) : a = x_0 < \cdots < x_n = x\right\}\\
&\quad = \sup\left\{\sum_{i=1}^{n} \big(F(x_i) - F(x_{i-1})\big)^{\pm} : a = x_0 < \cdots < x_n = x\right\}.
\end{aligned}$$

Ahora, si F toma valores complejos está claro que

$$\text{máx}\,\big(T_{Re\,F}(x), T_{Im\,F}(x)\big) \leq T_F(x) \leq T_{Re\,F}(x) + T_{Im\,F}(x),$$

de donde se sigue que $F \in BV(a,b)$ si y sólo $Re\,F,\ Im\,F \in BV(a,b)$. Como consecuencia de este hecho evidente, y de los teoremas 4.15 y 4.16, tenemos el siguiente resultado:

Proposición 4.17. *Sea $F \in BV(a,b)$, entonces:*

1) $F(x\pm) := \text{lím}_{y\to x^{\pm}} F(y)$ *existen para todo* $x \in (a,b)$, *al igual que* $F(a) := \text{lím}_{y\to a} F(y)$ *y* $F(b) := \text{lím}_{y\to b} F(y)$.

2) F *tiene a lo más una cantidad numerable de discontinuidades.*

3) F *y* $G(x) = F(x+)$ *son difernciables c.d.s. y* $F' = G'$ *c.d.s.*

De la proposición 4.17 tenemos que toda función de variación acotada tiene una modificación continua por la derecha.

Lema 4.18. *Si $F \in BV(a,b)$ es continua por la derecha también T_F es continua por la derecha.*

Demostración. Sean $x \in (a,b)$ y $\epsilon > 0$ arbitrarios. Notando que para $x < y < b$ fijo,

$$T_F(y) - T_F(x) = \sup\left\{\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| : x = x_0 < \cdots < x_n = y\right\},$$

podemos tomar $x = x_0 < \cdots < x_n = y$ tales que $\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| > T_F(y) - T_F(x) - \epsilon$. Por otro lado, si $x = x_0 < t < x_1$ tendremos que

$$\sum_{i=2}^{n} |F(x_i) - F(x_{i-1})| + |F(x_1) - F(t)| \leq T_F(y) - T_F(t) \leq T_F(y) - T_F(x+).$$

Dado que F es continua por la derecha, tomando el límite cuando $t \downarrow x$ obtenemos

$$T_F(y) - T_F(x+) \geq \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| > T_F(y) - T_F(x) - \epsilon.$$

Se sigue que $T_F(x+) - T_F(x) < \epsilon$ para todo $\epsilon > 0$ y podemos concluir. $\square$

Como consecuencia del lema 4.18, si $F \in BV(a,b)$ es continua por la derecha, tenemos

$$F = (Re\,F)^+ - (Re\,F)^- + i\big((Im\,F)^+ - (Im\,F)^-\big),$$

donde cada término es una función no-decreciente y continua por la derecha. Así, utilizando los resultados de la sección 1.4 podemos definir una única medida de Borel compleja μ_F en $\big((a,b),\mathcal{B}(a,b)\big)$ tal que para todo $a \leq c < d < b$,

$$\mu_F\big((c,d]\big) = F(d) - F(c).$$

La unicidad se sigue fácilmente del teorema de clases monótonas. Recíprocamente, para toda medida compleja μ en $\big((a,b),\mathcal{B}(a,b)\big)$, existe una $F \in BV(a,b)$ tal que $\mu\big((c,d]\big) = F(d) - F(c)$ dada por

$$F(x) = \mu\big((a,x]\big).$$

Teorema 4.19. *Sea $F \in BV(a,b)$ continua por la derecha, entonces $|\mu_F| = \mu_{T_F}$. Si además F es real y $F = F^+ - F^-$ es su descomposición de Jordan, entonces $\mu_F^\pm = \mu_{F^\pm}$.*

Demostración. Para la primera parte, por el lema de clases monótonas basta probar que para todo $a \leq c < d < b$,

$$|\mu_F|\big((c,d]\big) = \mu_{T_F}\big((c,d]\big) = T_F(d) - T_F(c).$$

Para probar esto usaremos la definición alternativa de $|\mu_F|$:

$$|\mu_F|(E) = \sup\left\{\sum_{i=1}^n |\mu_F(A_i)| : A_i \in \mathcal{B}(a,b),\ A_i \cap A_j = \emptyset \text{ si } i \neq j,\ E = \bigcup_{i=1}^n A_i\right\}.$$

También notamos, como en la demostración del lema 4.18, que

$$T_F(d) - T_F(c) = \sup\left\{\sum_{i=1}^n |F(x_i) - F(x_{i-1})| : c = x_0 < \cdots < x_n = d\right\},$$

De esta manera si $c = x_0 < \cdots < x_n = d$ se sigue que

$$\sum_{i=1}^n |F(x_i) - F(x_{i-1})| = \sum_{i=1}^n \big|\mu_F\big((x_{i-1},x_i]\big)\big| \leq |\mu_F|\big((c,d]\big).$$

Por lo tanto $T_F(d) - T_F(c) \leq |\mu_F|\big((c,d]\big)$. Para la otra desigualdad primero notamos que para todo intervalo $(x,y]$,

$$\big|\mu_F\big((x,y]\big)\big| = |F(y) - F(x)| \leq T_F(y) - T_F(x) = \mu_{T_F}\big((x,y]\big).$$

Luego, del lema 2.47 podemos ver que para el caso especial de $\mathbb{R}$, todo abierto U es la unión numerable de intervalos disjuntos de la forma $(x_i,y_i]$. De esta forma para todo abierto U,

$$\big|\mu_F(U)\big| = \Big|\sum_i \mu_F\big((x_i,y_i]\big)\Big| \leq \sum_i \mu_{T_F}\big((x_i,y_i]\big) = \mu_{T_F}(U).$$

Por la propiedad de regularidad por fuera de estas medidas podemos extender la desigualad anterior a todo Boreleano. Entonces si $(c,d] = \bigcup_{i=1}^{n} A_i$ con $A_i \in \mathcal{B}(\mathbb{R})$ disjuntos,

$$\sum_{i=1}^{n} |\mu_F(A_i)| \leq \sum_{i=1}^{n} \mu_{T_F}(A_i) = \mu_{T_F}\big((c,d]\big).$$

Por lo tanto $|\mu_F|\big((c,d]\big) \leq \mu_{T_F}\big((c,d]\big)$ y podemos concluir.

La segunda parte se sigue fácilmente de lo anterior y de la identidad $\mu_F^{\pm} = \frac{1}{2}\big(|\mu_F| \pm \mu_F\big)$. De esta manera para todo $a \leq c < d < b$,

$$\begin{aligned} \mu_F^{\pm}\big((c,d]\big) &= \frac{1}{2}\Big(\mu_{T_F}\big((c,d]\big) \pm \mu_F\big((c,d]\big)\Big) \\ &= \frac{1}{2}\big(T_F(d) - T_F(c) \pm (F(d) - F(c))\big) = F^{\pm}(d) - F^{\pm}(c). \end{aligned}$$

$\square$

Proposición 4.20. *Sea λ la medida de Lebesgue en $\big((a,b), \mathcal{B}(a,b)\big)$. Si $F \in BV(a,b)$ entonces $F' \in L^1\big((a,b),\lambda\big)$. Si además F es continua por la derecha, $\mu_F \perp \lambda$ si y sólo si $F' = 0$ c.d.s y $\mu_F \ll \lambda$ si y sólo si $F(x) = F(a) + \int_{(a,x]} F' d\lambda$.*

Demostración. Sin mayor pérdida de generalidad podemos suponer que F es continua por la derecha. Sea $\mu_F = \eta + f d\lambda$ la descomposición de Lebesgue-Radon-Nikodym de μ_F con respecto λ. Como μ_F es finita, $f \in L^1\big((a,b),\lambda\big)$. Luego, del teorema 4.14,

$$F'(x) = \lim_{h\to 0} \frac{\mu_F\big(E_h(x)\big)}{\lambda\big(E_h(x)\big)} = f(x) \ \ \textit{para casi todo } x \in (a,b),$$

con $E_h = (x-h, x]$ ó $(x, x+h]$. Queda claro que $\mu_F \perp \lambda$ si y sólo si $F' = f = 0$ c.d.s. Por otro lado $\mu_F \ll \lambda$ si y sólo si $d\mu_F = F' d\lambda$ si y sólo si $F(x) - F(a) = \mu_F\big((a,x]\big) = \int_{(a,x]} F' d\lambda$ para todo $x \in (a,b)$. $\square$

Para finalizar esta sección introducimos un concepto que nos dará la conexión entre las funciones de variación acotada y el teorema fundamental del cálculo:

Definición 4.21. Diremos que $F : [a,b] \to \mathbb{C}$ es absolutamente continua si para todo $\epsilon > 0$ existe $\delta > 0$ tal que si $(x_1,y_1), \ldots, (x_n,y_n) \subset (a,b)$ son intervalos disjuntos tales que $\sum_{i=1}^{n}(y_i - x_i) < \delta$, entonces

$$\sum_{i=1}^{n} |F(y_i) - F(x_i)| < \epsilon.$$

Nótese que esta definición sigue siendo válida para $F : (-\infty,\infty) \to \mathbb{C}$.
La continuidad absoluta es claramente más fuerte que la continuidad uniforme. En particular las funciones Lipschitz son absolutamente continuas. A continuación presentamos una caracterización para funciones de variación acotada.

Proposición 4.22. *Sea λ la medida de Lebesgue en $\big((a,b), \mathcal{B}(a,b)\big)$ y $F \in BV(a,b)$ continua por derecha. Entonces F es absolutamente continua si y sólo si $\mu_F \ll \lambda$.*

Demostración. Supongamos que F es absolutamente continua. Dado $E \in \mathcal{B}(a,b)$ tal que $\lambda(E) = 0$ queremos probar que $\mu_F(E) = 0$. Dado $\epsilon > 0$ tomemos $\delta > 0$ correspondiente a la definición de continuidad absoluta. Por regularidad por fuera existen abiertos $U_k \subset (a,b)$ tales que $E \subset U_{k+1} \subset U_k$, $\lambda(U_1) < \delta$ y $\mu_F(U_k) \to \mu_F(E)$. Por el lema 2.47 podemos descomponer a cada U_k como una unión numerable de intervalos disjuntos $(x_i^k, y_i^k]$. En consecuencia para todo $k, n \in \mathbb{N}$,

$$\sum_{i=1}^{n}(y_i^k - x_i^k) \leq \lambda(U_k) < \delta \quad y \quad \sum_{i=1}^{n} \left|\mu_F\big((x_i^k, y_i^k]\big)\right| = \sum_{i=1}^{n} |F(y_i^k) - F(x_i^k)| < \epsilon.$$

Por lo tanto para todo $k \in \mathbb{N}$, $|\mu_F(U_k)| \leq \sum_{i=1}^{\infty} \left|\mu_F\big((x_i^k, y_i^k]\big)\right| \leq \epsilon$ que a su vez implica $|\mu_F(E)| \leq \epsilon$. Como ϵ es arbitrario concluimos que $\mu(E) = 0$. El recíproco es consecuencia directa de la proposición 3.16. $\square$

Lema 4.23. *Sea $[a,b]$ un intervalo acotado y $F : [a,b] \to \mathbb{C}$ absolutamente continua, entonces $F \in BV(a,b)$.*

Demostración. Para $\epsilon = 1$ tomemos una $\delta > 0$ que corresponda a la definición de continuidad absoluta. Podemos particionar al intervalo $[a,b]$ en una cantidad finita de intervalos $[c_1, d_1], \ldots, [c_N, d_N]$ de longitud menor a δ. Si $a < x_0 < \cdots < x_n \leq b$ podemos refinar esta partición para que incluya los puntos c_j, d_j y así vemos que

$$\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| = \sum_{j=1}^{N} \sum_{c_j < x_i \leq d_j} |F(x_i) - F(x_{i-1})| < N.$$

De aquí podemos deducir que $T_F(b) \leq N$. $\square$

Teorema 4.24. *(Teorema Fundamental del Cálculo para la medida de Lebesgue) Sea $[a,b]$ un intervalo acotado y $F : [a,b] \to \mathbb{C}$. Las siguientes son equivalentes:*

1) F es absolutamente continua.

2) $F(x) = F(a) + \int_{(a,x]} f \, d\lambda$ para alguna $f \in L^1\big([a,b], \lambda\big)$.

3) F es diferenciable c.d.s., $F' \in L^1\big([a,b], \lambda\big)$ y $F(x) = F(a) + \int_{(a,x]} F' \, d\lambda$

Demostración. Comencemos probando que 1) implica 3). Por el lema anterior $F \in BV(a,b)$. Luego, por las proposiciones 4.20 y 4.22 tenemos $\mu_F \ll \lambda$, $F' \in L^1\big([a,b], \lambda\big)$ y $F(x) = F(a) + \int_{(a,x]} F' \, d\lambda$. 3) implica 2) es trivial. Para probar 2) implica 1) notamos que F es de variación acotada y $\mu_F\big((a,x]\big) = \int_{(a,x]} f \, d\lambda$ para todo $x \in [a,b]$. Por el lema de clases monótonas $d\mu_F = f \, d\lambda$. Como $\mu_F \ll \lambda$, de la proposición 4.22 podemos concluir que F es absolutamente continua. $\square$

Capítulo 5

Teoría de la Probabilidad

En esencia los fundamentos de la teoría de probabilidad moderna, axiomatizada por Andrey Kolmogorov en su libro "*Grundbegriffe der Wahrscheinlichkeitsrechnung*" publicado durante el año 1933, son los mismos que los de la teoría de la medida; sin embargo, el paradigma bajo el que se labora en estas áreas es ligeramente distinto. Toda la teoría desarrollada hasta aquí se enfoca en los espacios de medida y las funciones medibles. Por su parte, en la teoría de probabilidad, el espacio de medida $(\Omega, \mathcal{F}, \mathbb{P})$ será la base sobre la que se presentan los objetos estudiados por aquella y no será el objeto de interés *per se*.

5.1. Elementos aleatorios

Un **espacio de probabilidad** es un espacio de medida $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $\mathbb{P}(\Omega) = 1$. En este caso decimos que $\mathbb{P}$ es una **medida de probabilidad** y a la medida de un conjunto $A \in \mathcal{F}$, $\mathbb{P}(A)$, se le llama la *probabilidad* de A.

Para nosotros una **variable aleatoria** será cualquier función $X : \Omega \to E$, donde $(E, \mathcal{E})$ es un espacio medible, que sea $(\mathcal{F}, \mathcal{E})$-medible. Si X es una variable aleatoria entonces $X^{-1}(A) \in \mathcal{F}$ para cualquier $A \in \mathcal{E}$, y podremos considerar las probabilidades asociadas $\mathbb{P}(X^{-1}(A))$. Por la proposición 2.37 sabemos que ésta es la medida de probabilidad inducida por X, denotada aquí por $\mathbb{P}_X$, a la cual llamaremos **ley**, o **distribución**, de X. Aquí se presenta el primer cambio entre la teoría de probabilidad y la teoría de medida, pues usaremos el término *distribución* como sinónimo de medida de probabilidad, incluso cuando no se haya introducido un elemento aleatorio que la genere. Así vemos que el espacio $(\Omega, \mathcal{F}, \mathbb{P})$ no es más que un artefacto que no tiene un rol predominante.

Como estaremos trabajando con las distribuciones de variables aleatorias se acostumbra utilizar notación que nos permita reducir escritura y que sea intuitiva, por ejemplo:

1) Dados $A \in \mathcal{E}$ y X una variable aleatoria,

$$\{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\}, \qquad \mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

2) Si X es una variable aleatoria que toma valores en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, o $\big(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\big)$,

$$\{X \leq a\} := \{\omega \in \Omega : X(\omega) \leq a\}, \qquad \mathbb{P}_X\left((-\infty, a]\right) = \mathbb{P}(X \leq a).$$

Otro cambio en la notación surge cuando X cumple una propiedad $\mathbb{P}$ casi donde sea. En vez, se acostumbra decir que X cumple $\mathbb{P}$ casi seguramente (abreviado c.s.).

Por otra parte, si X es una variable aleatoria con valores en $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ o $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$ entonces definimos la *esperanza*, *valor esperado* o *medida* de X con respecto a $\mathbb{P}$ por

$$\mathbb{E}[X] = \mathbb{E}^{\mathbb{P}}[X] = \int X d\,\mathbb{P},$$

que no es más que la integral de X con respecto a $\mathbb{P}$ en el sentido de teoría de la medida.

El relacionar los conceptos de probabilidad con aquellos de teoría de la medida nos permite obtener resultados para probabilidad directamente de los resultados análogos de teoría de la medida. Un primer ejemplo de un teorema que se tropicaliza en probabilidad es aquel conocido como la *ley del estadístico incosciente*, que no es más que el teorema 2.38: si $X : \Omega \to E$ es una variable aleatoria $(\mathcal{F}, \mathcal{E})$-medible y $g : E \to \mathbb{R}$ una función medible, entonces $g(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ si y sólo si $g \in L^1(E, \mathcal{E}, \mathbb{P}_X)$, en cuyo caso

$$\mathbb{E}^{\mathbb{P}}[g(X)] = \int g \, d\,\mathbb{P}_X = \mathbb{E}^{\mathbb{P}_X}[g].$$

Otro ejemplo de teorema que se particulariza a probabilidad es la caracterización de las distribuciones sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ por medio de *funciones de distribución*.

Definición 5.1. Una **función de distribución** es una función $F : \mathbb{R} \to [0,1]$ que es no-decreciente, continua por la derecha y

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = 1.$$

Por el teorema 1.24 tenemos que F induce una medida de probabilidad, μ_F, ya que

$$\mu_F(\mathbb{R}) = \lim_{n\to\infty} \mu_F\left((-n, n]\right) = \lim_{n\to\infty} \left(F(n) - F(-n)\right) = 1.$$

Considerando $X = \mathrm{Id}$ sobre $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$ vemos que $\mathbb{P}_X = \mu_F$ y que $\mathbb{P}(X \leq x) = F(x)$.

Por otra parte, si μ es una distribución sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ entonces $F(x) = \mu((-\infty, x])$ será una función de distribución tal que $\mu_F = \mu$. Si en particular $\mu = \mathbb{P}_X$ para alguna variable aleatoria X, nuevamente vemos que $\mathbb{P}_X = \mu_F$. En este último caso diremos que F es la función de distribución asociada a X y se denotará por F_X.

Ejemplos

1. Sea $x_0 \in \mathbb{R}$ fijo, entonces

$$F_1(x) = \mathbb{1}_{[x_0,\infty)}(x) = \begin{cases} 0 & \textit{si } x < x_0, \\ 1 & \textit{si } x \geq x_0, \end{cases}$$

es claramente una función de distribución. Además tendremos que

$$\begin{aligned} \mathbb{P}_1(\{x_0\}) &= \mathbb{P}_1\left(\bigcap_{n\in\mathbb{N}} (x_0 - 1/n, x_0]\right) = \lim_{n\to\infty} \mathbb{P}_1\left((x_0 - 1/n, x_0]\right) \\ &= \lim_{n\to\infty} F_1(x_0) - F_1(x_0 - 1/n) = 1, \end{aligned}$$

por lo que la medida inducida por F_1 es δ_{x_0}.

2. Si ahora consideramos $\{x_1 < \cdots < x_n\} \subset \mathbb{R}$ y $\{a_i\}_{1\leq i\leq n} \subset \mathbb{R}^+ - \{0\}$ podemos definir

$$F_2(x) = \begin{cases} 0 & \textit{si } x < x_1, \\ \frac{\sum_{i=1}^m a_i}{\sum_{i=1}^n a_i} & \textit{si } x_m \leq x < x_{m+1} \textit{ para } m \in \{1, \dots, n-1\}, \\ 1 & \textit{si } x_n \leq x, \end{cases}$$

que por definición será una función de distribución que podremos reescbribir como

$$F_2(x) = \frac{1}{\sum_{i=1}^{n} a_i} \sum_{j=1}^{n} a_j \mathbb{1}_{[x_j,\infty)}(x).$$

Análogo al inciso anterior, vemos que para $i \in \{1, \ldots, n\}$ se tiene que

$$\mathbb{P}_2(\{x_i\}) = \frac{a_i}{\sum_{j=1}^{n} a_j}$$

y que $\sum_{i=1}^{n} \mathbb{P}_2(\{x_i\}) = 1$, de donde la medida inducida por F_2 estará dada por

$$\frac{1}{\sum_{i=1}^{n} a_i} \sum_{j=1}^{n} a_j \delta_{x_j}$$

Este tipo de función de distribución es la que subyace a distribuciones comunes en probabilidad como son: *binomial*, *beta-binomial*, *uniforme discreta* e *hipergeométrica*.

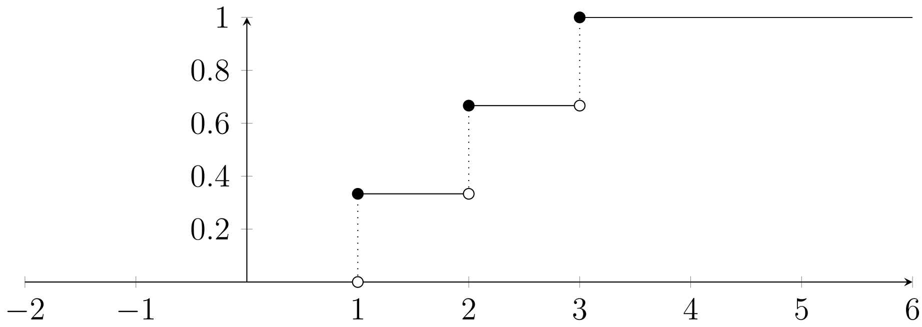

Figura 5.1: Distribución uniforme discreta con $\{x_1, x_2, x_3\} = \{1, 2, 3\}$.

3. Ahora consideraremos $\{x_n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}$ tal que $x_n \neq x_m$ cuando $n \neq m$ y $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^+ - \{0\}$ tal que $\sum_{n=1}^{\infty} a_n = 1$. Definimos

$$F_3(x) = \sum_{n=1}^{\infty} a_n \mathbb{1}_{[x_n,\infty)}(x).$$

Para ver que es una función de distribución primero notemos que para $\epsilon > 0$ dada existe $N \in \mathbb{N}$ tal que $\left|\sum_{j=1}^{n} a_j - \sum_{j=1}^{\infty} a_j\right| < \epsilon$ cuando $n \geq N$, y al considerar $x \geq \text{máx}\{x_1, \ldots, x_N\}$ se sigue que $|F_3(x) - 1| < \epsilon$, por lo que $\lim_{x\to\infty} F_3(x) = 1$. Por la misma razón, si $x < \text{mín}\{x_1, \ldots, x_N\}$ tendremos que $0 \leq F_3(x) < \epsilon$, de donde $\lim_{x\to-\infty} F_3(x) = 0$. Que es continua por la derecha y no decreciente es evidente. Este tipo de funciones de distribución se hacen presentes en las distribuciones *Poisson* y *binomial negativa*.
4. Si consideramos $f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ con $f \geq 0$ y $\int f d\lambda = 1$, entonces sabemos que $\nu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ dada por $\nu(A) = \int_A f d\lambda$ es una medida de probabilidad y se puede verificar fácilmente que su función de distribución asociada está dada por $F(x) = \int_{(-\infty,x]} f d\lambda$. Esto cubre la mayoría de distribuciones que se ven en los cursos básicos de probabilidad como: *gamma*, *beta*, *uniforme*, etcétera. Algunos ejemplos de este caso son:

I. $f(x) = \frac{r(x-a)^{r-1}}{(b-a)^r}\mathbb{1}_{[a,b]}(x)$ para $a, b \in \mathbb{R}$ y $r \geq 0$ con $a < b$. La función de distribución está dada por

$$F(x) = \begin{cases} 0 & \textit{si } x < a, \\ \frac{(x-a)^r}{(b-a)^r} & \textit{si } a \leq x < b, \\ 1 & \textit{si } b \leq x. \end{cases}$$

Se ilustran algunos casos en la figura 5.2.

II. $f(x) = \lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}^+}(x)$ con $\lambda > 0$. En este caso la función de distribución está dada por

$$F(x) = \begin{cases} 0 & \textit{si } x < 0, \\ 1 - e^{-\lambda x} & \textit{si } x \geq 0. \end{cases}$$

Si una variable aleatoria tienen función de distribución de este tipo, se dice que tiene distribución *exponencial*.

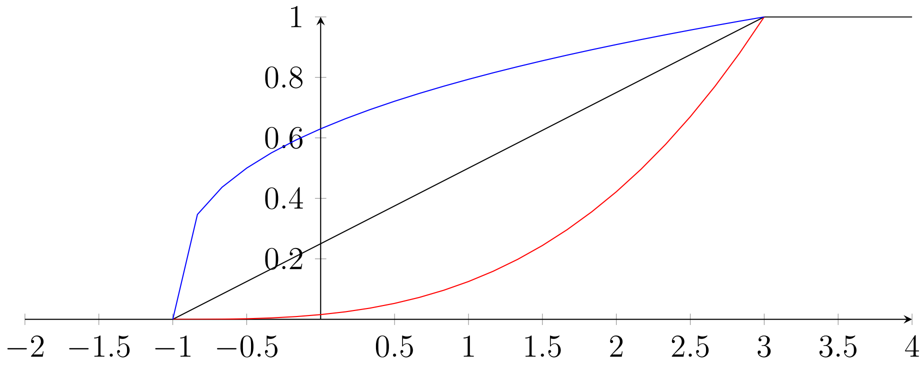

Figura 5.2: Ejemplos de F asociada a $f(x) = \frac{r(x-a)^{r-1}}{(b-a)^r}\mathbb{1}_{[a,b]}(x)$ para $a = -1, b = 3$. En azul $r = 1/3$, en rojo $r = 3$ y en negro $r = 1$.

5. Sea $f_0 : [0,1] \to [0,1]$ el mapeo identidad y definamos de forma recursiva

$$f_{n+1}(x) = \begin{cases} \frac{f_n(3x)}{2} & \textit{si } 0 \leq x < 1/3, \\ \frac{1}{2} & \textit{si } \frac{1}{3} \leq x \leq 2/3, \\ \frac{1}{2} + \frac{f_n(3x-2)}{2} & \textit{si } 2/3 < x \leq 1. \end{cases}$$

Es fácil ver que $\sup_{x\in[0,1]}|f_2(x) - f_1(x)| \leq 2^{-1}\sup_{x\in[0,1]}|f_1(x) - f_0(x)|$ y de forma inductiva se tiene que $\sup_{x\in[0,1]}|f_{n+1}(x) - f_n(x)| \leq 2^{-1}\sup_{x\in[0,1]}|f_n(x) - f_{n-1}(x)|$ para $n \in \mathbb{N}$. Así, para $n < m$ y $x \in [0,1]$, se tiene que

$$\begin{aligned} |f_n(x) - f_m(x)| \leq \sum_{k=0}^{m-n-1}|f_{n+k}(x) - f_{n+k+1}(x)| &\leq \sum_{k=0}^{m-n-1} 2^{-k}|f_n(x) - f_{n+1}(x)| \\ = 2|f_n(x) - f_{n+1}(x)|(1 - 2^{n-m}) \leq 2|f_n(x) - f_{n+1}(x)| &\leq 2^{-n+1}|f_0(x) - f_1(x)|, \end{aligned}$$

de donde $\sup_{x\in[0,1]}|f_n(x) - f_m(x)| \leq 2^{-n+1}|f_0(x) - f_1(x)|$ y así, $\{f_n\}$ es una sucesión uniformemente de Cauchy y por tanto es uniformemente convergente. El límite de esta sucesión, denotado por c y conocido como *función de Cantor*, es tal que es no decreciente, continuo, $c(0) = 0$ y $c(1) = 1$. Por lo tanto si definimos

$$F(x) = \begin{cases} 0 & \textit{si } x < 0, \\ c(x) & \textit{si } 0 \leq x < 1, \\ 1 & \textit{si } 1 \leq x, \end{cases}$$

ésta será una función de distribución, conocida como la *escalera del diablo*. Se puede comprobar fácilmente que $F' = 0$ λ-c.d.s., por lo que resulta que esta función de distribución, a pesar de ser continua, no se obtiene como la integral de una función de densidad.

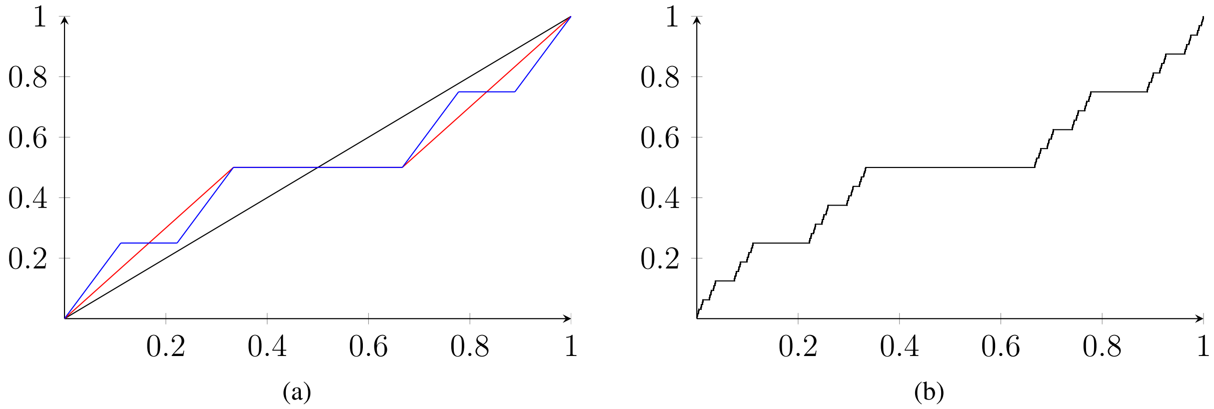

Figura 5.3: En la figura (a) se presentan f_0 en negro, f_1 en rojo y f_2 en azul. En la figura (b) se presenta la función de Cantor.

Ahora consideremos los espacios medibles $(E_i, \mathcal{E}_i)$ y las variables aleatorias $X_i : \Omega \to E_i$ para $1 \leq i \leq n$. Por la proposición 2.7, el mapeo $\omega \mapsto X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ es una nueva variable aleatoria que toma valores en $(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i)$. De manera recíproca, si $X : \Omega \to \prod_{i=1}^n E_i$ es una variable aleatoria entonces $X_i = \pi_i \circ X$ es una variable aleatoria que toma valores en $(E_i, \mathcal{E}_i)$. A X se le suele decir *vector aleatorio* y, así como, una variable aleatoria induce una distribución, también el vector lo hará; para ver que esto es cierto basta considerar $(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i) = (E, \mathcal{E})$. Sin embargo, aquí podremos obtener la ley, no solamente del vector, sino de cada una de sus componentes. Las leyes $\mathbb{P}_{X_i}$ serán llamadas las *leyes marginales* de X; la razón de este nombre es que para todo $A_j \in \mathcal{E}_j$,

$$\mathbb{P}_{X_j}(A_j) = \mathbb{P}(X_j \in A_j) = \mathbb{P}(X_j \in A_j, X_i \in E_i \textit{ para } i \neq j) = \mathbb{P}_X \left(A_j \times \prod_{i \neq j} E_i \right).$$

Otra de las estructuras importantes de variable aleatoria es aquella de *proceso estocástico*. Consideremos un conjunto T no vacío y el espacio de funciones de T a E: $E^T = \prod_{t \in T} E$. Una σ-álgebra en este conjunto es $\mathcal{E}^T = \bigotimes_{t \in T} \mathcal{E}$, la cual es generada por los mapeos de evaluación: $\pi_t : x \mapsto x_t$. Un proceso estocástico con espacio de estados E y trayectorias en E^T será cualquier variable aleatoria $X : \Omega \to E^T$. En este caso tendremos que $X_t = \pi_t \circ X$ será una variable aleatoria $(\mathcal{F}, \mathcal{E})$-medible para cada $t \in T$, por lo que podemos recuperar la definición usual de proceso estocástico como colección indexada de variables aleatorias E-valuadas: $\{X_t : t \in T\}$. Ahora, así como podemos introducir la ley del proceso X, también podremos introducir las distribuciones finito-dimensionales de X: si $t_1, \ldots t_n \in T$ con $n \in \mathbb{N}$ definimos la distribución finito dimensional de X en $t_1, \ldots t_n$ por

$$\mathbb{P}_{t_1, \ldots, t_n} = \mathbb{P} \circ (X_{t_1}, \ldots, X_{t_n})^{-1} = \mathbb{P}_X \circ (\pi_{t_1}, \ldots, \pi_{t_n})^{-1}.$$

El siguiente resultado nos dirá que la ley de un proceso estocástico está determinada por sus distribuciones finito dimensionales. Para introducirlo, diremos que dos variables aleatorias X y Y son iguales en distribución, denotado por $X \stackrel{d}{=} Y$ si y sólo si $\mathbb{P}_X = \mathbb{P}_Y$.

Proposición 5.2. *Sea $(E, \mathcal{E})$ un espacio medible y T un conjunto de índices. Si X y Y son procesos estocásticos con espacio de estados E y trayectorias en E^T, entonces $X \stackrel{d}{=} Y$ si y sólo si para cualquier $n \in \mathbb{N}$ y cualesquiera $t_1, \ldots, t_n \in T$,*

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (Y_{t_1}, \ldots, Y_{t_n}).$$

Demostración. La suficiencia es evidente de la definición de las distribuciones finito-dimensionales. Para la parte necesaria de la prueba, definamos $\mathcal{D} = \{U \in \mathcal{E}^T : \mathbb{P}_X(U) = \mathbb{P}_Y(U)\}$ y $\mathcal{C}$ como la clase de todos los conjuntos de la forma

$$\bigcap_{i=1}^{n} \pi_{t_i}^{-1}(F_{t_i}) = \prod_{i=1}^{n} F_{t_i} \times \prod_{t \in T - \{t_1, \ldots, t_n\}} E,$$

con $n \in \mathbb{N}$, $t_1, \ldots, t_n \in T$ y $F_{t_i} \in \mathcal{E}$. Es fácil ver que $\mathcal{C}$ es un π-sistema y que $\mathcal{D}$ es un λ-sistema. Por hipótesis tendremos que $\mathcal{C} \subset \mathcal{D}$ y por el teorema de clases monótonas se concluye que $\sigma(\mathcal{C}) \subset \mathcal{D} \subset \mathcal{E}^T$, pero $\mathcal{E}^T = \sigma(\mathcal{C})$, lo que nos permite conlcuir que, en efecto, $X \stackrel{d}{=} Y$. □

Finalmente enunciamos y probamos un resultado que es la forma funcional del teorema 1.11, el cual será útil para probar que una propiedad se cumple para una clase de funciones.

Definición 5.3. Sea $\mathcal{H}$ una colección de funciones de un espacio Ω a $\overline{\mathbb{R}}$. Denotamos por $\mathcal{H}^+$ y $\mathcal{H}^b$ por la sucolección de funciones positivas y acotadas de $\mathcal{H}$ de forma respectiva. Diremos que $\mathcal{H}$ es una **clase monótona de funciones** si se cumplen las siguientes condiciones:

I. $1 \in \mathcal{H}$.

II. $f, g \in \mathcal{H}^b$ y $a, b \in \mathbb{R}$ implica que $af + bg \in \mathcal{H}$.

III. $\{f_n\}_{n \in \mathbb{N}} \subset \mathcal{H}^+$ es tal que $f_n \nearrow f$ implica que $f \in \mathcal{H}$.

Teorema 5.4 (de clases monótonas funcional)**.** *Sea $\mathcal{H}$ una clase monótona de funciones de un espacio medible $(\Omega, \mathcal{F})$. Sea $\mathcal{C}$ un π-sistema tal que $\sigma(\mathcal{C}) = \mathcal{F}$. Si $\mathbb{1}_A \in \mathcal{H}$ para cada $A \in \mathcal{C}$ entonces $L^+(\Omega, \mathcal{F}) \subset \mathcal{H}$ y además las funciones $\mathcal{F}$-medibles acotadas también pertenecen a $\mathcal{H}$.*

Demostración. Definamos $\mathcal{D} = \{A \in \mathcal{F} : \mathbb{1}_A \in \mathcal{H}\}$; notemos que por el inciso I de la definición 5.3 $\Omega \in \mathcal{D}$ pues $\mathbb{1}_\Omega = 1 \in \mathcal{D}$; si $A, B \in \mathcal{D}$ tales que $A \subset B$ entonces $\mathbb{1}_{B-A} = \mathbb{1}_B - \mathbb{1}_A \in \mathcal{H}$ por el inciso II de la definición 5.3 y por tanto $B - A \in \mathcal{D}$; finalmente si $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{D}$ es creciente entonces al definir $A = \bigcup_{n=1}^{\infty} A_n$, el inciso III de la definición 5.3, $\mathbb{1}_A = \lim_{n \nearrow \infty} \mathbb{1}_{A_n} \in \mathcal{H}$ por lo que $A \in \mathcal{H}$. Acabamos de probar que $\mathcal{D}$ es un λ-sistema y por hipótesis $\mathcal{C} \subset \mathcal{D}$. Por el teorema de clases monótonas esto implica que $\mathcal{F} = \sigma(\mathcal{C}) \subset \mathcal{D}$. Por lo tanto $\mathbb{1}_A \in \mathcal{H}$ para cada $A \in \mathcal{F}$ y por el inciso II de la definición de clase monótona, se sigue que las funciones simples están en $\mathcal{H}$.

Ahora consideremos $f \in L^+(\Omega, \mathcal{F})$, por el teorema 2.18 sabemos que existe una sucesión de funciones simples $\{\varphi_n\}_{n \in \mathbb{N}} \subset \mathcal{H}^+$ tal que $\varphi_n \nearrow f$. Por el inciso III de la definición de clase monótona $f \in \mathcal{H}$. Finalmente para f acotada, tendremos que $f^+, f^- \in \mathcal{H}^b$ y por el inciso II de la definición de clase monótona se sigue que $f^+ - f^- = f \in \mathcal{H}$. □

5.2. Independencia

El primer concepto propio de teoría de probabilidad que no se tiene como tal en teoría de la medida es el concepto de *independencia*. Intuitivamente podemos pensar en que dos objetos independientes si la ocurrencia de uno de ellos no modifica nuestra información sobre la ocurrencia del otro. Matemáticamente esto se traduce de la siguiente manera: dos eventos $A, B \in \mathcal{F}$ son *independientes* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

La generalización de este concepto a una cantidad finita de eventos no es tan directa; diremos que $A_1, \dots, A_n \in \mathcal{F}$, con $n \in \mathbb{N}$ es independiente si para cada $I \subset \{1, \dots, n\}$ se cumple que

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Es decir que requerimos el cumplimiento de $\sum_{j=2}^{n} \binom{n}{j} = 2^n - n - 1$ igualdades. Puede darse el caso que unas se cumplan y otras no, como se verá en los siguientes ejemplos:

1) Consideremos tres lanzamientos de una moneda justa. Definamos A_1 como el evento en que el primer y el segundo lanzamiento coincidan, A_2 como el evento de que el primer y tercer lanzamiento coincidan y A_3 como el evento en el que el segundo y tercer lanzamiento coincidan. Es fácil ver que $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) = 1/4$ con $i \neq j$ pero $1/4 = \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_2 \cap A_3) \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3) = 1/8$ pues $A_1 \cap A_2 \subset A_3$.

2) Consideremos un lanzamiento de un dado justo con $\Omega = \{1, 2, 3, 4, 5, 6\}$, definamos $A = \{1, 2, 3, 4\}$ y $A_2 = A_3 = \{4, 5, 6\}$. Entonces $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 1/6 = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$ pero $1/6 = \mathbb{P}(A_1 \cap A_2) \neq \mathbb{P}(A_1)\mathbb{P}(A_2) = 1/3$ y $1/2 = \mathbb{P}(A_2 \cap A_3) \neq \mathbb{P}(A_2)\mathbb{P}(A_3) = 1/4$.

Nota: Para una colección finita de eventos $A_1, \dots, A_n \in \mathcal{F}$ podremos reescribir la condición de independencia como sigue: si $B_1, \dots, B_n$ son conjuntos tales que para $i \in \{1, \dots, n\}$ o $B_i = A_i$ o $B_i = \Omega$ entonces

$$\mathbb{P}\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} \mathbb{P}(B_i),$$

pues podremos ignorar $B_i = \Omega$ en la intersección y $\mathbb{P}(\Omega) = 1$ en el producto.

Una colección arbitraria de eventos $\{A_\alpha\}_{\alpha \in \Lambda} \subset \mathcal{F}$ es independiente si toda subcolección finita lo es.

Cuando se habla de independencia en probabilidad es usual que se hable de ésta entre variables aleatorias y no solamente para eventos de una σ-álgebra. Para poder introducir esta noción de independencia, primero extendemos la noción de independencia a clases $\mathcal{A}_1, \dots, \mathcal{A}_n \subset \mathcal{F}$, las cuales diremos son independientes si para cualquier elección de eventos $A_i \in \mathcal{A}_i$ con $i \in \{1, \dots, n\}$, la colección $A_1, \dots, A_n$ es independiente. Eso se extiende a cualquier colección de clases $\{\mathcal{A}_\alpha\}_{\alpha \in \Lambda}$, que diremos es independiente si cada colección de eventos $\{A_\alpha \in \mathcal{A}_\alpha : \alpha \in \Lambda\}$ es independiente. Por cómo se definió la independencia para colecciones arbitrarias de conjuntos es inmediato que podremos asegurar la independencia de $\{\mathcal{A}_\alpha\}_{\alpha \in \Lambda}$ si para cualquier $F \subset \Lambda$ finito, $\{\mathcal{A}_\alpha\}_{\alpha \in F}$ es independiente. Si además $\mathcal{A}_\alpha$ es un π-sistema para cada $\alpha \in \Lambda$ entonces podremos extender la independencia a las σ-álgebras generadas.

Proposición 5.5. *Sea $\mathcal{A}_\alpha \subset \mathcal{F}$ un π-sistema para cada $\alpha \in \Lambda$ tal que $\{\mathcal{A}_\alpha\}_{\alpha \in A}$ sea independiente. Entonces $\{\sigma(\mathcal{A}_\alpha)\}_{\alpha \in A}$ es un sistema independiente.*

Demostración. Consideremos $\alpha_1, \dots, \alpha_n \in \Lambda$ arbitrarios y notemos que para cualesquiera $A_{\alpha_i} \in \mathcal{A}_{\alpha_i}$ se tiene que

$$\mathbb{P}\left(\bigcap_{i\in I} A_{\alpha_i}\right) = \prod_{i\in I} \mathbb{P}(A_{\alpha_i}) \tag{5.1}$$

con $I \subset \{1, \dots, n\}$. Fijemos $A_{\alpha_j} \in \mathcal{A}_{\alpha_j}$ para $j \in \{2, \dots, n\}$ y definamos $\mathcal{D}$ como la clase de conjutos de $\mathcal{F}$ tales que (5.1) se cumpla. Por hipótesis $\mathcal{A}_{\alpha_1} \subset \mathcal{D}$ y además $\Omega \in \mathcal{D}$. Si $A, B \in \mathcal{D}$ son tales que $A \subset B$ entonces para cualquier $I \subset \{2, \dots, n\}$ tendremos

$$\begin{aligned}\mathbb{P}\left((B - A) \cap \bigcap_{i\in I} A_{\alpha_i}\right) &= \mathbb{P}\left(B \cap \bigcap_{i\in I} A_{\alpha_i}\right) - \mathbb{P}\left(A \cap \bigcap_{i\in I} A_{\alpha_i}\right) = (\mathbb{P}(B) - \mathbb{P}(A)) \prod_{i\in I} \mathbb{P}(A_{\alpha_i}) \\ &= \mathbb{P}(B - A) \prod_{i\in I} \mathbb{P}(A_{\alpha_i}),\end{aligned}$$

por ser $\mathbb{P}$ medida de probabilidad, de donde $B - A \in \mathcal{D}$. Si además $\{B_n\}_{n\in\mathbb{N}} \subset \mathcal{D}$ es creciente, entonces por propiedades de continuidad de medida, para cualquier $I \subset \{2, \dots, n\}$, se tendrá

$$\begin{aligned}\mathbb{P}\left(\left(\bigcup_{n\in\mathbb{N}} B_n\right) \cap \bigcap_{i\in I} A_{\alpha_i}\right) &= \mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\left(B_n \cap \bigcap_{i\in I} A_{\alpha_i}\right)\right) = \lim_{n\to\infty} \mathbb{P}\left(B_n \cap \bigcap_{i\in I} A_{\alpha_i}\right) \\ &= \lim_{n\to\infty} \mathbb{P}(B_n) \prod_{i\in I} \mathbb{P}(A_{\alpha_i}) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) \prod_{i\in I} \mathbb{P}(A_{\alpha_i}),\end{aligned}$$

de donde $\bigcup_{n\in\mathbb{N}} B_n \in \mathcal{D}$. Por lo tanto $\mathcal{D}$ es un λ-sistema; así para cualesquiera $B_{\alpha_1} \in \sigma(\mathcal{A}_{\alpha_1})$ y $A_{\alpha_i} \in \mathcal{A}_{\alpha_i}$, con $i \in \{2, \dots, n\}$, se satisface (5.1). Procediendo de forma recursiva se concluye que $\{\sigma(\mathcal{A}_{\alpha_i})\}_{i=1}^n$ es un sistema independiente, lo que prueba el resultado. $\square$

Una consecuencia inmediata de la proposición anterior es un lema de agrupamiento de sistemas independientes. Antes de enunciarlo y probarlo, si tenemos una colección de sub-σ-álgebras de $\mathcal{F}$, $\{\mathcal{F}_\alpha\}_{\alpha\in\Lambda}$, denotaremos a la σ-álgebra generada por éstas por

$$\mathcal{F}_\Lambda := \bigvee_{\alpha\in\Lambda} \mathcal{F}_\alpha = \sigma\left(\bigcup_{\alpha\in\Lambda} \mathcal{F}_\alpha\right).$$

Corolario 5.6. *Sea $\{\mathcal{F}_\alpha\}_{\alpha\in\Lambda}$ una colección de sub-σ-álgebras de $\mathcal{F}$ independiente. Si $\mathcal{T}$ es una partición de Λ entonces el sistema de σ-álgebras $\{\mathcal{F}_S\}_{S\in\mathcal{T}}$ es independiente.*

Demostración. Para cada $S \in \mathcal{T}$ definamos $\mathcal{C}_S$ como el conjunto de intersecciones finitas de los elementos de $\bigcup_{\alpha\in S} \mathcal{F}_\alpha$, el cual es un π-sistema. Por hipótesis se tendrá la independencia de $\{\mathcal{C}_S\}_{S\in\mathcal{T}}$; además $\mathcal{F}_S = \sigma(\mathcal{C}_S)$, por lo que el resultado se sigue de la proposición anterior. $\square$

Antes de proseguir, hacemos notar que en los resultados anteriores se habla de varias σ-álgebras a la vez, lo cual es característico de la teoría de probabilidad, pues las sub-σ-álgebras son de vital importancia; por su parte, en la teoría de medida que no está dirigida al estudio de la probabilidad solamente suele considerarse una única σ-álgebra.

Intuitivamente podemos pensar en en una sub-σ-álgebra como información parcial: si para $\omega \in \Omega$ conocemos $\mathbb{1}_A(\omega)$ para toda $A \in \mathcal{A} \subset \mathcal{F}$, entonces nuestro conocimiento de ω será mayor mientras $\mathcal{A}$ sea más grande.

El siguiente resultado es un primer acercamiento a las leyes cero-uno, que nos permiten decidir si un evento tiene probabilidad nula o uno, conocido como los lemas de Borel–Cantelli. El primero de éstos fue probado en el entorno general de teoría de la medida, mientras que el segundo es un resultado parcialmente recíproco, pues necesita de la noción de independencia.

Teorema 5.7 (Lemas de Borel–Cantelli)**.** *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Entonces se cumplen:*

1) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ es tal que $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ entonces $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$.

2) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ es una colección independiente y $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ entonces se tendrá que $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$.

Demostración. 1) Notemos que $\limsup_{n\to\infty} A_n \subset \bigcup_{k\geq m} A_k$ para cada $m \in \mathbb{N}$ y que $\mathbb{P}\left(\bigcup_{k\geq m} A_k\right) \leq \sum_{k\geq m} \mathbb{P}(A_k)$. Por hipótesis para cada $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que para toda $m \geq N$ se cumple que

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \leq \sum_{k\geq m} \mathbb{P}(A_k) < \varepsilon.$$

Se concluye que $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$.

2) Por ser $\mathbb{P}$ medida de probabilidad, basta probar que $\mathbb{P}\left(\liminf_{n\to\infty} A_n^c\right) = 0$, que podremos obtener por σ-subaditividad si probamos que $\mathbb{P}\left(\bigcap_{k\geq m} A_k^c\right) = 0$ para cada $m \in \mathbb{N}$. Usando independencia y el hecho de que $1 - x \leq e^{-x}$, tenemos que

$$\mathbb{P}\left(\bigcap_{k=m}^{m+j} A_k^c\right) = \prod_{k=m}^{m+j} (1 - \mathbb{P}(A_k)) \leq \exp\left\{-\sum_{k=m}^{m+j} \mathbb{P}(A_k)\right\}.$$

Tomando el límite con respecto a j, usando la continuidad por arriba de las medidas de probabilidad y la hipótesis de divergencia de la serie $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$, concluimos que

$$\mathbb{P}\left(\bigcap_{k\geq m} A_k^c\right) \leq \exp\{-\infty\} = 0,$$

como queríamos probar. □

Antes de introducir la siguiente ley 0-1, definiremos un concepto auxiliar y se dará un resultado elemental ligado a éste. Diremos que una σ-álgebra $\mathcal{F}$ es $\mathbb{P}$-trivial si para todo elemento $A \in \mathcal{F}$, $\mathbb{P}(A) = 0$ o $\mathbb{P}(A) = 1$.

Proposición 5.8. *Una σ-álgebra $\mathcal{F}$ es $\mathbb{P}$-trivial si y sólo si es independiente de sí misma. En este caso, si $X : \Omega \to E$ es una variable aleatoria, con (E, d) un espacio métrico separable, entonces $X = x$ c.s. para alguna $x \in E$.*

Demostración. Si $\mathcal{F}$ es $\mathbb{P}$-trivial entonces $\mathbb{P}(A \cap B) = \mathbb{P}(A) \wedge \mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)$ para cualesquiera $A, B \in \mathcal{F}$, de donde $\mathcal{F}$ es independiente de sí misma. Recíprocamente notemos que $\mathbb{P}(A) = \mathbb{P}(A \cap A) = (\mathbb{P}(A))^2$, por lo que $\mathcal{F}$ es claramente $\mathbb{P}$-trivial.

Para la segunda aserción, notemos que para cada n podemos particionar a E en una colección contable de conjuntos medibles disjuntos, $\{F_{nj}\}_{j\in\mathbb{N}}$, de diámetro menor que n^{-1}. Por la $\mathbb{P}$-trivialidad de $\mathcal{F}$, tendremos que $\mathbb{P}(X \in F_{nj})$ es cero o uno para toda n y toda j, por lo que para cada n, $X \in F_{nj}$ c.s. para una única $j = j_n$. Así $X \in \bigcap_{n\in\mathbb{N}} F_{nj_n}$ c.s. y el diámetro de $\bigcap_{n\in\mathbb{N}} F_{nj_n}$ es cero, por lo que consiste en un único punto $x \in E$. Por lo tanto $X = x$ c.s. □

Para una sucesión de σ-álgebras $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ definimos las σ-álgebras

$$\mathcal{D}_n = \bigvee_{k\geq n} \mathcal{F}_k \quad \text{y} \quad \mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{D}_n.$$

A $\mathcal{T}$ se le conoce como la σ-álgebra cola.

Teorema 5.9 (Ley 0-1 de Kolmogorov). *Si $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ es una sucesión de σ-álgebras independiente entonces la σ-álgebra cola, $\mathcal{T}$, es $\mathbb{P}$-trivial.*

Demostración. Para cada n, $\mathcal{D}_n$ es independiente de $\mathcal{F}_m$ cuando $m < n$, de donde $\mathcal{D}_n$ es independiente de $\bigvee_{m<n} \mathcal{F}_m$. Así, para $A \in \mathcal{T}$ se sigue que A es independiente de $\mathcal{F}_m$ para toda m y por lo tanto A es independiente de $\bigvee_{n\in\mathbb{N}} \mathcal{F}_n \supset \mathcal{T}$. Por la última contención $\mathcal{T}$ es independiente de sí misma y por lo tanto $\mathcal{T}$ es $\mathbb{P}$-trivial. □

Supongamos que para todo $\alpha \in \Lambda$, $X_\alpha : \Omega \to E_\alpha$ es una variable aleatoria $(\mathcal{F}, \mathcal{E}_\alpha)$-medible. Definimos a la σ-álgebra generada por la familia de variables aleatorias $\{X_\alpha\}_{\alpha\in\Lambda}$ como

$$\sigma(X_\alpha; \alpha \in \Lambda) := \sigma\left(\left\{X_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha,\ \alpha \in \Lambda\right\}\right).$$

Es inmediato que $\sigma(X_\alpha; \alpha \in \Lambda) \subset \mathcal{F}$ es la mínima σ-álgebra en Ω en la que las variables aleatorias $\{X_\alpha\}_{\alpha\in\Lambda}$ son medibles. En el caso de una única variable aleatoria, X, es fácil ver que

$$\sigma(X) = \left\{X^{-1}(E) : E \in \mathcal{E}\right\}.$$

Saber la σ-álgebra generada por variables aleatorias nos permitirá establecer algunas relaciones entre éstas. Por ejemplo, el siguiente resultado nos dirá que una función medible g es función de f si y sólo la σ-álgebra generada por g está contenida en aquella generada por f cuando el contradominio de g es bien portado.

Lema 5.10 (Doob–Dynkin). *Sean $(\mathcal{X}, \Sigma_\mathcal{X})$, $(\mathcal{Y}, \Sigma_\mathcal{Y})$ y $(\mathcal{Z}, \Sigma_\mathcal{Z})$ espacios medibles tales que $\{z\} \in \Sigma_\mathcal{Z}$ para todo $z \in \mathcal{Z}$. Supongamos que $f : \mathcal{X} \to \mathcal{Y}$ es $(\Sigma_\mathcal{X}, \Sigma_\mathcal{Y})$-medible. Sea $\mathcal{A} = \Sigma_\mathcal{Y} \cap f(\mathcal{X})$. Entonces $g : \mathcal{X} \to \mathcal{Z}$ es $(\sigma(f), \Sigma_\mathcal{Z})$-medible si y sólo si existe $h : f(\mathcal{X}) \to \mathcal{Z}$ $(\mathcal{A}, \Sigma_\mathcal{Z})$-medible tal que $g(x) = h(f(x))$ para toda $x \in \mathcal{X}$.*

Demostración. Primero supongamos que existe $h : f(\mathcal{X}) \to \mathcal{Z}$ $(\mathcal{A}, \Sigma_\mathcal{Z})$-medible tal que $g(x) = h(f(x))$ para toda $x \in \mathcal{X}$. Queremos probar que si $B \in \Sigma_\mathcal{Z}$ entonces $g^{-1}(B) \in \sigma(f)$. Para esto primero notemos que $h^{-1}(B) = A \cap f(\mathcal{X})$ para alguna $A \in \Sigma_\mathcal{Y}$ y que $f^{-1}(A) = f^{-1}(A \cap f(\mathcal{X}))$, por lo que $g^{-1}(B) = f^{-1}(h^{-1}(B)) = f^{-1}(A) \in \sigma(f)$.

Recíprocamente supongamos que $g : \mathcal{X} \to \mathcal{Z}$ es $(\sigma(f), \Sigma_\mathcal{Z})$-medible. Para toda $z \in \mathcal{Z}$ podemos definir $C_z = g^{-1}(\{z\}) \in \sigma(f)$; entonces existe $A_z \in \Sigma_\mathcal{Y}$ tal que $f^{-1}(A_z \cap f(\mathcal{X})) = f^{-1}(A_z) = C_z$. Definamos $h(y) = z$ para toda $y \in A_z \cap f(\mathcal{X})$. Para ver que $h : f(\mathcal{X}) \to \mathcal{Z}$ está bien definida notemos que para $z_1 \neq z_2$, tendremos que $C_{z_1} \cap C_{z_2} = \emptyset$ de donde $A_{z_1} \cap A_{z_2} \cap f(\mathcal{X}) = \emptyset$. Ahora, si $g(x) = z$ entonces $x \in C_z = f^{-1}(A_z)$, por lo que $f(x) \in A_z \cap f(\mathcal{X})$ y así $h(f(x)) = z = g(x)$, mostrando que $g = h \circ f$.

Falta ver que h es $(\mathcal{A}, \Sigma_\mathcal{Z})$-medible. Con este fin en mente, consideremos $A \in \Sigma_\mathcal{Z}$ y notemos que $g^{-1}(A) \in \sigma(f)$, por lo que existirá $B \in \Sigma_\mathcal{Y}$ tal que $g^{-1}(A) = f^{-1}(B)$. Si $y \in h^{-1}(A)$ entonces $z = h(y) \in A$ y $y = f(x)$ para alguna $x \in C_z \subset g^{-1}(A) = f^{-1}(B)$. Así $y = f(x) \in B$, por lo que $y \in B \cap f(\mathcal{X})$, implicando que $h^{-1}(A) \subset B \cap f(\mathcal{X})$. Por otra parte, si $y \in B \cap f(\mathcal{X})$ entonces $y = f(x)$ para alguna $x \in f^{-1}(B) = g^{-1}(A)$ y $h(y) = h(f(x)) = g(x) \in A$ por construcción; así $y \in h^{-1}(A)$, probando que $h^{-1}(A) \supset B \cap f(\mathcal{X})$. Por lo tanto $h^{-1}(A) = B \cap f(\mathcal{X}) \in \mathcal{A}$. □

La hipótesis de que los conjuntos unitarios estén en $\Sigma_{\mathcal{Z}}$ se requiere para evitar situaciones patólogicas como en el siguiente ejemplo: si $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$, con $\Sigma_{\mathcal{X}} = \Sigma_{\mathcal{Y}} = \mathcal{B}(\mathbb{R})$ y $\Sigma_{\mathcal{Z}} = \{\emptyset, \mathbb{R}\}$ entonces cualquier función $g : \mathcal{X} \to \mathcal{Z}$ será $\sigma(f)$ medible para cualquier función $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$-medible f. Sin embargo $g(x) = x$ no es función de $f(x) = x^2$.

Finalmente podemos introducir la noción de independencia para variables aletorias: decimos que una colección de variables aleatorias, $\{X_\alpha\}_{\alpha\in\Lambda}$ es *independiente* si el sistema de σ-álgebras $\{\sigma(X_\alpha)\}_{\alpha\in\Lambda}$ es independiente. Con esta definición, obtenemos el siguiente resultado que relaciona la independencia con las medidas producto.

Proposición 5.11. *Sea $X_i : \Omega \to E_i$ una variable aleatoria $(\mathcal{F}, \mathcal{E}_i)$-medible con distribución $\mathbb{P}_{X_i}$, con $i \in \{1, \ldots, n\}$. Entonces $X_1, \ldots, X_n$ son independientes si y sólo si $X = (X_1, \ldots, X_n)$ tiene distribución $\bigotimes_{i=1}^n \mathbb{P}_{X_i} := \mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n}$.*

Demostración. Si la ley de X es $\bigotimes_{i=1}^n \mathbb{P}_{X_i}$, es claro que $X_1, \ldots, X_n$ son independientes, estableciendo la parte necesaria de la prueba. Para la parte suficiente, si $A_i \in \mathcal{E}_i$ para $i \in \{1, \ldots, n\}$ entonces, por independencia, la probabilidad del rectángulo medible $A = A_1 \times \cdots \times A_n$ está lada por

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i) = \prod_{i=1}^n \mathbb{P}_{X_i}(A_i) = \bigotimes_{i=1}^n \mathbb{P}_{X_i}(A).$$

La igualdad se extiende fácilmente a $\bigotimes_{i=1}^n \mathcal{E}_i$ por medio del lema de clases monótonas. $\square$

Por virtud de la proposición 5.5, en el caso que X_i sea variable aleatoria con valores en $\mathbb{R}^d$ para toda i, para establecer su independencia, basta probar que

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i)$$

para cualesquiera $x_1, \ldots, x_n \in \mathbb{R}^d$. Observemos que tanto en la proposición anterior como hablando de variables aleatorias con valores en $\mathbb{R}^d$, únicamente estamos considerando una cantidad finita de variables aleatorias. Si bien es cierto que se puede extender esta proposición a una cantidad arbitraria de variables aleatorias independientes, el caso general resulta ser un corolario del *teorema de consistencia de Kolmogorov*, el cual no probaremos aquí. Lo que sí probaremos, después de dar una consecuencia del teorema de Tonelli–Fubini, es la existencia de sucesiones de variables aleatorias independientes con valores en $\mathbb{R}$.

Proposición 5.12. *Sean $X : \Omega \to \mathcal{X}$ y $Y : \Omega \to \mathcal{Y}$ variables aleatorias independientes, $(\mathcal{F}, \Sigma_{\mathcal{X}})$ y $(\mathcal{F}, \Sigma_{\mathcal{Y}})$-medibles de forma respectiva. Si $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ es $(\Sigma_{\mathcal{X}} \otimes \Sigma_{\mathcal{Y}}, \mathcal{B}(\mathbb{R}))$-medible y*

$$\int \mathbb{E}[|f(x, Y)|] d\mathbb{P}_X(x) < \infty,$$

entonces $\mathbb{E}[f(X, Y)] = \int \mathbb{E}[f(x, Y)] d\mathbb{P}_X(x)$.

Demostración. Primero supongamos que $f \geq 0$, por el teorema de Tonelli y el teorema 2.38,

$$\begin{aligned} \mathbb{E}[f(X, Y)] &= \int f(x, y) d(\mathbb{P}_X \times \mathbb{P}_Y)(x, y) = \int \int f(x, y) d\mathbb{P}_Y(y) d\mathbb{P}_X(x) \\ &= \int \mathbb{E}[f(x, Y)] d\mathbb{P}_X(x). \end{aligned}$$

En el caso general, podemos aplicar el mismo argumento a $|f|$ y así $\mathbb{E}[|f(X,Y)|] < \infty$; el resultado se sigue por el teorema de Fubini. $\square$

En particular, si $f = \mathbb{1}_B$ y $g = \mathbb{1}_{A\times\mathcal{Y}}\mathbb{1}_B$ con $A \in \Sigma_\mathcal{X}$ $B \in \Sigma_\mathcal{X} \otimes \Sigma_\mathcal{Y}$ entonces

$$\mathbb{P}((X,Y) \in B) = \int \mathbb{P}((x,Y) \in B) d\mathbb{P}_X(x) \quad \text{y}$$
$$\mathbb{P}((X,Y) \in B; X \in A) = \int_A \mathbb{P}((x,Y) \in B) d\mathbb{P}_X(x).$$

Definición 5.13. Diremos que una variable aleatoria $U : \Omega \to \mathbb{R}$ es *uniforme* (a,b), denotado $\mathrm{U}(a,b)$, si para todo $A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_U(A) = \int_A \mathbb{1}_{(a,b)} d\lambda$. Una variable aleatoria $Y : \Omega \to \mathbb{R}$ es *Bernoulli* con parámetro p, denotado $\mathrm{Ber}(p)$, si $\mathbb{P}_Y = p\delta_1 + (1-p)\delta_0$.

Para $x \in (0,1)$ definimos la expansión binaria de x como la sucesión $\{d_n\}_{n\in\mathbb{N}}$, con $d_n \in \{0,1\}$, tal que $x = \sum_{n\in\mathbb{N}} d_n 2^{-n}$ y $\sum_{n\in\mathbb{N}} d_n = \infty$. (La última condición se pide para que la expansión sea única.)

Lema 5.14. *Sea $U : \Omega \to \mathbb{R}$ una variable aleatoria con valores en $(0,1)$ y expansión binaria $\{\xi_n\}_{n\in\mathbb{N}}$. Entonces U es $\mathrm{U}(0,1)$ si y sólo si $\{\xi_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes, todas con distribución $\mathrm{Ber}(1/2)$.*

Demostración. Primero supongamos que U se distribuye $\mathrm{U}(0,1)$. Entonces tendremos, para cada n, que $\{\xi_1 = k_1, \dots, \xi_n = k_n\} = \{\sum_{i=1}^n k_i 2^{-i} < U < \sum_{i=1}^n k_i 2^{-i} + 2^{-n}\}$. Así obtenemos, para toda n, $\mathbb{P}(\xi_1 = k_1, \dots, \xi_n = k_n) = 2^{-n}$. Sumando sobre $k_1, \dots, k_{n-1}$ se sigue que $\mathbb{P}(\xi_n = k_n) = 1/2$ para $k_n = 0$ o $k_n = 1$. La independencia se obtiene del mismo cálculo. Por lo tanto $\{\xi_n\}_{n\in\mathbb{N}}$ es una sucesión independiente de variables aleatorias $\mathrm{Ber}(1/2)$.

Si ahora consideramos que $\{\xi_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes con distribución $\mathrm{Ber}(1/2)$, por lo que se acaba de probar, al considerar una variable aleatoria $\mathrm{U}(0,1)$, X, su expansión binaria $\{\eta_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes $\mathrm{Ber}(1/2)$. Entonces $\{\xi_n\}_{n\in\mathbb{N}} \overset{d}{=} \{\eta_n\}_{n\in\mathbb{N}}$. Por lo tanto $U = \sum_{n\in\mathbb{N}} \xi_n 2^{-n} \overset{d}{=} \sum_{n\in\mathbb{N}} \eta_n 2^{-n} = X$. Así U es $\mathrm{U}(0,1)$. $\square$

Lema 5.15. *Existe una sucesión, $\{f_n\}_{n\in\mathbb{N}}$, de funciones de $(0,1)$ a $\mathbb{R}$, medibles, tales que $\{f_n(U)\}_{n\in\mathbb{N}}$ es una sucesión independiente de variables aleatorias $\mathrm{U}(0,1)$ cuando U es $\mathrm{U}(0,1)$.*

Demostración. Para cada $x \in (0,1)$ consideremos la expansión binaria $\{g_n(x)\}_{n\in\mathbb{N}}$. Se puede comprobar que $g_n : (0,1) \to \mathbb{R}$ es medible. Consideremos una biyección $b : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$ y definamos $h_{nm} := g_{b(n,m)}$. Entonces $f_n = \sum_{m\geq 1} h_{nm} 2^{-m}$ es medible y por el lema anterior $\{h_{nm}(U)\}_{n,m\in\mathbb{N}}$ es una colección de variables aleatorias independientes $\mathrm{Ber}(1/2)$ cuando U es $\mathrm{U}(0,1)$. Por el lema de Doob–Dynkin y el corolario 5.6, $\{f_n(U)\}_{n\in\mathbb{N}}$ es una colección de variables aleatorias independientes. Finalmente, por el lema anterior se tiene que $f_n(U)$ es $\mathrm{U}(0,1)$ para cada $n \in \mathbb{N}$. $\square$

Teorema 5.16. *Sea $\{\mu_n\}_{n\in\mathbb{N}}$ una sucesión de medidas de probabilidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Entonces existe una sucesión de variables aleatorias independientes $\{X_n\}_{n\in\mathbb{N}}$, definida en un espacio $(\Omega, \mathcal{F}, \mathbb{P})$, tal que para cada n, μ_n es la ley de X_n.*

Demostración. Por el lema anterior existe una sucesión de variables aleatorias independiente $\mathrm{U}(0,1)$, $\{U_n\}_{n\in\mathbb{N}}$. Para cada $n \in \mathbb{N}$ definimos la función $Q_n : (0,1) \to \mathbb{R}$ por

$$Q_n(x) := \text{ínf}\{y \in \mathbb{R} : x \leq \mu_n((-\infty, y])\} = \sup\{y \in \mathbb{R} : \mu_n((-\infty, y]) < x\},$$

con $x \in (0,1)$, la cual es $(\mathcal{B}((0,1)), \mathcal{B}(\mathbb{R}))$-medible. Por Doob–Dynkin, al definir $X_n := Q_n(U_n)$ para cada n, obtenemos que $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes. Notando que para cada $n \in \mathbb{N}$ y $x \in \mathbb{R}$, $\{Q_n(U_n) \leq x\} = \{U_n \leq \mu_n((-\infty, x])\}$, se sigue que

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(Q_n(U_n) \leq x) = \mathbb{P}(U_n \leq \mu_n((-\infty, x])) = \mu_n((-\infty, x]),$$

lo que prueba que la distribución de X_n es μ_n para toda $n \in \mathbb{N}$. □

Notemos que en esta serie de resultados, solamente se presupone la existencia de variables aleatorias $\mathrm{U}(0,1)$, lo cual es inmediato: si consideramos $a, b \in \mathbb{R}$ con $a < b$ y $(\Omega, \mathcal{F}, \mathbb{P}) = \left((a,b), \mathcal{B}((a,b)), \frac{1}{b-a}\lambda^{(a,b)}\right)$ y $X : (a,b) \to \mathbb{R}$ por $\omega \mapsto \frac{\omega - a}{b-a}$, entonces obtendremos una variable aleatoria $\mathrm{U}(0,1)$.

5.3. Función característica

El objeto de estudio en esta sección se conoce como la función característica o transformada de Fourier de las medidas de probabilidad y distribuciones sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Definición 5.17. Sea μ una medida sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Definimos su **función característica**, $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$, por

$$\hat{\mu}(t) := \int e^{i\langle t,x\rangle} d\mu(x),$$

donde $\langle \cdot, \cdot \rangle$ es el producto interior usual en $\mathbb{R}^d$.

Si X es una variable aleatoria con valores en $\mathbb{R}^d$ y ley $\mathbb{P}_X$, definimos su función característica como la función característica de $\mathbb{P}_X$,

$$\varphi_X(t) = \widehat{\mathbb{P}}_X(t) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right].$$

Notemos que para toda medida de probabilidad, o en general cualquier medida finita, podemos asegurar la existencia de su función característica pues $\left|e^{i\langle t,x\rangle}\right| = 1$ para cualesquiera $x, t \in \mathbb{R}^d$.

Teorema 5.18. *Si $\mu : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ es una medida de probabilidad entonces*

1) $\hat{\mu}$ es uniformemente continua sobre $\mathbb{R}^d$,

2) $\hat{\mu}(0) = 1$,

3) $|\hat{\mu}(t)| \leq 1$,

4) $\hat{\mu}(-t) = \overline{\hat{\mu}(t)}$.

Demostración. 1) Primero notemos que para $t, h \in \mathbb{R}^d$ arbitrarios se tiene

$$\begin{aligned}
|\hat{\mu}(t+h) - \hat{\mu}(t)| &\leq \int \left|e^{i\langle t+h,x\rangle} - e^{i\langle t,x\rangle}\right| d\mu(x) = \int \left|e^{i\langle h,x\rangle} - 1\right| d\mu(x) \\
&= \int \left|e^{i\langle h,x\rangle/2} - e^{-i\langle h,x\rangle/2}\right| = 2\int \left|\operatorname{sen}\left(\tfrac{1}{2}\langle h,x\rangle\right)\right| d\mu(x) \\
&\leq 2\left(\mu\left(\{x \in \mathbb{R}^d : \|x\|_2 \geq M\}\right) + \int_{\{x\in\mathbb{R}^d:\|x\|_2<M\}} \left|\operatorname{sen}\left(\tfrac{1}{2}\langle h,x\rangle\right)\right| d\mu(x)\right).
\end{aligned}$$

Consideremos $\epsilon > 0$ arbitrario. Por ser μ finita existirá $M > 0$ tal que

$$\mu(\{x \in \mathbb{R}^d : \|x\|_2 \geq M\}) < \epsilon/4$$

; con M fija, y por continuidad del seno, existirá $\delta > 0$ tal que si $\|h\|_2 < \delta$ entonces

$$\left|\operatorname{sen}\left(\tfrac{1}{2}\langle h,x\rangle\right)\right| < \epsilon/4$$

para cualesquiera $x \in \mathbb{R}^d$ tales que $\|x\|_2 \leq M$. Entonces

$$|\hat{\mu}(t+h) - \hat{\mu}(t)| < 2\left(\frac{\epsilon}{4} + \frac{\epsilon}{4}\right) = \epsilon$$

para toda $t \in \mathbb{R}^d$, probando que $\hat{\mu}$ es uniformemente continua.

2) Se sigue de $\hat{\mu}(0) = \int e^{i\langle 0,x\rangle} d\mu(x) = \mu(\mathbb{R}^d) = 1$.

3) Basta notar que $|\hat{\mu}(t)| \leq \int \left|e^{i\langle t,x\rangle}\right| d\mu(x) = \mu(\mathbb{R}^d) = 1$.

4) Un cálculo directo da

$$\begin{aligned}
\hat{\mu}(-t) &= \int e^{i\langle -t,x\rangle} d\mu(x) = \int \cos(\langle -t,x\rangle) d\mu(x) + i\int \operatorname{sen}(\langle -t,x\rangle) d\mu(x) \\
&= \int \cos(\langle t,x\rangle) d\mu(x) - i\int \operatorname{sen}(\langle t,x\rangle) d\mu(x) = \overline{\int e^{i\langle t,x\rangle} d\mu(x)} \\
&= \overline{\hat{\mu}(t)}.
\end{aligned}$$

□

Teorema 5.19. *Sea $\mu : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ una medida de probabilidad tal que para alguna $m \in \mathbb{N}$, $\int |x_j|^m \, d\mu(x) < \infty$ para cada $j \in \{1,\ldots,d\}$. Entonces existen $\int \prod_{j=1}^d x_j^{k_j} d\mu(x)$ con $k_j \in \mathbb{N} \cup \{0\}$ y $\sum_{j=1}^d k_j \leq m$. Además $\hat{\mu}$ es de clase C^m y para $\alpha \leq m$,*

$$\frac{\partial^\alpha}{\partial t_{j1} \cdots \partial t_{j_\alpha}} \hat{\mu}(t) = i^\alpha \int \prod_{l=1}^{\alpha} x_{j_l} e^{i\langle t,x\rangle} d\mu(x).$$

Demostración. Notemos que si $\int |x_j|^m \, d\mu(x) < \infty$ entonces $\int |x_j|^{m'} \, d\mu(x) < \infty$ para toda $m' \leq m$ por ser μ medida de probabilidad. Por lo tanto basta establecer la primera parte del enunciado para m. Por la desigualdad de Hölder,

$$\left|\prod_{j=1}^d x_j^{k_j}\right| \leq \left(\sum_{j=1}^d |x_j|\right)^m \leq n^{m-1} \sum_{j=1}^d |x_j|^m,$$

con $k_j \in \mathbb{N} \cup \{0\}$ y $\sum_{j=1}^{d} k_j = m$, estableciendo la primera parte.

Para la segunda parte, basta establecer el resultado para $\alpha = 1$ y el caso general se sigue de forma recursiva. Así, primero veamos que para $h \neq 0$, $x \in \mathbb{R}^d$ y $e_j \in \mathbb{R}^d$ el j-ésimo vector canónico,

$$\left| \frac{e^{i\langle he_j, x\rangle} - 1}{h} \right| = \left| \frac{e^{i\langle he_j, x\rangle/2} - e^{-i\langle he_j, x\rangle/2}}{h} \right| = 2 \left| \frac{\operatorname{sen}\left(\frac{1}{2}\langle he_j, x\rangle\right)}{h} \right| \leq \left| \frac{\langle he_j, x\rangle}{h} \right| = |x_j| \,.$$

Por lo tanto, al definir $g_n(t) = \frac{\hat{\mu}(t+h_n e_j) - \hat{\mu}(t)}{h_n}$, con $\{h_n\}_{n\in\mathbb{N}} \subset \mathbb{R} - \{0\}$ tal que $h_n \to 0$, por una parte tendremos que $g_n(t) \to \frac{\partial}{\partial t_j}\hat{\mu}(t)$; por otra parte, debido al teorema de convergencia dominada,

$$\lim_{n\to\infty} g_n(t) = \int \lim_{n\to\infty} \frac{e^{i\langle t+h_n e_j, x\rangle} - e^{i\langle t,x\rangle}}{h_n} d\mu(x) = \int i x_j e^{i\langle t,x\rangle} d\mu(x) = i \int x_j e^{i\langle t,x\rangle} d\mu(x). \quad \square$$

Ahora probaremos que la función característica caracteriza a una medida, es decir que dos medidas serán iguales si y sólo si su función característica es la misma. Para este cometido primero mostraremos un lema, para el cual será necesario definir unos conceptos auxiliares. Si $(\mathcal{X}, \tau)$ es un espacio topológico entonces definimos

$$\begin{aligned} C(\mathcal{X}) &= \{f : \mathcal{X} \to \mathbb{C} : f \textit{ es continua}\}, \\ C(\mathcal{X}, \mathbb{R}) &= \{f : \mathcal{X} \to \mathbb{R} : f \textit{ es continua}\} \qquad \text{y} \\ C_b(\mathcal{X}) &= \{f \in C(\mathcal{X}) : f \textit{ es acotada}\}. \end{aligned}$$

Además definimos el **soporte** de $f \in C(\mathcal{X})$ como la cerradura del conjunto $\{f \neq 0\}$ y se denota por $\operatorname{supp} f$; cuando $\operatorname{supp} f$ sea compacto, decimos que f tiene **soporte compacto** y definimos

$$C_c(\mathcal{X}) = \{f \in C(X) : \operatorname{supp} f \textit{ es compacto}\}.$$

Más aún, decimos que $f \in C(X)$ se **desvanece en el infinito** si para cada $\epsilon > 0$ el conjunto $\{|f| \geq \epsilon\}$ es compacto y $|f| < \epsilon$ en su complemento; definimos

$$C_0(\mathcal{X}) = \{f \in C(X) : f \textit{ se desvanece en el infinito}\}.$$

Notemos que de igual forma podemos definir $C_b(\mathcal{X}, \mathbb{R}), C_c(\mathcal{X}, \mathbb{R})$ y $C_0(\mathcal{X}, \mathbb{R})$. Además es fácil ver que $C_c(\mathcal{X}) \subset C_0(\mathcal{X}) \subset C_b(\mathcal{X})$. Finalmente, podemos definir la norma $\|\cdot\|_u : C_b(\mathcal{X}) \to \mathbb{R}$ por $\|f\|_u := \sup\{|f(x)| : x \in X\}$, llamada la **norma uniforme**.

Lema 5.20. *Si $\mu, \nu : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ son dos medidas de probabilidad entonces son equivalentes los siguientes incisos.*

1) $\mu \equiv \nu$.

2) $\int f d\mu = \int f d\nu$ *para toda* $f \in C_b(\mathbb{R}^d, \mathbb{R})$.

3) $\int f d\mu = \int f d\nu$ *para toda* $f \in C_0(\mathbb{R}^d, \mathbb{R})$.

4) $\int f d\mu = \int f d\nu$ *para toda* $f \in C_c(\mathbb{R}^d, \mathbb{R})$.

Demostración. Es claro que (1) implica (2), que (2) implica (3) y que (3) implica (4) por lo que basta ver que (4) implica (1) para concluir. Consideremos $a, b \in \mathbb{R}^d$ con $a < b$ arbitrarios y definamos $A = \{x \in \mathbb{R}^d : a \leq x \leq b\}$ y $A_n = \left\{x \in \mathbb{R}^d : a - \frac{b-a}{n} \leq x \leq b + \frac{b-a}{n}\right\}$. Para $n \in \mathbb{N}$ sea f_n una función continua tal que $f_n(x) = 1$ si $x \in A$ y $f_n(x) = 0$ si $x \in A_n^c$. Entonces notemos que $\lim_{n\to\infty} f_n = \mathbb{1}_A$ y por el teorema de convergencia dominada y la hipótesis

$$\mu(A) = \int \mathbb{1}_A d\mu = \int \lim_{n\to\infty} f_n d\mu = \int \lim_{n\to\infty} f_n d\nu = \int \mathbb{1}_A d\nu = \nu(A).$$

Como el conjunto de los rectángulos cerrados es un π-sistema que genera $\mathcal{B}(\mathbb{R}^d)$ y μ y ν son finitas, $\mu \equiv \nu$ por el lema de clases monótonas. □

Lema 5.21. *Si $\nu : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ tiene λ^d-densidad $f_\sigma(x) = (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right)$ entonces para todo $t \in \mathbb{R}^d$,*

$$\hat{\nu}(t) = \prod_{j=1}^{d} e^{-\frac{1}{2}\sigma^2 t_j^2} = e^{-\frac{1}{2}\sigma^2\|t\|_2^2}.$$

Además, para $v \in \mathbb{R}^d$ se cumple que

$$f_{\sigma,v}(t) := f_\sigma(t-v) = (2\pi\sigma^2)^{-d/2} \int f_\sigma(x) \exp\left(i\left\langle \frac{t-v}{\sigma^2}, x\right\rangle\right) d\lambda^d(x).$$

Demostración. Para la primera parte del enunciado, consideremos el caso $d = 1$, entonces

$$\hat{\nu}(t) = \int f_\sigma(x) e^{itx} d\lambda(x) = \int f_\sigma(x)\cos(tx) d\lambda(x),$$

por definición de la integral y por ser sen función impar. Por el teorema 5.19,

$$\frac{d}{dt}\hat{\nu}(t) = -\int x f_\sigma(x) \operatorname{sen}(tx) d\lambda(x) = -\int \sigma^2 t f_\sigma(x)\cos(tx) d\lambda(x) = -\sigma^2 t\hat{\nu}(t),$$

donde la segunda igualdad se da por integración por partes. Resolviendo la ecuación diferencial, con la condición $\hat{\nu}(0) = 1$, obtenemos

$$\hat{\nu}(t) = \exp\left(-\frac{1}{2}\sigma^2 t^2\right).$$

El caso general se sigue por inducción y el teorema de Fubini.

Para la segunda parte del enunciado notemos que

$$f_{\sigma,v}(t) = (2\pi\sigma^2)^{-d/2}\hat{\nu}\left(\frac{t-v}{\sigma^2}\right) = (2\pi\sigma^2)^{-d/2} \int f_\sigma(x) \exp\left(i\left\langle \frac{t-v}{\sigma^2}, x\right\rangle\right) d\lambda^d(x). \qquad \square$$

Teorema 5.22. *Si $\mu, \nu : \mathcal{B}(\mathbb{R}^d) \to [0,\infty]$ son medidas de probabilidad entonces $\mu \equiv \nu$ si y sólo si $\hat{\mu} = \hat{\nu}$.*

Demostración. Que $\mu \equiv \nu$ implica que $\hat{\mu} = \hat{\nu}$ es claro. Recíprocamente sea $f_{\sigma,v}$ una función como en el lema anterior. Por el lema anterior y el teorema de Fubini tendremos que

$$\begin{aligned}\int f_{\sigma,v}(t)d\mu(t) &= \int\int(2\pi\sigma^2)^{-\frac{d}{2}}f_\sigma(x)e^{i\frac{\langle t-v,x\rangle}{\sigma^2}}d\lambda^d(x)d\mu(t)\\ &= \int(2\pi\sigma^2)^{-\frac{d}{2}}f_\sigma(x)\int e^{i\frac{\langle t-v,x\rangle}{\sigma^2}}d\mu(t)d\lambda^d(x)\\ &= \int(2\pi\sigma^2)^{-\frac{d}{2}}f_\sigma(x)e^{-i\frac{\langle v,x\rangle}{\sigma^2}}\hat{\mu}\left(\frac{x}{\sigma^2}\right)d\lambda^d(x)\\ &= \int(2\pi\sigma^2)^{-\frac{d}{2}}f_\sigma(x)e^{-i\frac{\langle v,x\rangle}{\sigma^2}}\hat{\nu}\left(\frac{x}{\sigma^2}\right)d\lambda^d(x) = \int f_{\sigma,v}(t)d\nu(t).\end{aligned}$$

Por linealidad la igualdad se extiende al subespacio de $C_0(X,\mathbb{R})$, $\mathcal{H}$, generado por $\{f_{\sigma,v} : \sigma \in \mathbb{R}-\{0\}, v \in \mathbb{R}^d\}$, el cual es una subálgebra que separa puntos. Por el teorema de Stone–Weierstrass $\mathcal{H}$ es denso en $C_0(X,\mathbb{R})$ y por el teorema de convergencia dominada, se extiende la igualdad a $C_0(X,\mathbb{R})$, lo que concluye la demostración por el lema 5.20. □

Corolario 5.23. *Las variables aleatorias $X_1,\dots,X_n$, con valores en $\mathbb{R}$, son independientes si y sólo si $\varphi_X(t) = \prod_{j=1}^n \varphi_{X_j}(t_j)$, con $X = (X_1,\dots,X_n)$ y $t \in \mathbb{R}^n$.*

Demostración. Las variables aleatorias son independientes si y sólo si $\mathbb{P}_X = \mathbb{P}_{X_1}\otimes\cdots\otimes\mathbb{P}_{X_n}$; por Fubini

$$\varphi_X(t) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right] = \prod_{j=1}^n \mathbb{E}\left[e^{it_jX_j}\right] = \prod_{j=1}^n \varphi_{X_j}(t_j).$$

El teorema de unicidad nos permite concluir. □

Proposición 5.24. *Si X es un vector aleatorio con valores en $\mathbb{R}^d$, $x \in \mathbb{R}^m$ y A es una matriz de dimensiones $m \times d$ entonces para cada $t \in \mathbb{R}^m$,*

$$\varphi_{x+AX}(t) = e^{i\langle t,x\rangle}\varphi_X(A^*t).$$

Demostración. Se establece por un cálculo directo:

$$\varphi_{x+AX}(t) = \mathbb{E}\left[e^{i\langle t,x+AX\rangle}\right] = e^{i\langle t,x\rangle}\mathbb{E}\left[e^{i\langle A^*t,X\rangle}\right] = e^{i\langle t,x\rangle}\varphi_X(A^*t).$$ □

Proposición 5.25. *Si $X_1,\dots,X_n$ son vectores aleatorios, con valores en $\mathbb{R}^d$, independientes y $S = \sum_{j=1}^n X_j$, entonces para todo $t \in \mathbb{R}^d$*

$$\varphi_S(t) = \prod_{j=1}^n \varphi_{X_j}(t).$$

Demostración. Procedemos por inducción. Para $n = 2$ tendremos que

$$\varphi_S(t) = \mathbb{E}\left[e^{i\langle t,X_1+X_2\rangle}\right] = \int \mathbb{E}\left[e^{i\langle t,x+X_2\rangle}\right]d\,\mathbb{P}_{X_1}(x) = \int \varphi_{X_2}(t)e^{i\langle t,x\rangle}d\,\mathbb{P}_{X_1}(x) = \varphi_{X_1}(t)\varphi_{X_2}(t).$$

Si suponemos que se cumple para algún $n \in \mathbb{N}$ entonces notemos que $S = S' + X_{n+1}$, donde $S' = \sum_{j=1}^n X_j$ es independiente de X_{n+1}. Así, por el mismo argumento de antes e hipótesis de inducción,

$$\varphi_S(t) = \varphi_{S'}(t)\varphi_{X_{n+1}}(t) = \prod_{i=1}^{n+1}\varphi_{X_i}(t).$$ □

El siguiente objetivo de esta sección será el dar una forma de recuperar la medida cuando conozcamos la función característica asociada. Para este cometido, necesitaremos de un par de lemas.

Lema 5.26. *Para toda* $n \in \mathbb{N} \cup \{0\}$ *y* $t \in \mathbb{R}$*,*

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| \leq \min \left\{ \frac{2\,|t|^n}{n!}, \frac{|t|^{n+1}}{(n+1)!} \right\}.$$

Demostración. El caso en el que $t = 0$ es trivial y se omite. Si $t \neq 0$ y $k \geq 0$, integrando por partes,

$$\int_0^t e^{ix}(t-x)^k dx = \frac{t^{k+1}}{k+1} + \frac{i}{k+1} \int_0^t e^{ix}(t-x)^{k+1} dx.$$

En particular, para $k = 0$, integrando directamente y usando el cálculo anterior,

$$\frac{e^{it} - 1}{i} = \int_0^t e^{ix} dx = t + i \int_0^t e^{ix}(t-x) dx,$$

de donde se sigue que $e^{it} = 1 + it + i^2 \int_0^t e^{ix}(t-x)dx$. Procediendo de forma inductiva se deduce que para $n \in \mathbb{N} \cup \{0\}$,

$$e^{it} = \sum_{k=0}^{n} \frac{(it)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^t e^{ix}(t-x)^n dx. \tag{5.2}$$

De esta misma ecuación se sigue, usando $n - 1$ en lugar de n, que

$$\begin{aligned} e^{it} &= \sum_{k=0}^{n-1} \frac{(it)^k}{k!} + \frac{(it)^n}{n!} - \frac{(it)^n}{n!} + \frac{i^n}{(n-1)!} \int_0^t e^{ix}(t-x) dx \\ &= \sum_{k=0}^{n} \frac{(it)^k}{k!} + \frac{i^n}{(n-1)!} \int_0^t (e^{ix} - 1)(t-x)^{n-1} dx. \end{aligned} \tag{5.3}$$

Si $t > 0$, de (5.2) y (5.3) obtenemos las siguientes cotas:

$$\begin{aligned} \left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| &\leq \frac{1}{n!} \int_0^t \left| e^{ix}(t-x)^n \right| dx = \frac{t^{n+1}}{(n+1)!}, \\ \left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| &\leq \frac{1}{(n-1)!} \int_0^t \left| (e^{ix} - 1)(t-x)^{n-1} \right| dx = \frac{2t^n}{n!}. \end{aligned}$$

Obteniendo las cotas de forma análoga cuando $t < 0$, se obtiene el resultado. $\square$

Lema 5.27. *Para* $\alpha \in \mathbb{R}$ *y* $t > 0$ *se tendrá que*

$$\left| \int_0^t \frac{\operatorname{sen}(\alpha x)}{x} dx \right| \leq \int_0^\pi \frac{\operatorname{sen} x}{x} dx \leq \pi.$$

Si además $\alpha \neq 0$ *entonces*

$$\lim_{t \to \infty} \int_0^t \frac{\operatorname{sen}(\alpha x)}{x} dx = \frac{\pi}{2} \operatorname{sgn}(\alpha).$$

Demostración. Se omite el caso $\alpha = 0$. Es claro que si se establece el resultado para $\alpha > 0$, por ser sen una función impar, se obtendrá el resultado para $\alpha < 0$. Más aún, usando la transformación $y = \alpha x$, basta establecer el resultado para $\alpha = 1$. Dicho esto, notemos que $\left\{\int_{(n-1)\pi}^{n\pi} \frac{\operatorname{sen}(x)}{x}\right\}_{n\in\mathbb{N}}$ es una sucesión que alterna signos y es decreciente en valor absoluto, además de que $\operatorname{sen} x \le x$ cuando $x \ge 0$, por lo que se sigue que para $t > 0$,

$$0 \le \int_0^t \frac{\operatorname{sen} x}{x} dx \le \int_0^\pi \frac{\operatorname{sen} x}{x} dx \le \int_0^\pi dx = \pi,$$

estableciendo la primera afirmación del enunciado.

Recordemos ahora que si $x > 0$, $x^{-1} = \int_0^\infty e^{-xy} dy$. Por Tonelli entonces tendremos que

$$\int_0^t \int_0^\infty e^{-xy} |\operatorname{sen} x|\, dydx = \int_0^t \frac{|\operatorname{sen} x|}{x} dx \le \int_0^t dx = t.$$

Así, por Fubini,

$$\int_0^\infty \int_0^t e^{-xy} \operatorname{sen}(x) dxdy = \int_0^t \frac{\operatorname{sen} x}{x} dx. \tag{5.4}$$

La integral interior se puede calcular por medio de integración por partes,

$$\begin{aligned} \int_0^t e^{-xy} \operatorname{sen}(x) dx &= \left[-e^{-xy} \cos x\right]_0^t - y \int_0^t e^{-xy} \cos(x) dx \\ &= (1 - e^{-ty} \cos t) - \left[y e^{-xy} \operatorname{sen} x\right]_0^t - y^2 \int_0^t e^{-xy} \operatorname{sen}(x) dx \\ &= \frac{1 - e^{-ty}(\cos t + y \operatorname{sen} t)}{1 + y^2}. \end{aligned}$$

Sustituyendo esto último en el lado derecho de (5.4) obtenemos que

$$\int_0^t \frac{\operatorname{sen} x}{x} dx = \frac{\pi}{2} - \int_0^\infty \frac{\cos t + y \operatorname{sen} t}{1 + y^2} e^{-ty} dy.$$

Se sigue del teorema de convergencia dominada que

$$\lim_{t\to\infty} \int_0^t \frac{\operatorname{sen} x}{x} dx = \frac{\pi}{2}. \qquad \square$$

Teorema 5.28. *Sea $\mu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ una medida de probabilidad con función característica $\hat{\mu}$. Para $a, b \in \mathbb{R}$, con $a < b$, arbitrarios*

$$\mu((a, b]) + \frac{1}{2}(\mu(\{a\}) - \mu(\{b\})) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{(-T,T)} \frac{e^{-ita} - e^{-itb}}{it} \hat{\mu}(t) d\lambda(t).$$

Demostración. Notemos que por el lema 5.26 aplicado con $n = 0$,

$$\left|\frac{e^{-ita} - e^{-itb}}{t}\right| = \left|\frac{e^{it(b-a)} - 1}{t}\right| \le b - a,$$

de donde se sigue que

$$\left|\int_{(-T,T)} \frac{e^{-ita}-e^{-itb}}{it}\hat{\mu}(t)d\lambda(t)\right| \leq \int_{(-T,T)} \left|\frac{e^{-ita}-e^{-itb}}{t}\right| d\lambda(t) \leq 2(b-a)T.$$

Así podremos aplicar Fubini en el siguiente cálculo:

$$\begin{aligned}
I_T &= \frac{1}{2\pi}\int_{(-T,T)} \frac{e^{-ita}-e^{-itb}}{it}\hat{\mu}(t)d\lambda(t) \\
&= \frac{1}{2\pi}\int_{(-T,T)}\int \frac{e^{-ita}-e^{-itb}}{it}e^{itx}d\mu(x)d\lambda(t) \\
&= \frac{1}{2\pi}\int\int_{(-T,T)} \frac{e^{-it(a-x)}-e^{-it(b-x)}}{it}d\lambda(t)d\mu(x) \\
&= \frac{1}{\pi}\int\int_0^T \left(\frac{\operatorname{sen}(t(b-x))}{t} - \frac{\operatorname{sen}(t(a-x))}{t}\right) dt d\mu(x) \\
&= \frac{1}{\pi}\int_{(-\infty,a)\cup(b,\infty)} H(a,b,x,T)d\mu(x) + \frac{1}{\pi}\mu(\{a\})\int_0^T \frac{\operatorname{sen}(t(b-a))}{t}dt \\
&\quad + \frac{1}{\pi}\int_{(a,b)} H(a,b,x,T)d\mu(x) - \frac{1}{\pi}\mu(\{b\})\int_0^T \frac{\operatorname{sen}(t(a-b))}{t}dt.
\end{aligned}$$

donde

$$\begin{aligned}
H(a,b,x,T) &= \int_0^T \left(\frac{\operatorname{sen}(t(b-x))}{t} - \frac{\operatorname{sen}(t(a-x))}{t}\right) dt \\
&= \int_{(0,T)} \left(\frac{\operatorname{sen}(t(b-x))}{t} - \frac{\operatorname{sen}(t(a-x))}{t}\right) d\lambda(t).
\end{aligned}$$

La última igualdad se da por el teorema 2.36.

Finalmente, por el lema 5.27 y por convergencia dominada, se tiene que

$$\lim_{T\to\infty} I_T = \frac{1}{2}\mu(\{a\}) + \mu((a,b)) + \frac{1}{2}\mu(\{b\}) = \mu((a,b]) + \frac{1}{2}(\mu(\{a\}) - \mu(\{b\})). \qquad \square$$

Observación 5.29. Si $a, b \in \mathbb{R}$, con $a < b$ son tales que $\mu(\{a\}) = \mu(\{b\}) = 0$ entonces

$$\mu((a,b]) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{(-T,T)} \frac{e^{-ita}-e^{-itb}}{it}\hat{\mu}(t)d\lambda(t).$$

Observación 5.30. Como μ es una medida finita, $\mathcal{D}_\mu = \{a \in \mathbb{R} : \mu(\{a\}) > 0\}$ es a lo más numerable, entonces $\mathcal{C} = \{(a,b] : a, b \notin \mathcal{D}_\mu\}$ es un π-sistema tal que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$, por lo cual podremos recuperar la medida μ bajo la observación anterior. Esto nos da otra prueba del teorema de unicidad para el caso $d = 1$, pues si $\hat{\mu} = \hat{\nu}$ entonces $\mu((a,b]) = \nu((a,b])$ para todo $(a,b] \in \mathcal{C}$ y por el lema de clases monótonas se sigue que $\mu \equiv \nu$.

Teorema 5.31. *Sea $\mu : \mathcal{B}(\mathbb{R}) \to [0,\infty]$ una medida de probabilidad con función característica $\hat{\mu}$. Si $\hat{\mu} \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ entonces $\mu \ll \lambda$ y tiene derivada de Radon–Nikodym*

$$f(x) = \frac{1}{2\pi}\int e^{-itx}\hat{\mu}(t)d\lambda(t).$$

Demostración. Primero observemos que para $a < b$,

$$\left|\int \frac{e^{-ita} - e^{-itb}}{it} \hat{\mu}(t) d\lambda(t)\right| \leq \int \left|\frac{e^{-ita} - e^{-itb}}{it} \hat{\mu}(t)\right| d\lambda(t) \leq (b-a) \int |\hat{\mu}|\, d\lambda = I\lambda((a,b)) < \infty,$$

donde $I = \int |\hat{\mu}|\, d\lambda$. Así, por el teorema anterior tendremos que

$$\mu((a,b)) + \frac{1}{2}(\mu(\{a\}) + \mu(\{b\})) \leq \frac{I}{2\pi}\lambda((a,b)) < \infty, \tag{5.5}$$

lo cual implica que $\mu(\{a\}) = 0$ para toda $a \in \mathbb{R}$. Por el lema 1.25 sabemos que para $A \in \mathcal{B}(\mathbb{R})$

$$\lambda(A) = \text{ínf}\left\{\sum_{n\in\mathbb{N}} \lambda((a_n, b_n)) : a_n, b_n \in \mathbb{R}; A \subset \bigcup_{n\in\mathbb{N}} (a_n, b_n)\right\}.$$

En particular, si $\lambda(A) = 0$ y $\epsilon > 0$ es arbitrario entonces existe una sucesión de intervalos $\{(a_n, b_n)\}_{n\geq 0}$ tal que $A \subset \bigcup_{n\in\mathbb{N}} (a_n, b_n)$ y

$$\sum_{n\in\mathbb{N}} \lambda((a_n, b_n)) < \frac{2\pi}{I}\epsilon.$$

Por lo tanto, por monotonía y la cota (5.5),

$$\mu(A) \leq \sum_{n\in\mathbb{N}} \mu((a_n, b_n)) \leq \sum_{n\in\mathbb{N}} \frac{I}{2\pi}\lambda((a_n, b_n)) < \epsilon.$$

Por lo tanto $\mu(A) = 0$ y así $\mu \ll \lambda$.

Sea f la derivada de Radon–Nikodym de μ con respecto a λ. Notemos que para cada $n \in \mathbb{N}$ y $x \in \mathbb{R}$,

$$n \int_{(x,x+1/n]} f d\lambda = n\mu((x, x+1/n]) = \frac{1}{2\pi} \int \frac{e^{-itx} - e^{-it(x+1/n)}}{it/n} \hat{\mu}(t) d\lambda(t).$$

Considerando el límite cuando $n \to \infty$, usando convergencia dominada,

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \hat{\mu}(t) d\lambda(t). \qquad \square$$

Para dar un ejemplo de cómo aplicar el teorema anterior consideremos la medida de probabilidad $\mu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ dada por

$$\mu(A) = \int_A (1 - |x|) \mathbb{1}_{[-1,1]}(x) d\lambda(x).$$

Su función característica está dada por

$$\hat{\mu}(t) = \int_{[-1,1]} e^{itx}(1 - |x|) d\lambda(x) = 2\int_0^1 \cos(tx)(1-x)dx = 2\frac{1-\cos t}{t^2}.$$

Por el lema 5.27 se sigue que

$$\int \frac{1 - \cos t}{t^2} d\lambda(t) = \pi$$

y por el teorema anterior

$$(1-|x|)\mathbb{1}_{[-1,1]}(x) = \frac{1}{2\pi}\int 2\frac{1-\cos t}{t^2}e^{-itx}d\lambda(t) = \int \frac{1-\cos t}{\pi t^2}e^{-itx}d\lambda(t).$$

Usando el hecho de que $f(x) = \frac{1-\cos x}{\pi x^2}$ es no negativa y $\int f d\lambda = 1$ y la paridad de f, se sigue que $\hat{\nu}(t) = (1-|t|)\mathbb{1}_{[-1,1]}(t)$, donde $\nu : \mathcal{B}(\mathbb{R}) \to [0,\infty]$ está dada por

$$\nu(A) = \int_A \frac{1-\cos x}{\pi x^2}d\lambda(x).$$

5.4. Vectores gaussianos

Antes de definir una de las familias de variables aleatorias más importantes en la teoría de probabilidad, se darán algunas definiciones de conceptos importantes como lo son la *covarianza* y *varianza*, así como extender el operador de esperanza a variables aleatorias con valores en $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Definición 5.32. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $X, Y \in L^2$ dos variables aleatorias. Se define la **covarianza** de X con Y como

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])].$$

Asimismo se define la **varianza** de $X \in L^2$ por

$$\operatorname{Var}(X) := \operatorname{Cov}(X,X).$$

Definición 5.33. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $X : \Omega \to \mathbb{R}^d$ una variable aleatoria $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$-medible tal que para toda $j \in \{1, \ldots, n\}$, $X_j \in L^1$. Se define la esperanza de X como

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]),$$

es decir el vector de esperanzas de sus entradas. De forma análoga se puede definir la esperanza de una matriz X de dimensión $d_1 \times d_2$ con $X_{jk} \in L^1$ con $j \in \{1, \ldots, d_1\}$ y $k \in \{1, \ldots, d_2\}$.

Si además $X_j \in L^2$ para $j \in \{1, \ldots, d\}$ entonces se define $\operatorname{Cov}(X)$, la **matriz de covarianzas**, como la matriz Q donde $Q_{jk} = \operatorname{Cov}(X_j, X_k)$, con $j, k \in \{1, \ldots, n\}$.

Definición 5.34. Diremos que $X : \Omega \to \mathbb{R}$ es una variable aleatoria **gaussiana** con parámetros (m, σ^2) si su ley tiene λ-densidad $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}$, cuando $\sigma^2 > 0$. Si $\sigma^2 = 0$ entonces la ley de X es δ_m y diremos que es *degenerada*. Cuando $(m, \sigma^2) = (0, 1)$ hablaremos de una variable aleatoria gaussiana *estándar*.

Observación 5.35. Si X es una variable aleatoria gaussiana con parámetros (m, σ^2) entonces su función característica queda dada por

$$\varphi_X(t) = e^{imt - \frac{1}{2}\sigma^2 t^2}.$$

La importancia de esta colección de distribuciones es que, bajo ciertas condiciones, la suma de variables aleatorias independientes será aproximadamente gaussiano, resultado que se puede extender a vectores aleatorios, por lo que nos gustaría definir lo que es un *vector gaussiano* en $\mathbb{R}^d$. Si bien uno quisiera extrapolar la definición dada para $\mathbb{R}$, esta extensión no resulta directa.

Definición 5.36. Un vector aleatorio X, $\mathbb{R}^d$-valuado, es **gaussiano** o **normal multivariado** si para cualquier $a \in \mathbb{R}^d$, $\langle a, X\rangle$ es una variable aleatoria gaussiana (posiblemente degenerada).

Los vectores gaussianos quedarán caracterizados por su función característica.

Teorema 5.37. *X es un vector gaussiano $\mathbb{R}^d$-valuado si y sólo si su función característica está dada por*

$$\varphi_X(t) = \exp\left(i\langle t, m\rangle - \frac{1}{2}\langle t, Qt\rangle\right),$$

con $m \in \mathbb{R}^d$ y Q una matriz simétrica positiva semidefinida. Entonces Q es la matriz de covarianzas de X y μ es el vector de medias de X.

Demostración. Primero supongamos que X es un vector gaussiano con media m y matriz de covarianzas Q. Para $a \in \mathbb{R}^d$ arbitrario, definamos $Y = \langle a, X\rangle$ que es gaussiana por definición, con media $\mathbb{E}[Y] = \langle a, m\rangle$ y varianza $\sigma^2(Y) = \langle a, Qa\rangle$. Entonces Y tiene función característica

$$\varphi_Y(t) = \exp\left(it\langle a, m\rangle - \frac{1}{2}t^2\langle a, Qa\rangle\right).$$

Finalmente notemos que $\varphi_Y(1) = \varphi_X(a)$.

Recíprocamente, si $Y = \langle a, X\rangle$ con $a \in \mathbb{R}^d$ y $t \in \mathbb{R}$ entonces

$$\varphi_Y(t) = \varphi_X(ta) = \exp\left(it\langle a, m\rangle - \frac{1}{2}t^2\langle a, Qa\rangle\right),$$

que es la función característica de una variable aleatoria gaussiana con parámetros $(\langle a, m\rangle, \langle a, Qa\rangle)$ y por el teorema de unicidad se sigue que Y es gaussiana. □

Si consideramos dos variables aleatorias $X, Y \in L^2$, diremos que *no están correlacionadas* si $\mathrm{Cov}(X, Y) = 0$. En general es fácil ver que si X es independiente de Y entonces $\mathrm{Cov}(X, Y) = 0$, sin embargo la afirmación recíproca es falsa en general. Sin embargo, para vectores gaussianos, gracias al teorema anterior, tendremos el siguiente resultado.

Corolario 5.38. *Si $X : \Omega \to \mathbb{R}^d$ es un vector gaussiano entonces $X_1, \ldots, X_d$ son independientes si y sólo si $\mathrm{Cov}(X)$ es una matriz diagonal.*

Antes de probar el teorema, notemos que si $d = 2$, $\mathrm{Cov}(X_1, X_2) = 0$ si y sólo si X_1 es independiente de X_2, obteniendo un recíproco parcial a la afirmación hecha antes de enunciar el corolario.

Demostración. Por el corolario 5.23 sabemos que $X_1, \ldots, X_d$ son independientes si y sólo si $\varphi_X(t) = \prod_{j=1}^d \varphi_{X_j}(t_j)$ para todo $t \in \mathbb{R}^d$. Por el teorema anterior tenemos que

$$\varphi_X(t) = \exp\left\{i\langle t, m\rangle - \frac{1}{2}\langle t, Qt\rangle\right\},$$

donde $m = \mathbb{E}[X]$ y $Q = \mathrm{Cov}(X)$. Por definición, si $e_j \in \mathbb{R}^d$ es el j-ésimo vector canónico, $X_j = \langle e_j, X\rangle$ es gaussiano con parámetros $(m_j, \sigma_j^2) = (\langle e_j, m_j\rangle, \langle e_j, Qe_j\rangle)$ y, nuevamente por el teorema anterior,

$$\varphi_{X_j}(t_j) = \exp\left\{im_jt_j - \frac{1}{2}\sigma_j^2t_j^2\right\}.$$

Por lo tanto, $X_1, \dots, X_d$ son independientes si y sólo si

$$\langle t, Qt\rangle = \sum_{j=1}^{d} \sigma_j^2 t_j^2$$

para todo $t \in \mathbb{R}^d$, de donde se sigue el resultado. $\square$

Teorema 5.39. *Si $X : \Omega \to \mathbb{R}^d$ es un vector gaussiano entonces existen $Y_1, \dots, Y_n : \Omega \to \mathbb{R}$, una colección de variables gaussianas donde Y_j tiene parámetros $(0, \beta_j)$ (con $\beta_j \geq 0$), y A una matriz ortogonal tal que $X = AY + m$, donde $Y = (Y_1, \dots, Y_d)$ y $m = \mathbb{E}[X]$.*

Demostración. Sea $Q = \operatorname{Cov}(X)$ la cual es una matriz simétrica positiva semidefinida de entradas reales. Sabemos que existen A y B, una matriz ortogonal, *i.e.* $A^*A = AA^* = \mathrm{Id}$, y una matriz diagonal tal que $\beta_j := B_{jj} \geq 0$ para toda $j \in \{1, \dots, d\}$ de forma respectiva tales que $Q = ABA^*$. Entonces $Y = A^*(X - m)$ es un vector gaussiano tal que

$$\begin{aligned}
\varphi_Y(t) &= e^{-i\langle t, A^*m\rangle}\varphi_X(At) \\
&= e^{-i\langle At, m\rangle} \exp\left\{ i\langle At, m\rangle - \frac{1}{2}\langle At, QAt\rangle \right\} \\
&= \exp\left\{ -\frac{1}{2}\langle t, A^*ABA^*At\rangle \right\} = \exp\left\{ -\frac{1}{2}\langle t, Bt\rangle \right\},
\end{aligned}$$

de donde se sigue el resultado. $\square$

Observemos que en el teorema que se acaba de probar, puede que $\beta_j = 0$ para alguna $j \in \{1, \dots, d\}$, lo que implicaría que $Y_j = 0$ casi seguramente. Por lo tanto, el número de variables aleatorias independientes requeridas puede ser menor que el número de componentes no degeneradas de X. Esto en particular se expresa en el siguiente resultado.

Corolario 5.40. *Sea $X : \Omega \to \mathbb{R}^d$ un vector gaussiano. La ley de X es absolutamente continua con respecto a λ^d si y sólo si $\operatorname{Cov}(X)$ no es singular.*

Demostración. Del teorema anterior $A^*BA = Q = \operatorname{Cov}(X)$ y $X = AY + m$ con Y un vector gaussiano con matriz de covarianzas diagonal. Primero supongamos que $Q = \operatorname{Cov}(X)$ no es singular, lo cual implica que $\beta_j = B_{jj} > 0$ para toda $j \in \{1, \dots, d\}$. Es elemental ver que la ley de Y_j es absolutamente continua con respecto a λ y su derivada de Radon–Nikodym está dada por

$$f_{Y_j}(y) = \frac{1}{\sqrt{2\pi\beta_j}} e^{-\frac{y_j^2}{2\beta_j}}.$$

Luego, la ley de Y es absolutamente continua con respecto a λ^d con derivada de Radon–Nikodym dada por

$$f_Y(y) = \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\beta_j}} e^{-\frac{y_j^2}{2\beta_j}} = (2\pi)^{-n/2} \left|\det B\right|^{-1/2} e^{-\frac{1}{2}\langle y, B^{-1}y\rangle}.$$

Notando que $Y = A^*(X - m)$, por el teorema 2.46, la ley de X será absolutamente continua con respecto a λ^d y su derivada de Radon–Nikodym será

$$\begin{aligned}
f_X(x) &= (2\pi)^{-n/2} \left|\det B\right|^{-1/2} \left|\det A^*\right| e^{-\frac{1}{2}\langle A^*(x-m), BA^*(x-m)\rangle} \\
&= (2\pi)^{-n/2} \left|\det Q\right|^{-1/2} e^{-\frac{1}{2}\langle x-m, Q(x-m)\rangle},
\end{aligned}$$

donde usamos que $|\det Q| = |\det B|$ y $|\det A| = |\det A^*| = 1$.

De forma recíproca, supongamos que Q es singular, ***i.e.*** $\det Q = 0$. Supongamos que el kernel de Q es de dimensión k y sea $\{a_1, \ldots, a_k\} \subset \ker Q$ una base del kernel de Q. Entonces, para cada $j \in \{1, \ldots, k\}$, la variable aleatoria $Z_j = \langle a_j, X\rangle$ será gaussiana con parámetros $(\langle a_j, m\rangle, 0)$, donde $m = \mathbb{E}[X]$, por lo que $Z_j = \langle a_j, m\rangle$ casi seguramente. Así, si $\mathcal{H} = (\ker Q)^{\perp} + m$, tendremos que $\mathbb{P}(X \in \mathcal{H}) = 1$. Notando que $\mathcal{H}$ es un hiperplano de dimensión $d-k$, $\lambda^d(\mathcal{H}) = 0$. Ahora supongamos que la ley de X es absolutamente continua con respecto a λ^d con derivada de Radon–Nikodym f; entonces tendríamos

$$1 = \mathbb{P}(X \in \mathcal{H}) = \int_{\mathcal{H}} f d\lambda^d = 0,$$

lo cual es una contradicción. Por lo tanto la ley de X no es absolutamente continua con respecto a λ^d. □

Concluimos la sección y el capítulo con un resultado que se deja como ejercicio al lector.

Proposición 5.41. *Si* $X : \Omega \to \mathbb{R}^{d_1}$ *y* $Y : \Omega \to \mathbb{R}^{d_2}$ *son vectores gaussianos independientes entonces* $Z = (X, Y) : \Omega \to \mathbb{R}^{d_1+d_2}$ *es un vector gaussiano.*

Capítulo 6

Convergencias

A lo largo de estas notas se han introducido las convergencias *casi donde sea* y *en* L^p para funciones con valores extendidos o con valores complejos. En este capítulo se introducen y estudian diversos modos de convergencia para variables aleatorias con valores en espacios más generales y se dan algunas relaciones entre estos modos.

6.1. Convergencia de variables aleatorias

Antes de definir las convergencias pertinentes para variables aleatorias, recordemos que si (S,d) es un espacio métrico separable entonces $\mathcal{B}(S\times S)=\mathcal{B}(S)\otimes\mathcal{B}(S)$. Como la métrica d es continua con respecto a la topología producto en $S\times S$, entonces será medible y, si X y Y son variables aleatorias con valores en S, así $d(X,Y)$ será una variable aleatoria, por lo que la siguiente definición tendrá sentido.

Definición 6.1. Sean $(\Omega,\mathcal{F},\mathbb{P})$ un espacio de probabilidad, (S,d) un espacio métrico separable, $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias $(\mathcal{F},\mathcal{B}(S))$-medibles y $X:\Omega\to S$ otra variable aleatoria.

1) $\{X_n\}_{n\in\mathbb{N}}$ **converge casi seguramente** a X, denotado por $X_n\xrightarrow{c.s.}X$, si existe $N\in\mathcal{F}$ con $\mathbb{P}(N)=0$ tal que $X_n(\omega)\mapsto X(\omega)$ para todo $\omega\in\Omega-N$.

2) $\{X_n\}_{n\in\mathbb{N}}$ **converge en probabilidad** a X, denotado por $X_n\xrightarrow{\mathbb{P}}X$, si para todo $\epsilon>0$,
$$\lim_{n\to\infty}\mathbb{P}(d(X_n,X)\geq\epsilon)=0.$$

3) $\{X_n\}_{n\in\mathbb{N}}$ **converge casi uniformemente** a X, denotado por $X_n\xrightarrow{c.u.}X$, si para todo $\delta>0$ existe $A\in\mathcal{F}$ con $\mathbb{P}(A)<\delta$, tal que $\{X_n\}_{n\in\mathbb{N}}$ converge uniformemente a X en $\Omega-A$.

Cuando nuestro espacio de medida es finito, como lo son los espacios de probabilidad, entonces tendremos el siguiente resultado, que nos permite caracterizar la convergencia casi segura en términos de la medida. El resultado se enuncia sin demostración, invitando al lector a realizarla.

Proposición 6.2. *Sean $(\Omega,\mathcal{F},\mathbb{P})$ un espacio de probabilidad, (S,d) un espacio métrico separable y $\{X_n\}_{n\in\mathbb{N}}$ y X variables aleatorias $(\mathcal{F},\mathcal{B}(S))$-medibles. Entonces $X_n\xrightarrow{c.s.}X$ si y sólo si para todo $\epsilon>0$,*
$$\lim_{n\to\infty}\mathbb{P}\left(\bigcup_{m\geq n}\{\omega\in\Omega: d(X_m(\omega),X(\omega))\geq\epsilon\}\right)=0.$$

Así como se necesitó que (S,d) fuese un espacio métrico separable para que $d(X,Y)$ fuese una variable aleatoria, si ahora $(S,\|\cdot\|)$ es un espacio de Banach, para que $\alpha X+\beta Y$ sea variable aleatoria, también pediremos que $(S,\|\cdot\|)$ sea separable, considerando la topología generada por la métrica inducida por $\|\cdot\|$: $d_{\|\cdot\|}(x,y)=\|x-y\|$. Notemos que esto en particular implica que $\mathcal{B}(S)=\mathcal{B}(S,d_{\|\cdot\|})$.

Definición 6.3. Sean $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, $(S, \|\cdot\|)$ un espacio de Banach separable, $p \in [1, \infty)$, $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles y X otra variable aleatoria $(\mathcal{F}, \mathcal{B}(S))$-medible. Si $\|X\| \in L^p$ y $\{\|X_n\|\}_{n\in\mathbb{N}} \subset L^p$, diremos que X_n **converge en media** $\boldsymbol{p}$ a X, denotado por $X_n \xrightarrow{L^p} X$, si

$$\lim_{n\to\infty} \mathbb{E}[\|X_n - X\|^p] = 0.$$

De ahora en adelante se entenderá que que las variables aleatorias que mencionaremos tienen valores en un espacio, si no se menciona explícitamente, de Banach separable cuando se mencione la convergencia en media p y de un espacio métrico separable en otro caso

Notemos que si $X_n \xrightarrow{*} X$ bajo cualquier modo, entonces para toda subsucesión $\{n_k\} \subset \mathbb{N}$, $X_{n_k} \xrightarrow{*} X$. Además, para todos los modos de convergencia se pueden definir las correspondientes nociones de sucesiones de Cauchy y se invita al lector a enunciar las definiciones correspondientes.

Salvo la convergencia es probabilidad, es claro que si una sucesión es convergente en un modo, entonces será de Cauchy en ese mismo modo. Para ver que esto también es cierto para la convergencia en probabilidad, notemos que

$$\begin{aligned}&\{\omega \in \Omega : d(X_n(\omega), X_m(\omega)) < \epsilon\}\\ &\qquad \supset \{\omega \in \Omega : d(X_n(\omega), X(\omega)) < \epsilon/2\} \cap \{\omega \in \Omega : d(X_m(\omega), X(\omega)) < \epsilon/2\},\end{aligned}$$

de donde se sigue, considerando complementos, que

$$\begin{aligned}&\{\omega \in \Omega : d(X_n(\omega), X_m(\omega)) \geq \epsilon\}\\ &\qquad \subset \{\omega \in \Omega : d(X_n(\omega), X(\omega)) \geq \epsilon/2\} \cup \{\omega \in \Omega : d(X_m(\omega), X(\omega)) \geq \epsilon/2\}.\end{aligned}$$

Un primer resultado que relaciona estos modos de convergencia es el siguiente.

Teorema 6.4. *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, $p \in [1, \infty)$ y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias que converge en media p a X. Entonces $X_n \xrightarrow{\mathbb{P}} X$.*

Demostración. Para $\epsilon > 0$ y $n \in \mathbb{N}$ arbitrarios sea $E_n(\epsilon) = \{\omega \in \Omega : d_{\|\cdot\|}(X_n(\omega), X(\omega)) \geq \epsilon\} = \{\omega \in \Omega : \|X_n(\omega) - X(\omega)\| \geq \epsilon\}$, entonces $\epsilon^p \mathbb{1}_{E_n(\epsilon)} \leq \|X_n - X\|^p\, \mathbb{1}_{E_n(\epsilon)}$, por lo que

$$\epsilon^p \mathbb{P}(E_n(\epsilon)) = \mathbb{E}[\epsilon^p \mathbb{1}_{E_n(\epsilon)}] \leq \mathbb{E}[\|X_n - X\|^p\, \mathbb{1}_{E_n(\epsilon)}] \leq \mathbb{E}[\|X_n - X\|^p],$$

de donde $\lim_{n\to\infty} \mathbb{P}(E_n(\epsilon)) = 0$. □

Para ver que el recíproco del resultado es falso, consideremos $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$ y $X_n = n\mathbb{1}_{[0,1/n]}$ y observemos que para cada $\epsilon > 0$ tendremos que

$$E_n(\epsilon) = \{\omega \in \Omega : |X_n(\omega)| \geq \epsilon\} = \begin{cases} \emptyset, & \text{si } \epsilon > n, \\ [0, 1/n], & \text{si } \epsilon \leq n, \end{cases}$$

por lo que $\mathbb{P}(E_n(\epsilon)) \leq 1/n$, y así $X_n \xrightarrow{\mathbb{P}} 0$. Sin embargo,

$$\mathbb{E}\big[\, |X_n - X_{2n}|^p \,\big] = n^p \mathbb{E}[\mathbb{1}_{[0,1/n]}] = n^{p-1},$$

por lo que $\{X_n\}_{n\in\mathbb{N}}$ no es Cauchy en media p y por lo tanto $\{X_n\}_{n\in\mathbb{P}}$ no converge en media p. En este mismo ejemplo, observemos que $X_n(\omega) \to 0$ para toda $\omega \in (0, 1]$, por lo que convergencia casi segura tampoco implicará convergencia dominada; sin embargo, tendremos el siguiente resultado cuya demostración se omite.

Teorema 6.5. *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad,* $p \in [1, \infty)$ *fijo y* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias que convergen casi seguramente a* X *tales que* $\{\|X_n\|\}_{n\in\mathbb{N}} \subset L^p$*. Si existe* $g \in L^1$ *tal que* $\|X_n\|^p \leq g$ *casi seguramente para toda* $n \in \mathbb{N}$*, entonces* $\|X\| \in L^p$ *y* $X_n \xrightarrow{L^p} X$.

De la definición es claro que, en un espacio de medida finita, convergencia casi segura implica convergencia en probabilidad; el resultado inverso, en general, es falso: si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ es una sucesión de eventos independientes tales que $\mathbb{P}(A_n) = 1/n$, $\mathbb{P}(A_n) \to 0$ entonces $\mathbb{1}_{A_n} \to 0$ en probabilidad, sin embargo $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ y por el segundo lema de Borel–Cantelli,

$$1 = \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{E}\left[\limsup_{n\to\infty} \mathbb{1}_{A_n}\right],$$

de donde $\mathbb{1}_{A_n}$ no converge casi seguramente a 0. Veremos que existe un resultado parcialmente inverso.

Teorema 6.6. *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad,* $\{X_n\}_{n\in\mathbb{N}}$ *y* X *variables aleatorias con valores en* (S, d)*. Entonces* $X_n \xrightarrow{\mathbb{P}} X$ *si y sólo si para cada subsucesión* $\{X_{n_k}\}_{k\in\mathbb{N}}$ *de* $\{X_n\}_{n\in\mathbb{N}}$ *existe una subsubsucesión* $\{X_{n_{k_r}}\}_{r\in\mathbb{N}}$ *tal que* $X_{n_{k_r}} \xrightarrow{c.s.} X$.

Demostración. Si $X_n \xrightarrow{\mathbb{P}} X$ entonces también toda subsucesión lo hará. Dada $\{X_{n_k}\}_{k\in\mathbb{N}} \subset \{X_n\}_{n\in\mathbb{N}}$, por definición de convergencia en probabilidad podemos construir $\{X_{n_{k_r}}\}_{r\in\mathbb{N}}$ tal que para toda $r \in \mathbb{N}$,

$$\mathbb{P}(d(X_{n_{k_r}}, X) \geq r^{-1}) \leq r^{-2}.$$

Para $\epsilon > 0$ arbitraria, por propiedad arquimedeana existe $N \in \mathbb{N}$ tal que $\{d(X_{n_{k_r}}, X) \geq \epsilon\} \subset \{d(X_{n_{k_r}}, X) \geq r^{-1}\}$ para toda $r \geq N$; por el primer lema de Borel–Cantelli,

$$\begin{aligned} 0 &= \mathbb{P}\left(\limsup_{r\to\infty} \left\{d(X_{n_{k_r}}, X) \geq r^{-1}\right\}\right) \\ &= \lim_{r\to\infty} \mathbb{P}\left(\bigcup_{m\geq r} \{d(X_{n_{k_m}}, X) \geq m^{-1}\}\right) \\ &\geq \lim_{r\to\infty} \mathbb{P}\left(\bigcup_{m\geq r} \{d(X_{n_{k_m}}, X) \geq \epsilon\}\right). \end{aligned}$$

Luego $X_{n_{k_r}} \xrightarrow{c.s.} X$.

Inversamente, supongamos que $\{X_n\}_{n\in\mathbb{N}}$ no converge en probabilidad a X; es decir que existen $\epsilon, \delta > 0$ tales que para cada $k \in \mathbb{N}$ existe $n_k \geq k$ tal que $\mathbb{P}(d(X_{n_k}, X) \geq \epsilon) \geq \delta$. Sea $\{X_{n_k}\}_{k\in\mathbb{N}}$ la subsucesión de $\{X_n\}_{n\in\mathbb{N}}$ que cumpla lo anterior, entonces ninguna subsubsucesión convergerá a X casi seguramente. $\square$

Para reducir la notación, si $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de probabilidad y (S, τ) es un espacio topológico, denotaremos por $\mathcal{L}^0(\Omega, S)$ al conjunto de funciones $(\mathcal{F}, \mathcal{B}(S))$-medibles y por $L^0(\Omega, S)$ a las clases de equivalencia de $\mathcal{L}^0(\Omega, S)$ inducidas por la relación de igualdad $\mathbb{P}$-c.s. Entonces podremos metrizar la convergencia en probabilidad de la siguiente manera.

Teorema 6.7. *Sea* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad y* (S, d) *un espacio métrico. Entonces* $\rho(X, Y) = \mathbb{E}\left[\frac{d(X,Y)}{1+d(X,Y)}\right]$ *es una métrica en* $L^0(\Omega, S)$ *que metriza la convergencia en probabilidad.*

Demostración. Que $\rho(X,Y)=0$ si y sólo si $X=Y$ c.s. y $\rho(X,Y)=\rho(Y,X)$ es claro. Para probar la desigualdad del triángulo, notemos que $g(x)=\frac{x}{1+x}$ es creciente cuando $x\geq 0$, por lo que

$$\frac{d(X,Y)}{1+d(X,Y)}\leq\frac{d(X,Z)+d(Z,Y)}{1+d(X,Z)+d(Z,Y)}\leq\frac{d(X,Z)}{1+d(X,Z)}+\frac{d(Z,Y)}{1+d(Z,Y)}.$$

Tomando esperanzas se sigue el resultado.

Ahora supongamos que $\{X_n\}_{n\in\mathbb{N}}$ y X son variables aleatorias $(\mathcal{F},\mathcal{B}(S))$-medibles tales que $\rho(X_n,X)\to 0$. Sea $\epsilon>0$ arbitrario; por hipótesis existe $N\in\mathbb{N}$ tal que para toda $n\geq N$, $\rho(X_n,X)<\frac{\epsilon^2}{1+\epsilon}$, de donde

$$\begin{aligned}\frac{\epsilon}{1+\epsilon}\mathbb{1}_{\{d(X_n,X)\geq\epsilon\}}&\leq\frac{d(X_n,X)}{1+d(X_n,X)}\mathbb{1}_{\{d(X_n,X)\geq\epsilon\}}\\&\leq\frac{d(X_n,X)}{1+d(X_n,X)}\end{aligned}$$

para cada $n\geq N$. Entonces, tomando esperanzas, tendremos que

$$\mathbb{P}(d(X_n,X)\geq\epsilon)\leq\frac{\epsilon}{1+\epsilon}\rho(X_n,X)\leq\epsilon.$$

Por lo tanto $X_n\xrightarrow{\mathbb{P}}X$.

Inversamente notemos que para $\epsilon>0$ dado, existe $N\in\mathbb{N}$ tal que

$$\mathbb{P}\left(d(X_n,X)\geq\frac{\epsilon}{2}\right)\leq\frac{\epsilon}{2}$$

para toda $n\geq N$. De esta manera tendremos que

$$\frac{d(X_n,X)}{1+d(X_n,X)}<\frac{\frac{\epsilon}{2}}{1+\frac{\epsilon}{2}}\mathbb{1}_{\{d(X_n,X)<\epsilon/2\}}+\frac{d(X_n,X)}{1+d(X,X)}\mathbb{1}_{\{d(X_n,X)\geq\epsilon/2\}}\leq\frac{\epsilon}{2}+\mathbb{1}_{\{d(X_n,X)\geq\epsilon/2\}}.$$

Por lo tanto, para cada $n\geq N$,

$$\rho(X_n,X)<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$ □

A la métrica introducida en el teorema anterior se le conoce como *métrica de Ky Fan*. Otra métrica de Ky Fan está dada por

$$\alpha(X,Y)=\text{ínf}\{\epsilon>0:\mathbb{P}(d(X,Y)\geq\epsilon)\leq\epsilon\}.$$

Queda como ejercicio al lector probar que esta métrica está bien definida y que sí es una métrica. Ahora veremos que haber metrizado la topología en probabilidad nos permitirá probar con mayor facilidad algunos resultados.

Teorema 6.8. *Sean* $(\Omega,\mathcal{F},\mathbb{P})$ *un espacio de probabilidad y* (S,d) *un espacio métrico completo.* $L^0(\Omega,S)$ *es completo bajo* ρ.

Demostración. Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de Cauchy con respecto a ρ. Para cada $k\in\mathbb{N}$ sea $n_k\in\mathbb{N}$ tal que para todo $m\geq n_k$,

$$\rho(X_m,X_{n_k})\leq\frac{2^{-2k}}{1+2^{-k}},$$

de donde se sigue que para $m \geq k$

$$\mathbb{P}(d(X_{n_k}, X_{n_m}) \geq 2^{-k}) \leq 2^{-k}.$$

Por el primer lema de Borel–Cantelli se sigue que

$$1 = \mathbb{P}\left(\liminf_{k\to\infty}\left\{d(X_{n_k}, X_{n_{k+1}}) < 2^{-k}\right\}\right).$$

Así, si $A = \liminf_{k\to\infty}\left\{d(X_{n_k}, X_{n_{k+1}}) < 2^{-k}\right\}$ notemos que para $\omega \in A$ existe $k \in \mathbb{N}$ tal que $d(X_{n_r}(\omega), X_{n_{r+1}}(\omega)) < 2^{-r}$ para toda $r \geq k$. Por lo tanto, para toda $r \geq k$, $d(X_{n_k}(\omega), X_{n_r}(\omega)) < 2^{-k+1}$, es decir que $\{X_{n_k}(\omega)\}_{k\in\mathbb{N}}$ es una sucesión de Cauchy y, como (S, d) es completo, entonces $\lim_{k\to\infty} X_{n_k}(\omega)$ existe. Definamos

$$Y(\omega) = \begin{cases} \lim_{k\to\infty} X_{n_k}(\omega), & \text{si } \omega \in A, \\ y, & \text{si } \omega \notin A. \end{cases}$$

con $y \in S$ arbitraria y fija. Entonces $\rho(X_{n_k}, Y) \to 0$ y en un espacio métrico cualquier sucesión de Cauchy con una subsucesión convergente converge al mismo límite, por lo que $\rho(X_n, Y) \to 0$. □

Observación 6.9. Sean $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $\{X_n\}_{n\in\mathbb{N}}$, $\{Y_n\}_{n\in\mathbb{N}}$, X y Y variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles.

1) Si $X_n \xrightarrow{\mathbb{P}} X$ y $X_n \xrightarrow{\mathbb{P}} Y$ entonces $X = Y$ casi seguramente.

2) Si $X_n \xrightarrow{\mathbb{P}} X$ y $X_n = Y_n$ casi seguramente para toda $n \in \mathbb{N}$ entonces $Y_n \xrightarrow{\mathbb{P}} X$.

Demostración. 1) Notemos que $\{X \neq Y\} = \bigcup_{n\in\mathbb{N}}\{d(X, Y) \geq n^{-1}\}$, por lo que basta probar que $\mathbb{P}(d(X, Y) \geq \epsilon) = 0$ para todo $\epsilon > 0$, lo cual se sigue de la relación

$$\{d(X, Y) \geq \epsilon\} \subset \left\{d(X, X_n) \geq \frac{\epsilon}{2}\right\} \cup \left\{d(X_n, Y) \geq \frac{\epsilon}{2}\right\}.$$

2) Para toda $n \in \mathbb{N}$ existe $N_n \in \mathcal{F}$ tal que $\mathbb{P}(N_n) = 0$ y $X_n(\omega) = Y_n(\omega)$ si $\omega \notin N_n$. Sea $N = \bigcup_{n\in\mathbb{N}} N_n$, por lo que tendremos que

$$\{d(Y_n, X) \geq \epsilon\} \subset N \cup \{d(X_n, X) \geq \epsilon\}.$$ □

Con el trabajo realizado hasta el momento podremos dar una generalización del teorema de convergencia dominada de Lebesgue, relajando la hipótesis de convergencia casi segura a convergencia en probabilidad.

Teorema 6.10. *Sean $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, Y variable aleatoria $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-medible, y $\{X_n\}_{n\in\mathbb{N}}$ y X variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles tales que $\|X_n\| \leq Y$ casi seguramente para cada $n \in \mathbb{N}$ y $Y \in L^p$. Si $X_n \xrightarrow{\mathbb{P}} X$ entonces $\|X\| \in L^p$ y $X_n \xrightarrow{L^p} X$.*

Demostración. Es claro que $\{\|X_n\|\}_{n\in\mathbb{N}} \subset L^p$. Por otra parte notemos que para $n, k \in \mathbb{N}$,

$$\{\|X\| \geq Y + k^{-1}\} \subset \{\|X\| \geq \|X_n\| + k^{-1}\} \subset \{\|X - X_n\| \geq k^{-1}\},$$

de donde $\|X\| \leq Y$ casi seguramente. Por lo tanto $\|X\| \in L^p$. Finalmente, observemos que por el teorema 6.6 para toda subsucesión $\{X_{n_k}\}_{k\in\mathbb{N}}$ existe una subsubsucesión $\{X_{n_{k_r}}\}_{r\in\mathbb{N}}$ tal que $X_{n_{k_r}} \xrightarrow{c.s.} X$. Esta subsubsucesión es tal que $\left\|X_{n_{k_r}} - X\right\| \xrightarrow{c.s.} 0$ y $\left\|X_{n_{k_r}} - X\right\| \leq 2Y$ casi seguramente para toda $r \in \mathbb{N}$. Por el teorema de convergencia dominada de Lebesgue, $X_{n_{k_r}} \xrightarrow{L^p} X$, de donde se sigue el resultado. □

Finalmente introducimos la relación entre la convergencia uniforme con las demás. Como se podrá intuir, este modo de convergencia será el más fuerte en general. Empero, como se verá, en espacios de medida finito, como lo son los espacios de probabilidad, resultará que es equivalente a la convergencia casi segura.

Proposición 6.11. *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad y* $\{X_n\}_{n\in\mathbb{N}}$ *y* X *variables aleatorias* $(\mathcal{F}, \mathcal{B}(S))$*-medibles tales que* $X_n \xrightarrow{c.u.} X$. *Entonces* $X_n \xrightarrow{c.s.} X$.

Demostración. Para $\delta = k^{-1}$ existe $F_k \in \mathcal{F}$ con $\mathbb{P}(F_k) < k^{-1}$ tal que $\{X_n\}_{n\in\mathbb{N}}$ converge uniformemente a X en $\Omega - F_k$. Sea $F = \bigcap_{k\in\mathbb{N}} F_k$, entonces $\mathbb{P}(F) = 0$ y si $\omega \notin F$ entonces existe $N \in \mathbb{N}$ tal que $\omega \notin F_N$, por lo que $X_n(\omega) \to X(\omega)$. □

Teorema 6.12 (Egorov)**.** *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad y* $\{X_n\}_{n\in\mathbb{N}}$ *y* X *variables aleatorias* $(\mathcal{F}, \mathcal{B}(S))$*-medibles tales que* $X_n \xrightarrow{c.s.} X$. *Entonces* $X_n \xrightarrow{c.u.} X$.

Demostración. Como $X_n \xrightarrow{c.s.} X$, entonces para todo $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m\geq n} \{d(X_n, X) \geq \epsilon\}\right) = 0.$$

Así, para $\delta > 0$ y $k \in \mathbb{N}$ dadas existe $n_k \in \mathbb{N}$ tal que

$$\mathbb{P}\left(\bigcup_{m\geq n_k} \{d(X_m, X) \geq k^{-1}\}\right) < 2^{-k}\delta.$$

Entonces $F = \bigcup_{k\in\mathbb{N}} \bigcup_{m\geq n_k} \{d(X_m, X) \geq k^{-1}\}$ es tal que $\mathbb{P}(F) < \delta$ y $\{X_n\}_{n\in\mathbb{N}}$ converge uniformemente a X en $\Omega - F$. □

Mediante el teorema de Egorov, podremos reducir el siguiente resultado al caso de convergencia segura y convergencia en probabilidad únicamente.

Teorema 6.13. *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad,* $\{X_n\}_{n\in\mathbb{N}}$ *y* X *variables aleatorias* $(\mathcal{F}, \mathcal{B}(S))$*-medibles y* $f : S \to E$ *una función continua, donde* (E, d_E) *es un espacio métrico separable.*

1) Si $X_n \xrightarrow{c.s.} X$ *entonces* $f(X_n) \xrightarrow{c.s.} f(X)$.

2) Si $X_n \xrightarrow{\mathbb{P}} X$ *entonces* $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Demostración. 1) Como $X_n \xrightarrow{c.s.} X$, existe $N \in \mathcal{F}$ con $\mathbb{P}(N) = 0$ y $X_n(\omega) \to X(\omega)$ para cualquier $\omega \notin N$. Por continuidad de f, $f(X_n(\omega)) \to f(X(\omega))$ para toda $\omega \notin N$, de donde $f(X_n) \xrightarrow{c.s.} f(X)$.

2) Sea $\{f(X_n)\}_{n\in\mathbb{N}}$ la sucesión de variables aleatorias que resulta de componer f y X_n. Consideremos cualquier subsucesión $\{f(X_{n_k})\}_{k\in\mathbb{N}}$; por el teorema 6.6 existe $\{X_{n_{k_r}}\}_{r\in\mathbb{N}}$ tal que $X_{n_{k_r}} \xrightarrow{c.s.} X$ y por (1), $f(X_{n_{k_r}}) \xrightarrow{c.s.} f(X)$. Por el mismo teorema 6.6 tendremos que $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. □

Para concluir la sección, se introducen dos conceptos nuevos, el de L^p*-acotado* y aquel de *uniformemente integrable*.

Definición 6.14. Diremos que $\{X_t\}_{t\in T}$, una colección de variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles, es ***$L^p$*-acotada** si

$$\sup_{t\in T} \mathbb{E}[\|X_t\|^p] < \infty.$$

Además diremos que es **uniformemente integrable** si

$$\lim_{K\to\infty} \sup_{t\in T} \mathbb{E}[\|X_t\| \, \mathbb{1}_{\{\|X_t\|>K\}}] = 0.$$

Lema 6.15. *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles tal que $\{\|X_n\|\}_{n\in\mathbb{N}} \subset L^1$. $\{X_n\}_{n\in\mathbb{N}}$ es uniformemente integrable si y sólo si*

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{E}[\|X_n\| \, \mathbb{1}_{\{\|X_n\|>K\}}] = 0.$$

Demostración. Como para cada $K \geq 0$, $\limsup_{n\to\infty} \mathbb{E}[X_n \mathbb{1}_{\{\|X_n\|>K\}}] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[X_n \mathbb{1}_{\{\|X_n\|>K\}}]$, tendremos que $\{X_n\}_{n\in\mathbb{N}}$ uniformemente integrable implica que

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{E}[\|X_n\| \, \mathbb{1}_{\{\|X_n\|>K\}}] = 0.$$

De forma recíproca sea $\epsilon > 0$ dado y notemos que existe K_∞ tal que

$$\inf_{n\in\mathbb{N}} \sup_{m\geq n} \mathbb{E}[\|X_m\| \, \mathbb{1}_{\{\|X_m\|>K_\infty\}}] \leq \frac{\epsilon}{2},$$

es decir que existe $N \in \mathbb{N}$ tal que para $n \geq N$, $\mathbb{E}[\|X_n\| \, \mathbb{1}_{\{\|X_n\|>K_\infty\}}] \leq \epsilon$. Como $\{\|X_j\|\}_{j=1}^{N-1} \subset L^1$, existen $K_1, \ldots, K_{N-1} > 0$ tales que $\mathbb{E}[\|X_j\| \, \mathbb{1}_{\{\|X_j\|>K_j\}}] \leq \epsilon$ para $j \in \{1, \ldots, N-1\}$, entonces para $K \geq \max\{K_1, \ldots, K_{N-1}, K_\infty\}$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}[\|X_n\| \, \mathbb{1}_{\{\|X_n\|>K\}}] \leq \epsilon.$$ □

Teorema 6.16. *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y $\{X_t\}_{t\in T}$ variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles. Entonces $\{X_t\}_{t\in T}$ es uniformemente integrable si y sólo si es L^1-acotada y para cada $\epsilon > 0$ existe un $\delta > 0$ tales que para cada $A \in \mathcal{F}$ con $\mathbb{P}(A) < \delta$, $\sup_{t\in T} \mathbb{E}[\|X_t\| \, \mathbb{1}_A] \leq \epsilon$.*

Demostración. Primero supongamos que $\sup_{t\in T} \mathbb{E}[\|X_t\|] \leq K < \infty$ para alguna $K > 0$. Para $\epsilon > 0$ sea $\delta > 0$ dada por las hipótesis. Sea $M > K/\delta$, entonces para cada $t \in T$,

$$\mathbb{P}(\|X_t\| > M) \leq \mathbb{P}(\|X_t\| > K/\delta) = \mathbb{E}[\mathbb{1}_{\{\|X_t\|>K/\delta\}}] < \frac{\delta}{K}\mathbb{E}[\|X_t\|] \leq \delta,$$

de donde $\sup_{t\in T} \mathbb{E}[\|X_t\| \, \mathbb{1}_{\{\|X_t\|>M\}}] < \epsilon$. Por lo tanto $\{X_t\}_{t\in T}$ es uniformemente integrable.

Inversamente sea $\{X_t\}_{t\in T}$ uniformemente integrable y $M > 0$ tal que

$$\sup_{t\in T} \mathbb{E}[\|X_t\| \, \mathbb{1}_{\{\|X_t\|>M\}}] \leq 1,$$

entonces para cada $t \in T$,

$$\mathbb{E}[\|X_t\|] = \mathbb{E}[\|X_t\| \, \mathbb{1}_{\{\|X_t\|\leq M\}}] + \mathbb{E}[\|X_t\| \, \mathbb{1}_{\{\|X_t\|>M\}}] \leq M + 1,$$

de donde $\{X_t\}_{t\in T}$ es L^1-acotada. Sea $\epsilon > 0$ arbitrario y $K > 0$ tal que

$$\sup_{t\in T} \mathbb{E}[\|X_t\|\, \mathbb{1}_{\{\|X_t\|>K\}}] \leq \frac{\epsilon}{2}.$$

Consideremos $\delta = \epsilon/(2K)$ y notemos que para cada $A \in \mathcal{F}$ con $\mathbb{P}(A) < \delta$ y para cada $t \in T$ tendremos que

$$\begin{aligned}\mathbb{E}[\|X_t\|\, \mathbb{1}_A] &= \mathbb{E}[\|X_t\|\, \mathbb{1}_A \mathbb{1}_{\{\|X_t\|\leq K\}}] + \mathbb{E}[\|X_t\|\, \mathbb{1}_A \mathbb{1}_{\{\|X_t\|>K\}}]\\ &\leq K\mathbb{E}[\mathbb{1}_A] + \mathbb{E}[\|X_t\|\, \mathbb{1}_{\{\|X_t\|>K\}}]\\ &< \epsilon.\end{aligned}$$

□

Proposición 6.17. *Sean* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espacio de probabilidad,* $p \in [1, \infty)$ *fija y variables aleatorias* $(\mathcal{F}, \mathcal{B}(S))$*-medibles,* X *y* $\{X_n\}_{n\in\mathbb{N}}$*, tales que* $\|X\| \in L^p$ *y* $\{\|X_n\|\}_{n\in\mathbb{N}} \subset L^p$*. Si* $X_n \xrightarrow{\mathbb{P}} X$ *entonces son equivalentes:*

1) $X_n \xrightarrow{L^p} X$*,*

2) $\lim_{n\to\infty} \mathbb{E}[\|X_n\|^p] = \mathbb{E}[\|X\|^p]$*,*

3) $\{\|X_n\|^p\}_{n\in\mathbb{N}}$ *es uniformemente integrable.*

Demostración. Que (1) implica (2) se sigue de la desigualdad de Minkowski, pues por ésta

$$\left|\mathbb{E}[\|X\|^p]^{\frac{1}{p}} - \mathbb{E}[\|X_n\|^p]^{\frac{1}{p}}\right| \leq \mathbb{E}[\|X_n - X\|^p]^{\frac{1}{p}}.$$

Para ver que (2) implica (3), notemos que para $K > 0$, $f(x) = \|x\|^p \wedge (K - \|x\|^p)_+$ es una función continua y acotada por $\frac{K}{2}$. Por el teorema 6.13, $\|X_n\|^p \wedge (K - \|X_n\|^p)_+ \xrightarrow{\mathbb{P}} \|X\|^p \wedge (K - \|X\|^p)_+$ y por el teorema 6.10, $\|X_n\|^p \wedge (K - \|X_n\|^p)_+ \xrightarrow{L^1} \|X\|^p \wedge (K - \|X\|^p)_+$. Por lo ya probado esto implica que

$$\lim_{n\to\infty} \mathbb{E}[\|X_n\|^p \wedge (K - \|X_n\|^p)_+] = \mathbb{E}[\|X\|^p \wedge (K - \|X\|^p)_+].$$

Además tendremos que $\|X_n\|^p\, \mathbb{1}_{\{\|X_n\|^p>K\}} \leq \|X_n\|^p - \|X_n\|^p \wedge (K - \|X_n\|^p)_+$, de donde

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{E}[\|X_n\|^p\, \mathbb{1}_{\{\|X_n\|^p>K\}}] = \lim_{K\to\infty} \mathbb{E}[\|X\|^p - \|X^p\| \wedge (K - \|X\|^p)_+] = 0,$$

donde la última igualdad se da por convergencia dominada de Lebesgue. Por el lema 6.15 se sigue el resultado al notar que $\{\|X_n\|^p\}_{n\in\mathbb{N}} \subset L^1$.

Finalmente, para ver que (3) implica (1), para $\epsilon > 0$ dado existen $K_1, K_2 > 0$ tales que

$$\sup_{n\in\mathbb{N}} \mathbb{E}[\|X_n\|^p\, \mathbb{1}_{\{\|X_n\|^p>2^{-p}K\}}] \leq 2^{-p-1}\epsilon$$

y $\mathbb{E}[\|X\|^p\, \mathbb{1}_{\{\|X\|^p>2^{-p}K\}}] \leq 2^{-p-1}\epsilon$ para toda $K \geq K_1 \vee K_2$. Entonces tendremos, para cada $n \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{E}[\|X_n - X\|^p\, \mathbb{1}_{\{\|X_n-X\|^p>K\}}]\\ &\qquad \leq 2^p\mathbb{E}[\max\{\|X_n\|^p, \|X\|^p\}\mathbb{1}_{\{\max\{\|X_n\|^p,\|X\|^p\}>2^{-p}K\}}]\\ &\qquad\qquad \leq 2^p\mathbb{E}[\|X_n\|^p\, \mathbb{1}_{\{\|X_n\|^p>K\}}] + 2^p\mathbb{E}[\|X\|^p\, \mathbb{1}_{\{\|X\|^p>2^{-p}K\}}] \leq \epsilon,\end{aligned}$$

de donde $\{\|X_n - X\|^p\}_{n\in\mathbb{N}}$ es uniformemente integrable. Además notemos que para cualquier $K > 0$, $\{\|X_n - X\|^p \wedge K\}_{n\in\mathbb{N}}$ es una sucesión uniformemente acotada tal que $\|X_n - X\|^p \wedge K \xrightarrow{\mathbb{P}} 0$. Para $\epsilon > 0$ dado sea $K' > 0$ tal que para toda $K \geq K'$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}[\|X_n - X\|^p \, \mathbb{1}_{\{\|X_n - X\|^p > K\}}] \leq \frac{\epsilon}{2}.$$

Por el teorema 6.10 se sigue que para $K > K'$ fija existe N_1 tal que para toda $n \geq N_1$,

$$\mathbb{E}[\|X_n - X\|^p \wedge K] < \frac{\epsilon}{2}.$$

Entonces tendremos que para $n \geq N_1$,

$$\begin{aligned}
\mathbb{E}[\|X_n - X\|^p] &= \mathbb{E}[\|X_n - X\|^p \, \mathbb{1}_{\{\|X_n - X\|^p > K\}}] + \mathbb{E}[\|X_n - X\|^p \, \mathbb{1}_{\{\|X_n - X\|^p \leq K\}}] \\
&\leq \frac{\epsilon}{2} + \mathbb{E}[\|X_n - X\|^p \wedge K] \\
&< \epsilon.
\end{aligned}$$

Así $X_n \xrightarrow{L^p} X$. □

6.2. Convergencia en ley

Las convergencias estudiadas hasta el momento, aunque útiles en la rama de probabilidad, no son exhaustivas. Existe otro tipo de convergencia que resulta ser funamentalmente distinto: la *convergencia débil* o *convergencia en ley*.

Como el mismo nombre lo indica, este modo de convergencia será más débil que los introducidos hasta ahora, por lo que se requerirá menos para que una sucesión de variables aleatorias tenga un límite en el sentido débil. Lo sorprendente, y distinto a los demás modos de convergencia, es que no importan los valores que tomen las variables aleatorias, sino las probabilidades de que tomen esos valores, es decir que nos importaran las distribuciones (o medidas inducidas).

En este sentido, sabemos que si (S, τ) es un espacio topológico y $\mu : \mathcal{B}(S) \to \mathbb{C}$ es una medida compleja, podemos definir su variación total $\|\mu\|_{TV} = |\mu|(S)$ y decir que $\{\mu_n\}_{n\in\mathbb{N}}$ converge en variación total a μ, $\mu_n \xrightarrow{TV} \mu$, si $\lim_{n\to\infty} \|\mu_n - \mu\|_{TV} = 0$. Esta noción de convergencia resulta ser demasiado fuerte en general: notemos que si $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R} - \{0\}$ es tal que $x_n \to 0$ entonces $\delta_{x_n} \overset{TV}{\not\longrightarrow} \delta_0$. Por lo tanto deberemos definir otra noción de convergencia de medidas que sea más débil. En lo que resta de la sección, todas las medidas que se consideren serán de probabilidad.

Definición 6.18. Si (S, τ) es un espacio topológico, diremos que $\{\mu_n\}_{n\in\mathbb{N}}$ **converge débilmente** a μ, denotado por $\mu_n \xrightarrow{d} \mu$, si para toda $f \in C_b(S)$,

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu.$$

Notemos que para todas las medidas de probabilidad, si f es continua y acotada entonces será integrable y retomando el ejemplo anterior a la definición, $\delta_{x_n} \xrightarrow{d} \delta_0$ por continuidad de las funciones $f \in C_b(S)$.

Definición 6.19. Una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ **converge en distribución** a X, todas con valores en (S,τ), denotado por $X_n \xrightarrow{d} X$, si $\mathbb{P}_{X_n} \xrightarrow{d} \mathbb{P}$.

Un resultado directo de las definiciones anteriores y el teorema 2.38 es el siguiente.

Proposición 6.20. *$X_n \xrightarrow{d} X$, donde todas las variables aleatorias toman valores en (S,τ), si y sólo si para toda función $f \in C_b(S)$,*

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

Notemos que, en general, no hay razón para que las variables aleatorias que consideremos estén definidas sobre un mismo espacio de probabilidad; en cambio, los modos de convergencia de la sección anterior requieren, *a priori*, que todas las variables aleatorias estén definidas en un mismo espacio de probabilidad. Cuando ocurra este fenómeno, podremos relacionar la convergencia en distribución con las demás con el siguiente resultado.

Teorema 6.21. *Sean $\{X_n\}_{n\in\mathbb{N}}$ y X variables aleatorias $(\mathcal{F},\mathcal{B}(S))$-medibles. Si $X_n \xrightarrow{\mathbb{P}} X$ entonces $X_n \xrightarrow{d} X$ y hay equivalencia si X es casi seguramente constante.*

Demostración. Sea $f \in C_b(S)$ arbitraria, entonces $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ y como $\{f(X_n)\}_{n\in\mathbb{N}}$ es una sucesión acotada, por el teorema 6.10, $f(X_n) \xrightarrow{L^1} f(X)$. Por lo tanto tendremos que

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \mathbb{E}[|f(X_n) - f(X)|] \xrightarrow{n\to\infty} 0.$$

Así las cosas, $X_n \xrightarrow{d} X$.

Recíprocamente, si $X = y \in S$ casi seguramente, como la función $f(x) = \frac{d(x,y)}{1+d(x,y)}$ es continua y acotada,

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{d(X_n,y)}{1+d(X_n,y)}\right] = \mathbb{E}\left[\frac{d(X,y)}{1+d(X,y)}\right] = 0.$$

El teorema 6.7 nos permite concluir. $\square$

A primera vista, parece tentador decir que $X_n \xrightarrow{d} X$ implica que $\mathbb{P}(X_n \in A)$ converge a $\mathbb{P}(X \in A)$ para todos los conjuntos $A \in \mathcal{B}(S)$, lo cual *es en general falso*, y los conjuntos en los que los haga serán especiales, como veremos en el siguiente teorema.

Teorema 6.22 (de Portmanteau)**.** *Para variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y X con valores en un espacio métrico (S,d) son equivalentes:*

1) $X_n \xrightarrow{d} X$,

2) $\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ *para cualquier $G \subset S$ abierto,*

3) $\limsup_{n\in\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ *para cualquier $F \subset S$ cerrado,*

4) $\lim_{n\to\infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$ *para cualquier $B \in \mathcal{B}(S)$ con $X \notin \partial B$ c.s.*

A los borelianos $B \in \mathcal{B}(S)$ tales que $X \notin \partial B$ c.s. se les conoce como *conjuntos de X-continuidad*.

Demostración. Primero supongamos que se satisface (1) y fijemos $G \subset S$ abierto; definamos $d(x, G^c) := \inf\{d(x, y) : y \in G^c\}$ y, para cada $k \in \mathbb{N}$, sea $f_k(x) = \min\{1, kd(x, G^c)\}$. Tenemos que $f_k \in C_b(S)$, $f_k \le f_{k+1}$ y $f_k \le \mathbb{1}_G$ para cada $k \in \mathbb{N}$, por lo que $\mathbb{E}[f_k(X_n)] \le \mathbb{P}(X_n \in G)$ para cada $n \in \mathbb{N}$, así para $k \in \mathbb{N}$ fija,

$$\mathbb{E}[f_k(X)] = \lim_{n\to\infty} \mathbb{E}[f_k(X_n)] \le \liminf_{n\to\infty} \mathbb{P}(X_n \in G).$$

Finalmente se sigue, por convergencia dominada, que

$$\mathbb{P}(X \in G) = \lim_{k\to\infty} \mathbb{E}[f_k(X)] \le \liminf_{n\to\infty} \mathbb{P}(X_n \in G),$$

obteniendo que (2) se cumple.

La equivalencia entre (2) y (3) se deduce al considerar complementos, por lo que podremos asumir que ambas se cumplen para probar (4). Notemos que para cualquier $B \in \mathcal{B}(S)$,

$$\mathbb{P}(X \in B^\circ) \le \liminf_{n\to\infty} \mathbb{P}(X_n \in B) \le \limsup_{n\to\infty} \mathbb{P}(X_n \in B) \le \mathbb{P}(X \in \overline{B}).$$

Si en particular $X \notin \partial B$ c.s., entonces se tiene que $\mathbb{P}(X \in B^\circ) = \mathbb{P}(X \in \overline{B}) = \mathbb{P}(X \in B)$ y se sigue el resultado.

Ahora supongamos que se satisface (4). Fijemos $F \subset S$ cerrado y definamos, para $\epsilon > 0$, $F^\epsilon := \{x \in S : d(x, F) \le \epsilon\}$, donde $d(x, F)$ se define como al inicio de la prueba. Notemos que $\partial F^\epsilon \subset \{x \in S : d(x, F) = \epsilon\}$ y que $\{x \in S : d(x, F) = \epsilon_1\} \cap \{x \in S : d(x, F) = \epsilon_2\} = \emptyset$ cuando $\epsilon_1 \ne \epsilon_2$. Usando este hecho, que $F^c = \bigcup_{\epsilon>0}\{x \in S : d(x, F) = \epsilon\}$ y que $\mathbb{P}(F^c) \le 1$ se sigue que $\mathbb{P}(X \in F^\epsilon) = 0$ λ-c.d.s. Para $\epsilon > 0$ que satisfaga la igualdad anterior tendremos que $\mathbb{P}(X_n \in F) \le \mathbb{P}(X_n \in F^\epsilon)$ y por lo tanto

$$\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \le \mathbb{P}(X \in F^\epsilon).$$

Al hacer $\epsilon \to 0$ se obtiene (3) y consecuentemente (2). Ahora sea $f \in C_b(S, \mathbb{R})$ no-negativa, entonces por (2) y el lema de Fatou,

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_{\mathbb{R}_+} \mathbb{P}(f(X) > t)d\lambda(t) \le \int_{\mathbb{R}_+} \liminf_{n\to\infty} \mathbb{P}(f(X_n) > t)d\lambda(t) \\ &\le \liminf_{n\to\infty} \int_{\mathbb{R}_+} \mathbb{P}(f(X_n) > t)d\lambda(t)) = \liminf_{n\to\infty} \mathbb{E}[f(X_n)].\end{aligned}$$

Si ahora consideramos $f \in C_b(S, \mathbb{R})$, entonces $|f| \le K < \infty$; aplicando la desigualdad anterior a $K \pm f$ tendremos que $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$. Para el caso general $f \in C_b(S)$ basta separar en parte real e imaginaria para concluir. $\square$

Teorema 6.23. *Si (S, d_S) y (E, d_E) son dos espacios métricos y $f : S \to E$ es una función continua. Si además $\{X_n\}_{n\in\mathbb{N}}$ y X son variables aleatorias con valores en S tales que $X_n \xrightarrow{d} X$ entonces $f(X_n) \xrightarrow{d} f(X)$.*

Demostración. Sea $F \subset E$ cerrado, entonces $f^{-1}(F) \subset S$ es cerrado. Entonces tendremos que

$$\limsup_{n\to\infty} \mathbb{P}(f(X_n) \in F) = \limsup_{n\to\infty} \mathbb{P}(X_n \in f^{-1}(F)) \le \mathbb{P}(X \in f^{-1}(F)) = \mathbb{P}(f(X) \in F),$$

y se concluye por el teorema anterior. $\square$

Definición 6.24. Diremos que $\{X_t\}_{t\in T}$, una colección de variables aleatorias $(\mathcal{F}, \mathcal{B}(S))$-medibles, es **tensa** si

$$\lim_{K\to\infty} \sup_{t\in T} \mathbb{P}(\|X_t\| > K) = 0.$$

De forma análoga, si $\{\mu_t\}_{t\in T}$ es una colección de medidas sobre $(S, \mathcal{B}(S))$, diremos que es tensa si

$$\lim_{K\to\infty} \sup_{t\in T} \mu_t(\{x \in S : \|x\| > K\}) = 0.$$

En el caso en el que $T = \mathbb{N}$, se obtiene el siguiente lema, cuya prueba es análoga a la del lema 6.15 y se omite.

Lema 6.25. *Una sucesión de variables aleatorias, $\{X_n\}_{n\in\mathbb{N}}$, es tensa si y sólo si*

$$\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}(\|X_n\| > K) = 0.$$

Proposición 6.26. *Sean $\{X_n\}_{n\in\mathbb{N}}$ y X variables aleatorias con valores en un espacio de Banach $(S, \|\cdot\|)$. Si $X_n \xrightarrow{d} X$ entonces $\{X_n\}_{n\in\mathbb{N}}$ es tensa.*

Demostración. Para cualquier $K > 0$ sea $f_K(x) = (1 - (K - \|x\|)_+)_+$, una función continua y acotada que satisface $\mathbb{1}_{\{\|Y\|>K\}} \leq f(Y) \leq \mathbb{1}_{\{\|Y\|>K-1\}}$ para cualquier variable aleatoria con valores en S. Entonces se sigue que

$$\limsup_{n\to\infty} \mathbb{P}(\|X_n\| > K) \leq \lim_{n\to\infty} \mathbb{E}[f_K(X_n)] = \mathbb{E}[f_K(X)] \leq \mathbb{P}(\|X\| > K - 1).$$

Tomando el límite cuando $K \to \infty$, se sigue el resultado por el lema anterior. $\square$

Lema 6.27. *Para cualesquiera medida de probabilidad μ sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y $K > 0$ se satisfacen*

$$\mu([-K, K]^c) \leq \frac{K}{2} \int_{-2/K}^{2/K} (1 - \hat{\mu}(t))dt,$$

$$\mu([-K, K]) \leq 2K \int_{-1/K}^{1/K} |\hat{\mu}(t)|\, dt.$$

Lema 6.28. *Sean μ una medida de probabilidad sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ y $\mathrm{pr}_j : \mathbb{R}^d \to \mathbb{R}^d$ la proyección al subespacio generado por el j-ésimo vector canónico, e_j. Entonces*

$$\hat{\mu}(\mathrm{pr}_j(t)) = \widehat{\mathrm{pr}_{j\,*}\mu}(t).$$

Demostración. Por el teorema de cambio de variable y la definición de función característica,

$$\hat{\mu}(\mathrm{pr}_j(t)) = \int e^{i\langle \mathrm{pr}_j(t), x\rangle} d\mu(x) = \int e^{i\langle t, \mathrm{pr}_j(x)\rangle} d\mu(x) = \int e^{i\langle t,x\rangle} d\,\mathrm{pr}_{j\,*}\mu(x) = \widehat{\mathrm{pr}_{j\,*}\mu}(t). \quad \square$$

Teorema 6.29 (de continuidad de Lévy)**.** *Sean $\{\mu_n\}_{n\in\mathbb{N}}$ y μ medidas de probabilidad sobre el espacio medible $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Una condición suficiente y necesaria para que $\mu_n \xrightarrow{d} \mu$ es $\lim_{n\to\infty} \hat{\mu}_n(t) = \hat{\mu}(t)$ para toda $t \in \mathbb{R}^d$.*

Demostración. Si $\mu_n \xrightarrow{d} \mu$ entonces, al ser $x \mapsto e^{i\langle t,x\rangle}$ continua y acotada para cada $t \in \mathbb{R}^d$, $\hat{\mu}_n(t) \to \hat{\mu}(t)$ cuando $n \to \infty$.

De forma recíproca, para cualesquiera $j \in \{1,\ldots,d\}$ y $K > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(\{x \in \mathbb{R}^d : \left|\mathrm{pr}_j(x)\right| > K\}) &\leq \lim_{n\to\infty} \frac{K}{2}\int_{-2/K}^{2/K} \big(1 - \hat{\mu}_n(te_j)\big)dt \\ &= \frac{K}{2}\int_{-2/K}^{2/K} \big(1 - \hat{\mu}(te_j)\big)dt,\end{aligned}$$

donde la última desigualdad de da por el teorema de convergencia dominada. Recordando que $\hat{\mu}$ es continua en cero, el lado derecho de la última desigualdad tiende a cero cuando $K \to \infty$, lo cual implica que $\{\mu_n\}_{n\in\mathbb{N}}$ es tensa, pues

$$\begin{aligned}&\lim_{K\to\infty} \limsup_{n\to\infty} \mu_n(\{x \in \mathbb{R}^d : |x| > K\}) \\ &\quad\leq \lim_{K\to\infty} \limsup_{n\to\infty} \sum_{j=1}^{d} \mu_n(\{x \in \mathbb{R}^d : \left|\mathrm{pr}_j(x)\right| > K\}) \\ &\quad\leq \sum_{j=1}^{d} \lim_{K\to\infty} \limsup_{n\to\infty} \mu_n(\{x \in \mathbb{R}^d : \left|\mathrm{pr}_j(x)\right| > K\}) = 0.\end{aligned}$$

Así, para cualquier $\epsilon > 0$ existirá $K > 0$ tal que $\mu_n(\{x \in \mathbb{R}^d : |x| > K\}) \leq \epsilon$ para toda $n \in \mathbb{N}$ y $\mu(\{x \in \mathbb{R}^d : |x| > K\}) \leq \epsilon$. Ahora consideremos $f \in C_b(\mathbb{R}^d, \mathbb{R})$, con $|f| \leq M < \infty$. Sea f_K su restricción a $\{x \in \mathbb{R}^d : |x| \leq K\}$. Extendamos f_K a una función $f_0 \in C_0(\mathbb{R}^d, \mathbb{R})$ tal que $|f_0| \leq M$. De la prueba del teorema 5.22 sea g una función en el espacio generado por $\{f_{\sigma,v}; \sigma \in \mathbb{R} - \{0\}, v \in \mathbb{R}^d\}$ tal que $\|f_0 - g\|_u < \epsilon$. Refiriéndonos a la prueba del teorema 5.22, se puede ver que para g,

$$\lim_{n\to\infty} \int g d\mu_n = \int g d\mu.$$

Además, para cualquier $n \in \mathbb{N}$,

$$\begin{aligned}\left|\int f d\mu_n - \int g d\mu_n\right| &\leq \int |f - f_0|\, d\mu_n + \int |f_0 - g|\, d\mu_n \\ &\leq \|f - f_0\|_u\, \mu(\{x \in \mathbb{R}^d : |x| > K\}) + \|f_0 - g\|_u \\ &\leq (2M+1)\epsilon.\end{aligned}$$

Intercambiando μ_n por μ se obtiene una desigualdad similar para μ. Por lo tanto tendremos que

$$\begin{aligned}\left|\int f d\mu_n - \int f d\mu\right| &\leq \left|\int f d\mu_n - \int g d\mu_n\right| + \left|\int g d\mu_n - \int g d\mu\right| + \left|\int g d\mu - \int f d\mu\right| \\ &\leq 2(2M+1)\epsilon + \left|\int g d\mu_n - \int g d\mu\right|,\end{aligned}$$

de donde se sigue que

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu,$$

y como $f \in C_b(\mathbb{R}^d, \mathbb{R})$ fue arbitraria se sigue que el resultado es válido para toda $f \in C_b(\mathbb{R})$ al separar en parte real e imaginaria. Por lo tanto $\mu_n \xrightarrow{d} \mu$. $\square$

Este teorema que se acaba de probar nos permitirá demostrar con facilidad uno de los teoremas más importantes de la probabilidad moderna: el teorema central del límite. Una primera aproximación al resultado se da en el siguiente teorema.

Teorema 6.30 (Teorema central del límite, IID). *Si* $\{X_n\}_{n\in\mathbb{N}} \subset L^2$ *y* $X \in L^2$ *son variables aleatorias* $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$*-medibles, independientes e idénticamente distribuidas, tales que* $\mathbb{E}[X] = m$ *y* $\mathrm{Var}(X) = \sigma^2 > 0$ *y* $Z \sim \mathrm{N}(0,1)$*, entonces*

$$\frac{\sum_{k=1}^n X_n - nm}{\sqrt{n\sigma^2}} \xrightarrow{d} Z.$$

Demostración. Supongamos, sin pérdida de generalidad, que $m = 0$ y que $\sigma^2 = 1$. Si denotamos por φ a la función característica de X y φ_n a aquella de $n^{-1/2}\sum_{k=1}^n X_k$, tendremos que

$$\varphi_n(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Combinando el teorema 5.19 y el lema 5.26, para toda $t \in \mathbb{R}$,

$$\left|\varphi(t) - \left(1 + it\mathbb{E}[X] - \frac{1}{2}t^2\mathbb{E}[X^2]\right)\right| \le \mathbb{E}\left[\mathrm{mín}\left\{t^2X^2, \frac{|tX|^3}{6}\right\}\right], \tag{6.1}$$

de donde, para n suficientemente grande,

$$\varphi_n(t) = \left(1 - \frac{1}{2}\cdot\frac{t^2}{n} + \mathrm{o}\left(\frac{|t|^3}{n^{3/2}}\right)\right)^n,$$

donde se usó que $\mathbb{E}[X] = 0$ y $\mathrm{Var}(X) = \mathbb{E}[X^2] = 1$. De esta última expresión se sigue que

$$\lim_{n\to\infty}\varphi_n(t) = \exp\left\{-\frac{1}{2}t^2\right\},$$

que es la función característica de Z. Concluimos por el teorema de continuidad de Lévy. $\square$

Para extender el teorema anterior a sucesiones que no sean necesariamente idénticamente distribuidas, probaremos antes un lema que nos ayudará en la demostración.

Lema 6.31. *Sean* $\{z_n\}_{n\in\mathbb{N}} \subset \mathbb{C}$ *y* $\{w_n\}_{n\in\mathbb{N}} \subset \mathbb{C}$ *tales que* $|z_n| \le 1$ *y* $|w_n| \le 1$ *para toda* $n \in \mathbb{N}$. *Entonces para cada* $n \in \mathbb{N}$,

$$\left|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right| \le \sum_{k=1}^n |z_k - w_k|.$$

Demostración. Procedemos por inducción. El caso $n = 1$ es evidente. Para el paso inductivo, notemos que

$$\begin{aligned}
\left|\prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} w_k\right| &= \left|(z_{n+1} - w_{n+1})\left(\prod_{k=1}^n z_k\right) - w_{n+1}\left(\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right)\right| \\
&\le |z_{n+1} - w_{n+1}| + \left|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right| \\
&\le \sum_{k=1}^{n+1} |z_k - w_k|,
\end{aligned}$$

donde la última se da por hipótesis de inducción. $\square$

Teorema 6.32 (Teorema central del límite, Lindeberg)**.** *Sean* $\{X_n\}_{n\in\mathbb{N}} \subset L^2$ *independientes, con* $\mathbb{E}[X_n] = m_n$ *y* $\mathrm{Var}(X_n) = \sigma_n^2 > 0$ *para toda* $n \in \mathbb{N}$. *Definamos* $S_n = \sum_{k=1}^n X_k$, $s_n^2 = \sum_{k=1}^n \sigma_k^2$ *y*

$$L_\epsilon(n) = \frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[(X_k - m_k)^2 \mathbb{1}_{\{|X_k - m_k| > \epsilon s_n\}}\right].$$

Si

$$\lim_{n\to\infty} L_\epsilon(n) = 0 \tag{CL}$$

entonces

$$\frac{1}{s_n}\sum_{k=1}^n (X_k - m_k) \xrightarrow{d} Z \sim \mathrm{N}(0,1).$$

A la condición (CL) se le conoce como *condición de Lindeberg*.

Demostración. Se puede suponer, sin pérdida de generalidad, que $m_n = 0$ para toda $n \in \mathbb{N}$. Así, el resultado se seguirá al probar que

$$\lim_{n\to\infty}\left|\varphi_{\frac{S_n}{s_n}}(t) - e^{-\frac{t^2}{2}}\right| = 0.$$

Notando que $e^{-t^2/2} = \prod_{k=1}^n \exp\left\{-\frac{\sigma_k^2 t^2}{2s_n^2}\right\}$ para cada $n \in \mathbb{N}$, será equivalente probar que

$$\lim_{n\to\infty}\left|\prod_{k=1}^n \varphi_{X_k}\left(\frac{t}{s_n}\right) - \prod_{k=1}^n \exp\left\{-\frac{\sigma_k^2 t^2}{2s_n^2}\right\}\right| = 0.$$

Por el lema anterior será suficiente probar que

$$\lim_{n\to\infty}\sum_{k=1}^n \left|\varphi_{X_k}\left(\frac{t}{s_n}\right) - \exp\left\{-\frac{\sigma_k^2 t^2}{2s_n^2}\right\}\right| = 0.$$

Por (6.1), los primeros dos términos distintos de cero de la expansión de Taylor de las funciones características y la exponencial coinciden y, por la desigualdad del triángulo, bastará probar que

$$\lim_{n\to\infty}\sum_{k=1}^n \left|\varphi_{X_k}\left(\frac{t}{s_n}\right) - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right)\right| = 0,$$

$$\lim_{n\to\infty}\sum_{k=1}^n \left|\exp\left\{-\frac{\sigma_k^2 t^2}{2s_n^2}\right\} - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right)\right| = 0.$$

Notando que el segundo límite es un caso particular del primero, bastará probar

$$\lim_{n\to\infty}\sum_{k=1}^n \left|\varphi_{X_k}\left(\frac{t}{s_n}\right) - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right)\right| = 0.$$

Así las cosas,

$$\begin{aligned}\sum_{k=1}^{n}\left|\varphi_{X_k}\left(\frac{t}{s_n}\right)-\left(1-\frac{\sigma_k^2 t^2}{2s_n^2}\right)\right| &\le \sum_{k=1}^{n}\mathbb{E}\left[\text{mín}\left\{\frac{t^2X_k^2}{s_n^2},\frac{|tX_k|^3}{6s_n^3}\right\}\right]\\ &\le \sum_{k=1}^{n}\mathbb{E}\left[\frac{|tX_k|^3}{6s_n^3}\mathbb{1}_{\{|X_k|\le\epsilon s_n\}}\right]+\mathbb{E}\left[\frac{t^2X_k^2}{s_n^2}\mathbb{1}_{\{|X_k|>\epsilon s_n\}}\right]\\ &\le t^2L_\epsilon(n)+\sum_{k=1}^{n}\frac{|t|^3\epsilon}{6s_n^2}\mathbb{E}\left[X_k^2\mathbb{1}_{\{|X_k|\le\epsilon s_n\}}\right]\\ &\le t^2L_\epsilon(n)+\frac{|t|^3\epsilon}{6}.\end{aligned}$$

Luego, por la condición (CL),

$$\limsup_{n\to\infty}\sum_{k=1}^{n}\left|\varphi_{X_k}\left(\frac{t}{s_n}\right)-\left(1-\frac{\sigma_k^2t^2}{2s_n^2}\right)\right|\le\frac{|t|^3\epsilon}{6}.$$

El resultado se sigue pues $\epsilon > 0$ es arbitrario. □

6.3. Ley fuerte de grandes números

En esta sección se prueba otro de los teoremas más importantes de la probabilidad moderna: la ley fuerte de grandes números.

Lema 6.33. *Si $\{x_n\}_{n\in\mathbb{N}}\subset\mathbb{C}$ es tal que $x_n\to x$ entonces $n^{-1}\sum_{k=1}^{n}x_n\to x$.*

Demostración. Como la sucesión $\{x_n\}_{n\in\mathbb{N}}$ es convergente, es acotada. Supongamos que $|x_n|\le M$ para toda $n\in\mathbb{N}$. Para $\epsilon>0$ sea $N\in\mathbb{N}$ tal que $|x-x_n|<\epsilon/2$ para toda $n\ge N$. Así, si $n>N\vee(4NM/\epsilon)$,

$$\left|x-\frac{1}{n}\sum_{k=1}^{n}x_k\right|\le\frac{1}{n}\sum_{k=1}^{N-1}2M+\frac{1}{n}\sum_{k=N}^{n}\frac{\epsilon}{2}<\epsilon.$$ □

Teorema 6.34. *(Ley fuerte de grandes números) Sea $\{X_n\}_{n\in\mathbb{N}}\subset L^1$ una sucesión de variables aleatorias reales independientes e idénticamente distribuidas tales que $\mathbb{E}[X_1]=m$. Si $S_n=\sum_{k=1}^{n}X_k$,*

$$\frac{1}{n}S_n\xrightarrow{c.s.}m.$$

Demostración. (Etemadi). Notemos primero que si el teorema es válido para variables aleatorias no negativas entonces

$$\frac{1}{n}S_n=\frac{1}{n}\sum_{k=1}^{n}X_k^+-\frac{1}{n}\sum_{k=1}^{n}X_k^-\xrightarrow{c.s.}\mathbb{E}[X_1^+]-\mathbb{E}[X_1^-]=m.$$

Así podremos suponer que $X_n\ge0$ para toda $n\in\mathbb{N}$.

Definamos $Y_n = X_n \mathbb{1}_{\{X_n \le n\}}$ y $T_n = \sum_{k=1}^n Y_k$ para cada $n \in \mathbb{N}$. Sean $\alpha > 1$ fija y $u_n = \lfloor \alpha^n \rfloor$ para cada $n \in \mathbb{N}$. Primero probaremos que para $\epsilon > 0$ arbitraria,

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left| \frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n} \right| > \epsilon \right) < \infty.$$

Con este fin, por independencia tendremos que

$$\operatorname{Var}(T_n) = \sum_{k=1}^n \operatorname{Var}(Y_k) \le \sum_{k=1}^n \mathbb{E}[Y_k^2] = \sum_{k=1}^n \mathbb{E}[X_1^2 \mathbb{1}_{\{X_1 \le k\}}] \le n\mathbb{E}[X_1^2 \mathbb{1}_{\{X_1 \le n\}}].$$

Entonces, por la desigualdad de Chebyshev,

$$\begin{aligned}\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left| \frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n} \right| > \epsilon \right) &\le \sum_{n\in\mathbb{N}} (\epsilon u_n)^{-2} \operatorname{Var}(T_{u_n}) \\ &\le \sum_{n\in\mathbb{N}} \frac{1}{\epsilon^2 u_n} \mathbb{E}[X_1^2 \mathbb{1}_{\{X_1 \le u_n\}}] \\ &= \epsilon^{-2} \mathbb{E}\left[X_1^2 \sum_{n\in\mathbb{N}} \frac{1}{u_n} \mathbb{1}_{\{X_1 \le u_n\}} \right].\end{aligned}$$

Ahora sean $K = 2\alpha/(\alpha - 1)$ y $x > 0$. Si $N = \operatorname{ínf}\{n \in \mathbb{N} : u_n \ge x\}$, $\alpha^N \ge x$ y de $y \le 2\lfloor y \rfloor$ para $y \ge 1$ se seguirá que

$$\sum_{u_n \ge x} \frac{1}{u_n} \le 2 \sum_{n \ge N} \alpha^{-n} = K\alpha^{-N} \le \frac{K}{x}.$$

De esta manera, $\sum_{n\in\mathbb{N}} u_n^{-1} \mathbb{1}_{\{X_1 \le u_n\}} \le K X_1^{-1}$ cuando $X_1 > 0$. Jutando este resultado con la desigualdad que se tenía,

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left| \frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n} \right| > \epsilon \right) \le \epsilon^{-2} K m < \infty.$$

Por Borel–Cantelli tendremos entonces que

$$\mathbb{P}\left(\limsup_{n\to\infty} \left\{ \left| \frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n} \right| > \epsilon \right\} \right) = 0,$$

y

$$\mathbb{P}\left(\limsup_{n\to\infty} \left\{ \left| \frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n} \right| > \epsilon \right\} \right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m \ge n} \left\{ \left| \frac{T_{u_m} - \mathbb{E}[T_{u_m}]}{u_m} \right| > \epsilon \right\} \right) = 0,$$

es decir

$$\frac{T_{u_n} - \mathbb{E}[T_{u_n}]}{u_n} \xrightarrow{c.s.} 0.$$

Del lema anterior, como $\lim_{n\to\infty} \mathbb{E}[Y_n] = m$, $\lim_{n\to\infty} n^{-1}\mathbb{E}[T_n] = m$, y así

$$\frac{1}{u_n} T_{u_n} \xrightarrow{c.s.} m.$$

Ahora, notando que $\mathbb{P}(X_1 > n) = \int_{n-1}^{n} \mathbb{P}(X_1 > n)dt \leq \int_{n-1}^{n} \mathbb{P}(X_1 > t)dt$,

$$\sum_{n\in\mathbb{N}} \mathbb{P}(X_n \neq Y_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(X_1 > n) \leq \int_0^\infty \mathbb{P}(X_1 > t)dt = \mathbb{E}[X_1] < \infty.$$

Una nueva aplicación de Borel–Cantelli nos dice que

$$\mathbb{P}\left(\liminf_{n\to\infty} \{X_n = Y_n\}\right) = 1.$$

Por lo tanto, con probabilidad uno

$$\lim_{n\to\infty} \frac{T_n - S_n}{n} = 0.$$

Así las cosas,

$$\frac{S_{u_n}}{u_n} \xrightarrow{c.s.} m.$$

Para finalizar, si $u_n \leq k \leq n+1$, como $X_i \geq 0$,

$$\frac{u_n}{u_{n+1}} \cdot \frac{S_{u_n}}{u_n} = \frac{S_{u_n}}{u_{n+1}} \leq \frac{S_k}{k} \leq \frac{S_{u_{n+1}}}{u_n} = \frac{S_{u_{n+1}}}{u_{n+1}} \cdot \frac{u_{n+1}}{u_n}$$

y, dado que $\frac{u_{n+1}}{u_n} \to \alpha$, con probabilidad uno se cumplirá que

$$\frac{1}{\alpha} m \leq \liminf_{k\to\infty} \frac{S_k}{k} \leq \limsup_{k\to\infty} \frac{S_k}{k} \leq \alpha m.$$

Esto es,

$$E(\alpha) = \left\{ \frac{1}{\alpha} m \leq \liminf_{k\to\infty} \frac{S_k}{k} \leq \limsup_{k\to\infty} \frac{S_k}{k} \leq \alpha m \right\}$$

es tal que $\mathbb{P}(E(\alpha)) = 1$. Definiendo $E = \bigcap_{\alpha\in\mathbb{Q}\cap(1,\infty)} E(\alpha)$, tendremos que $\mathbb{P}(E) = 1$ y

$$E = \left\{ \lim_{n\to\infty} \frac{S_n}{n} = m \right\}. \qquad \square$$

6.4. Convergencia vaga

En esta sección se introduce un nuevo tipo de convergencia, más débil que los introducidos hasta el momento. Esta convergencia, llamada *convergencia vaga*, nos permitirá ver bajo una nueva luz el teorema de continuidad de Lévy y, por lo tanto, la convergencia en distribución.

Con este fin en mente consideremos (S, d), un espacio métrico completo y separable, con su σ-álgebra de Borel $\mathcal{B}(S)$ y una medida μ sobre $(S, \mathcal{B}(S))$. Diremos que μ es *localmente finita* si para todo $A \in \mathcal{B}(S)$ acotado, *i.e.* existen $x \in A$ y $r > 0$ tales que $A \subset B_r(x)$, $|\mu(A)| < \infty$. Notemos que toda medida compleja es localmente finita. Al espacio de medidas localmente finitas sobre (S, d) lo denotaremos por $\mathcal{M}_l(S)$. En este espacio podemos introducir la topología generada por los mapeos $\hat{f} : \mathcal{M}_l(S) \to \mathbb{C}$ definidos por $\mu \mapsto \int f d\mu$, con $f \in C_c(S)$; esto es la mínima topología tal que $\hat{f}$ es continua para toda $f \in C_c(S)$. A esta topología se le conoce como **topología vaga**.

Definición 6.35. Diremos que $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(S)$ *converge vagamente* a $\mu \in \mathcal{M}_l(S)$, denotado por $\mu_n \xrightarrow{v} \mu$, si para cada $f \in C_c(S)$ se satisface

$$\lim_{n\to\infty} f d\mu_n = \int f d\mu.$$

Si en particular $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ son medidas de probabilidad tales que $\mu_n \xrightarrow{v} \mu$ y para cada $n \in \mathbb{N}$ sea $f_n \in C_c(\mathbb{R}^d)$ con $\operatorname{supp} f_n = \overline{B_n(0)}$, $\mathbb{1}_{B_{n-1}(0)} \leq f_n \leq 1$. Entonces

$$0 \leq \int f_m d\mu = \lim_{n\to\infty} \int f_m d\mu_n \leq 1,$$

de donde $0 \leq \mu(\mathbb{R}^d) \leq 1$, es decir que no podremos asegurar la convergencia a una medida de probabilidad como era en el caso de la convergencia débil, cuya definición es suficientemente similar a la de convergencia vaga como para ser confundidas. Los siguientes resultados están enfocados en garantizar que la convergencia vaga de probabilidades sea a una medida de probabilidad. El primero de estos resultados es conocido como el *teorema de selección de Helly*, para su prueba se requerirán un par de lemas preliminares.

Lema 6.36. *Si $\mu \in \mathcal{M}_l(\mathbb{R}^d)$ una medida finita no-negativa, existe una colección a lo más numerable de hiperplanos $\mathcal{H}_\mathcal{D}$ tales que la función $F(x) = \mu((-\infty, x])$ es continua para $x \notin \bigcup_{H\in\mathcal{H}_\mathcal{D}} H$.*

Demostración. Definamos $\mathcal{H}_i^y = \{x \in \mathbb{R}^d : x_i = y\}$; como μ es una medida finita

$$\mathcal{H}_\mathcal{D} = \bigcup_{i=1}^{d} \bigcup_{y\in\mathbb{R}} \{\mathcal{H}_i^y : \mu(\{\mathcal{H}_i^y\}) > 0\}$$

contiene a lo más una cantidad numerable de hiperplanos. Sea $\mathfrak{D}_\mu = \bigcup_{H\in\mathcal{H}_\mathcal{D}} H$, si $x \notin \mathfrak{D}_\mu$ entonces $\mu(\partial(-\infty, x]) = 0$, de donde

$$\lim_{y\uparrow x} F(y) = F(x) = \lim_{z\downarrow x} F(z).$$

Por tanto, para $\epsilon > 0$ existirán $x_1 < x < x_2$ tales que $0 \leq F(x) - F(x_1) \leq \epsilon$ y $0 \leq F(x_2) - F(x) \leq \epsilon$. Cuando $y \in [x_1, x_2]$ se tendrá que

$$F(x) - \epsilon \leq F(x_1) \leq F(y) \leq F(x_2) \leq F(x) + \epsilon$$

o equivalentemente $|F(x) - F(y)| \leq \epsilon$, probando el resultado. □

Lema 6.37. *Sean $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ y $\mu \in \mathcal{M}_l(\mathbb{R}^d)$ medidas de probabilidad. Si $\mathfrak{D}_\mu$ se define como en el lema pasado y $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ para todo $x \notin \mathfrak{D}_\mu$, $\mu_n \xrightarrow{d} \mu$.*

Demostración. Para $\epsilon > 0$ podemos encontrar $K > 0$ tal que $I = (-K, K]^d$ tenga sus vértices en $\mathfrak{D}_\mu^c$ y $\mu(I) > 1 - \epsilon$. Por hipótesis tendremos que existirá $N \in \mathbb{N}$ tal que para toda $n \geq N$ se cumpla que $\mu_n(I) > 1 - 2\epsilon$. Si $f \in C_b(\mathbb{R}^d)$ entonces será uniformemente continua en $[-K, K]^d$, por lo que podremos descomponer I en una cantidad finita de rectángulos $(u, v] \subset \mathbb{R}^d$ con vértices en $\mathfrak{D}_\mu^c$ de tal manera que $|f(x) - f(y)| < \epsilon$ cuando x y y pertenezcan al mismo rectángulo. Por lo tanto podremos aproximar f en I por una función g, constante en cada rectángulo $(u, v]$, tal que $|f(x) - g(x)| < \epsilon$ para $x \in I$. Por lo tanto tendremos

$$\begin{aligned}\left|\int f d\mu_n - \int f d\mu\right| &\leq \left|\int_I f d\mu_n - \int_I f d\mu\right| + \left|\int_{I^c} f d\mu_n - \int_{I^c} f d\mu\right| \\ &\leq \int_{I^c} |f|\, d\mu_n + \int_{I^c} |f|\, d\mu + \int_I |f - g|\, d\mu_n + \int_I |f - g|\, d\mu_n \\ &\quad + \left|\int_I g d\mu_n - \int_I g d\mu\right|,\end{aligned}$$

de donde $\int f d\mu_n \to \int f d\mu$, probando que $\mu_n \xrightarrow{d} \mu$. □

Teorema 6.38 (de selección de Helly). *Si $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ es una sucesión de medidas de probabilidad entonces existe una subsucesión $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ vagamente convergente.*

Demostración. Para cada $n \in \mathbb{N}$ sea $F_n : \mathbb{R}^d \to \mathbb{R}$ dada por $F_n(x) = \mu_n((-\infty, x])$. Sea $\{r_n\}_{n\in\mathbb{N}}$ una enumeración de $\mathbb{Q}^d$. Como $\{F_n(r_1)\}_{n\in\mathbb{N}} \subset [0, 1]$, existe una subsucesión $\{n_{k,1}\}_{k\in\mathbb{N}}$ de $\mathbb{N}$ tal que el límite de $\{F_{n_{k,1}}(r_1)\}_{k\in\mathbb{N}}$ existe. Denotemos a éste por $G(r_1)$. Asimismo, existe una subsucesión $\{n_{k,2}\}_{k\in\mathbb{N}}$ de $\{n_{k,1}\}_{k\in\mathbb{N}}$ tal que existe el límite de $\{F_{n_{k,2}}(r_2)\}_{k\in\mathbb{N}}$, denotado por $G(r_2)$. Procediendo de forma inductiva, para $j \in \mathbb{N}$, sea $\{n_{k,j}\}_{k\in\mathbb{N}}$ una subsucesión de $\{n_{k,j-1}\}_{k\in\mathbb{N}}$ tal que exista el límite de $\{F_{n_{k,j}}(r_j)\}$, denotado por $G(r_j)$. Para $k \in \mathbb{N}$, sea $n_k = n_{k,k}$; por construcción se tendrá que para $j \in \mathbb{N}$,

$$G(r_j) = \lim_{k\to\infty} F_{n_k}(r_j).$$

Luego, para $x \in \mathbb{R}^d$ definamos

$$F(x) = \inf\{G(r) : x < r, r \in \mathbb{Q}^d\}.$$

Dado que cada F_n es d-creciente, G también lo será y por lo tanto F también; más aún, G es monótona, *i.e.* $G(x) \leq G(y)$ cuando $x \leq y$ con $x, y \in \mathbb{Q}^d$, y así, por construcción F será continua por la derecha. Por el teorema 1.29 existirá una medida finita $\mu : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ tal que $\mu((a, b]) = \sum_{k=0}^{d} (-1)^k \sum_{c\in\Delta_{a,b}^k} F(c)$ para todo $a, b \in \mathbb{R}^d$. Ahora probaremos que $\mu_{n_k} \xrightarrow{v} \mu$.

Notemos primero que si $\mu(\partial(-\infty, x]) = 0$, entonces x es un punto de continuidad de F. Para $\epsilon > 0$ existirán $q, r \in \mathbb{Q}^d$ tales que $q < x < r$ y

$$F(x) - \epsilon \leq G(q) \leq F(x) \leq G(r) \leq F(x) + \epsilon.$$

Más aún, existe $N \in \mathbb{N}$ tal que para toda $n_k \geq N$,

$$F(x) - 2\epsilon \leq F_{n_k}(q) \leq F_{n_k}(x) \leq F_{n_k}(r) \leq F(x) + 2\epsilon.$$

De esta última cadena de desigualdades obtendremos que

$$F(x) - 2\epsilon \leq \liminf_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(x) \leq F(x) + 2\epsilon,$$

es decir, $\lim_{k\to\infty} F_{n_k}(x) = F(x)$. Esto en particular implica que $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ para toda $x \notin \mathfrak{D}_\mu^c$, donde $\mathfrak{D}_\mu$ se define como en los lemas anteriores. Más aún, si $(u, v] \subset \mathbb{R}^d$ es un rectángulo con vértices en $\mathfrak{D}_\mu^c$, $\mu_{n_k}((u, v]) \to \mu((u, v])$.

Así las cosas, para $f \in C_c(\mathbb{R}^d)$, por tener soporte compacto, podemos encontrar un rectángulo $(u,v] \subset \mathbb{R}^d$ con vértices en $\mathfrak{D}_\mu^c$ tal que $\operatorname{supp} f \subset (u,v]$. Si $\mu((u,v]) = 0$, es claro que $\int f d\mu_{n_k} \to \int f d\mu$, por lo que supondremos que $\mu((u,v]) > 0$. En este caso definamos $\nu(A) = \mu(A \cap (u,v])/\mu((u,v])$ y $\nu_n(A) = \mu_n(A \cap (u,v])/\mu_n((u,v])$ para cada $n \in \mathbb{N}$. Es claro que $\nu_{n_k}((-\infty,x]) \to \nu((-\infty,x])$ para toda $x \notin \mathfrak{D}_\nu \subset \mathfrak{D}_\mu$ y, por el lema anterior, se sigue que $\nu_{n_k} \xrightarrow{d} \nu$. Así

$$\lim_{k\to\infty} \int f d\nu_{n_k} = \int f d\nu.$$

Por lo tanto tendremos

$$\begin{aligned}
\left|\int f d\mu_{n_k} - \int f d\mu\right| &= \left|\int_{(u,v]} f d\mu_{n_k} - \int_{(u,v]} f d\mu\right| \\
&= \left|\mu_{n_k}((u,v]) \int f d\nu_{n_k} - \mu((u,v]) \int f d\nu\right| \\
&\leq \mu_{n_k}((u,v]) \left|\int f d\nu_{n_k} - \int f d\nu\right| + \left|\int f d\nu\right| \left|\mu_{n_k}((u,v]) - \mu((u,v])\right| \\
&\leq \left|\int f d\nu_{n_k} - \int f d\nu\right| + \|f\|_u \left|\mu_{n_k}((u,v]) - \mu((u,v])\right|,
\end{aligned}$$

de donde finalamente concluimos que

$$\lim_{k\to\infty} \int f d\mu_{n_k} = \int f d\mu,$$

de donde $\mu_{n_k} \xrightarrow{v} \mu$. $\square$

Lo que se ha probado en este teorema es que el espacio de medidas de probabilidad sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ es relativamente compacto con respecto a la topología vaga en $\mathcal{M}_l(\mathbb{R}^d)$. Sin embargo, como ya se hizo mención, pudiese ser el caso que si $\mu_n \xrightarrow{v} \mu$, donde $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ son medidas de probabilidad, entonces $\mu(\mathbb{R}^d) < 1$. Para que éste no sea el caso, pediremos que las sucesiones de medidas de probabilidad que consideremos sean tensas.

Lema 6.39. *Sean $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ una sucesión de medidas de probabilidad. Si $\mu_n \xrightarrow{v} \mu$ entonces $\mu_n((u,v]) \to \mu((u,v])$ para todo $(u,v] \subset \mathbb{R}^d$ con vértices en $\mathfrak{D}_\mu^c$.*

Demostración. Sea $(u,v] \subset \mathbb{R}^d$ un rectángulo con vértices en $\mathfrak{D}_\mu^c$. Para cada $k \in \mathbb{N}$ sean $[x_k, y_k] \subset (u,v] \subset (w_k, z_k)$ con vértices en $\mathfrak{D}_\mu^c$ tales que $[x_k,y_k] \subset [x_{k+1}, y_{k+1}]$, $\bigcup_{k\in\mathbb{N}} [x_k,y_k] = (u,v)$, $(w_k,z_k) \supset (w_{k+1}, z_{k+1})$, $\bigcap_{k\in\mathbb{N}} (w_k,z_k) = [u,v]$. Por el lema de Urysohn existen $\{h_k\}_{k\in\mathbb{N}}, \{H_k\}_{k\in\mathbb{N}} \subset C(\mathbb{R}^d, [0,1])$ tales que $h_k([x_k,y_k]) = 1$, $h_k((u,v)^c) = 0$, $H_k([u,v]) = 1$ y $H_k((w_k,z_k)^c) = 0$. Notemos que $h_k \leq \mathbb{1}_{(u,v]} \leq H_k$ para cada $k \in \mathbb{N}$ y que $\{h_k\}_{k\in\mathbb{N}}, \{H_k\}_{k\in\mathbb{N}} \subset C_K(\mathbb{R}^d)$, por lo que por monotonía de la integray y por hipótesis,

$$\int h_k d\mu \leq \liminf_{n\to\infty} \mu_n((u,v]) \leq \limsup_{n\to\infty} \mu_n((u,v]) \leq \int H_k d\mu.$$

Así, por convergencia monótona,

$$\mu((u,v)) \leq \liminf_{n\to\infty} \mu_n((u,v]) \leq \limsup_{n\to\infty} \mu_n((u,v]) \leq \mu([u,v]).$$

Basta observar que $\mu((u,v)) = \mu([u,v]) = \mu((u,v])$ por hipótesis para concluir. $\square$

Proposición 6.40. *Sea* $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ *una sucesión de medidas de probabilidad tal que* $\mu_n \xrightarrow{v} \mu$. $\mu(\mathbb{R}^d) = 1$ *si y sólo si* $\{\mu_n\}_{n\in\mathbb{N}}$ *es tensa, en cuyo caso* $\mu_n \xrightarrow{d} \mu$.

Demostración. Primero supongamos que $\mu(\mathbb{R}^d) = 1$. Por el lema anterior y el lema 6.37 tendremos que $\mu_n \xrightarrow{d} \mu$ y por la proposición 6.26, $\{\mu_n\}_{n\in\mathbb{N}}$ será tensa. Recíprocamente, para $\epsilon > 0$ existe $M > 0$ tal que $\sup_{n\in\mathbb{N}} \mu_n(\{x \in \mathbb{R}^d : |x| > M\}) \leq \epsilon$; entonces existirá $K > 0$ tal que $\mu_n((-K, K]^d) \geq 1 - \epsilon$ para toda $n \in \mathbb{N}$, donde $(K, K]^d$ es tal que sus vértices están en $\mathfrak{D}_\mu^c$. Por el lema anterior tendremos entonces que $\mu((-K, K]^d) \geq 1 - 2\epsilon$, de donde $\mu(\mathbb{R}^d) = 1$, concluyendo la demostración. □

Proposición 6.41. *Sea* $\{\mu\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ *una sucesión de medidas de probabilidad. Esta sucesión es tensa si y sólo si toda subsucesión tiene una subsubsucesión que convege débilmente.*

Demostración. Por el teorema de selección de Helly, para toda subsucesión $\{\mu_{n_k}\}_{k\in\mathbb{N}} \subset \{\mu_n\}_{n\in\mathbb{N}}$ existe una subsubsucesión $\{\mu_{n_{k_r}}\}_{r\in\mathbb{N}} \subset \{\mu_{n_k}\}_{k\in\mathbb{N}}$ vagamente convergente. Si $\{\mu_n\}_{n\in\mathbb{N}}$ es tensa, esto implica que $\{\mu_{n_{k_r}}\}_{r\in\mathbb{N}}$ es débilmente convergente.

Inversamente supongamos que para toda subsucesión de $\{\mu_n\}_{n\in\mathbb{N}}$ existe una subsubsucesión débilmente convergente y que $\{\mu_n\}_{n\in\mathbb{N}}$ no es tensa. Por el lema 6.25, existen $\epsilon > 0$ y una subsucesión $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ tal que $\mu_{n_k}(\{x \in \mathbb{R}^d : |x| > k\}) > \epsilon$. Por hipótesis, existe $\{\mu_{n_{k_r}}\}_{r\in\mathbb{N}} \subset \{\mu_{n_k}\}_{k\in\mathbb{N}}$ débilmente convergente y por lo tanto tensa, por lo que para $\epsilon > 0$ existe $K > 0$ tal que $\mu_{n_{k_r}}(\{x \in \mathbb{R}^d : |x| > K\}) \leq \epsilon$ para toda $r \in \mathbb{N}$, lo cual es una contradicción. Por lo tanto $\{\mu_n\}_{n\in\mathbb{N}}$ es tensa. □

Finalmente probamos la extensión del teorema de continuidad de Lévy.

Teorema 6.42. *Sea* $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_l(\mathbb{R}^d)$ *una sucesión de medidas de probabilidad con* $\hat{\mu}(t) \to \varphi(t)$ *para toda* $t \in \mathbb{R}$, *donde* φ *es una función continua en cero. Entonces* $\mu_n \xrightarrow{d} \mu$ *para alguna medida de probabilidad con* $\hat{\mu} = \varphi$.

Demostración. De manera similar a la prueba del teorema 6.29, $\{\mu_n\}_{n\in\mathbb{N}}$ es tensa. Por el teorema de selección de Helly y la proposición 6.40 existe una subsucesión $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ débilmente convergente, digamos, a una medida de probabilidad μ. Por el teorema de continuidad de Lévy, $\hat{\mu}_{n_k}(t) \to \hat{\mu}(t) = \varphi(t)$ y, nuevamente por el teorema de continuidad de Lévy, $\mu_n \xrightarrow{d} \mu$. □

Apéndices

Apéndice A. **Reales Extendidos**

En Teoría de la Medida será inevitable trabajar con el infinito. En muchos casos la medida de conjuntos tomará el valor ∞. También será necesario considerar que series y sucesiones pueden converger a ∞ ó $-\infty$. Principalmente por estos motivos introducimos a los **reales extendidos**, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. Extendemos el orden total de $\mathbb{R}$ a $\overline{\mathbb{R}}$ declarando que para todo $x \in \mathbb{R}$, $-\infty < x < \infty$. Una consecuencia inmediata de esto es que todo subconjunto A de $\mathbb{R}$ tiene un supremo y un ínfimo.

En particular, si $\{x_n\}_{n\in\mathbb{N}} \subset \overline{\mathbb{R}}$ es una sucesión monótono-creciente, se verifica fácilmente que el límite existe y $\lim_{n\to\infty} x_n = \sup_{n\geq 1} x_n$. Si la sucesión es monótona decreciente, $\lim_{n\to\infty} x_n = \inf_{n\geq 1} x_n$. Como consecuencia, para cualquier sucesión podemos definir el **límite inferior** y **límite superior**:

$$\liminf_n x_n := \lim_{n\to\infty}\Big(\inf_{k\geq n} x_k\Big), \qquad \limsup_n x_n := \lim_{n\to\infty}\Big(\sup_{k\geq n} x_k\Big).$$

Nótese que si $b_n = \inf_{k\geq n} x_k$ y $c_n = \sup_{k\geq n} x_k$, entonces $\{b_n\}_{n\in\mathbb{N}}$ es monótono-creciente y $\{c_n\}_{n\in\mathbb{N}}$ es monótono-decreciente. Además dado que $b_n \leq c_n$ para todo $n \in \mathbb{N}$, se sigue que $\liminf_n x_n \leq \limsup_n x_n$.

Proposición A.1. *Sea* $\{x_n\}_{n\in\mathbb{N}} \subset \overline{\mathbb{R}}$*. Entonces,* $\liminf_n x_n \leq \limsup_n x_n$ *y la igualdad se cumple si y sólo si el límite de* $\{x_n\}_{n\in\mathbb{N}}$ *existe y en este caso,*

$$\lim_{n\to\infty} x_n = \liminf_n x_n = \limsup_n x_n.$$

Demostración. Tomemos $b_n = \inf_{k\geq n} x_k$ y $c_n = \sup_{k\geq n} x_k$. Está claro que $b_n \leq x_n \leq c_n$; así que si $\liminf_n x_n = \limsup_n x_n$, podemos concluir la primera implicación. Ahora supongamos que $\lim_{n\to\infty} x_n = \alpha$. Si $\alpha = \infty$, para todo $M > 0$ existe $N \in \mathbb{N}$ tal que para todo $n \geq N$, $x_n > M$. Se sigue que $b_n, c_n \geq M$ para todo $n \geq M$. Por lo tanto

$$\liminf_n x_n = \limsup_n x_n = \infty.$$

El caso $\alpha = -\infty$ es análogo. Si $\alpha \in \mathbb{R}$, dado $\epsilon > 0$ sea N tal que $x_n \in (\alpha - \epsilon, \alpha + \epsilon)$ para todo $n \geq N$. Esto implica que $b_n \geq \alpha - \epsilon$ y $c_n \leq \alpha + \epsilon$ para todo $n \geq N$. En consecuencia

$$\alpha - \epsilon \leq \liminf_n x_n \leq \limsup_n x_n \leq \alpha + \epsilon.$$

Como ϵ es arbitrario, tenemos $\liminf_n x_n = \limsup_n x_n = \alpha$. $\square$

También extenderemos parcialmente las operaciones aritméticas a $\overline{\mathbb{R}}$:

1) $x \pm \infty = \pm\infty$ para todo $x \in \mathbb{R}$,

2) $\infty + \infty = \infty$ y $-\infty - \infty = -\infty$,

3) $x \cdot (\pm\infty) = \pm\infty$ si $x > 0$ y $x \cdot (\pm\infty) = \mp\infty$ si $x < 0$,

4) $0 \cdot (\pm\infty) = 0$.

Si $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^+ \cup \{\infty\}$ entonces $\sum_{n=1}^{\infty} x_n$ está definido. Más aún:

Proposición A.2. *Sea $x_{n,m} \geq 0$ para todo $n, m \in \mathbb{N}$ y $k \mapsto (n(k), m(k))$ cualquier biyección de $\mathbb{N}$ a $\mathbb{N} \times \mathbb{N}$. Entonces,*

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{n,m} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{n,m} = \sum_{k=1}^{\infty} x_{n(k),m(k)} = \sup\Big\{ \sum_{(n,m)\in F} x_{n,m} : F \subset \mathbb{N}\times\mathbb{N},\ F \textit{ es finito}\Big\}.$$

Demostración. Sea $S = \sup\Big\{ \sum_{(n,m)\in F} x_{n,m} : F \subset \mathbb{N}\times\mathbb{N},\ F \textit{ es finito}\Big\}$. Para $N, M, K \in \mathbb{N}$,

$$\sum_{n=1}^{N}\sum_{m=1}^{M} x_{n,m} = \sum_{m=1}^{M}\sum_{n=1}^{N} x_{n,m} \leq S, \quad \sum_{k=1}^{K} x_{n(k),m(k)} \leq S.$$

Las desigualdades anteriores se pueden extender a $N = M = K = \infty$ (tomando primero el límite con respecto a la suma de adentro y luego con la externa). Por otro lado, para todo $F \subset \mathbb{N}\times\mathbb{N}$ finito existen N y K lo suficientemente grandes para que

$$\begin{aligned}\sum_{(n,m)\in F} x_{n,m} &\leq \min\Big\{ \sum_{n=1}^{N}\sum_{m=1}^{N} x_{n,m},\ \sum_{k=1}^{K} x_{n(k),m(k)}\Big\} \\ &\leq \min\Big\{ \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{n,m},\ \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{n,m},\ \sum_{k=1}^{\infty} x_{n(k),m(k)}\Big\}.\end{aligned}$$

Por lo tanto $S \leq \min\Big\{ \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{n,m},\ \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{n,m},\ \sum_{k=1}^{\infty} x_{n(k),m(k)}\Big\}$ y podemos concluir. □

Apéndice B. **Espacios Métricos**

Definición B.1. Dado un espacio X, $d : X \times X \mapsto \mathbb{R}^+$ es una **métrica** si para todo $x, y, z \in X$,

1) $d(x, y) > 0$ si $x \neq y$ y $d(x, x) = 0$,

2) $d(x, y) = d(y, x)$,

3) $d(x, y) \leq d(x, z) + d(z, y)$.

Decimos que la dupla (X, d) es un **espacio métrico**. El ejemplo más común de espacio métrico es $\mathbb{R}^n$ con la métrica euclidiana

$$d(x, y) = |x - y| = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2}.$$

Si (X, d) es un espacio métrico, para $x \in X$ definimos a la *bola abierta* de radio $r > 0$ al rededor de x como

$$B_r(x) := \{y \in X : d(x, y) < r\}.$$

Definición B.2.

1) Decimos que $G \subset X$ es **abierto** si para todo $x \in G$ existe $r > 0$ tal que $B_r(x) \subset G$.

2) Dado $E \subset X$ decimos que x es **punto de acumulación** de E si para todo $r > 0$, $\big(B_r(x) - \{x\}\big) \cap E \neq \emptyset$.

3) Decimos que F es **cerrado** si contiene a todos sus puntos de acumulación.

4) Decimos que E es **denso** en X si todo punto en $X - E$ es un punto de acumulación de E.

Proposición B.3. *Para todo $x \in X$ y $r > 0$, $B_r(x)$ es abierto.*

Demostración. Dado $y \in B_r(x)$ sea $h = d(x, y) < r$. Veamos que $B_{r-h}(y) \subset B_r(x)$:
Para todo $z \in B_{r-h}(y)$ tenemos

$$d(x, z) \leq d(x, y) + d(y, z) < h + (r - h) = r.$$

Por lo tanto $z \in B_r(x)$ y podemos concluir. □

Proposición B.4. *Un conjunto $G \subset X$ es abierto si y sólo si G^c es cerrado.*

Demostración. Supongamos que G^c es cerrado. Si $x \in G$ entonces éste no puede ser punto de acumulación de G^c. Se sigue que existe $r > 0$ tal que $\big(B_r(x) - \{x\}\big) \cap G^c = \emptyset$; es decir, $B_r(x) \subset G$. Por lo tanto G es abierto.
Ahora supongamos que G es abierto. Si x es punto de acumulación de G^c, entonces para todo $r > 0$, $\big(B_r(x) - \{x\}\big) \cap G^c \neq \emptyset$. Si $x \in G$ lo anterior se contradice así que $x \in G^c$. Por lo tanto G^c es cerrado. □

Corolario B.5. *$F \subset X$ es cerrado si y sólo si F^c es abierto.*

Proposición B.6.

1) Para toda colección de abiertos, $\{G_\alpha\}_{\alpha\in\Lambda}$, $\bigcup_{\alpha\in\Lambda} G_\alpha$ es abierto.

2) Para toda colección de cerrados, $\{F_\alpha\}_{\alpha\in\Lambda}$, $\bigcap_{\alpha\in\Lambda} F_\alpha$ es cerrado.

3) Para toda colección finita de abiertos, $G_1, \ldots, G_n$, $\bigcap_{i=1}^n G_i$ es abierto.

4) Para toda colección finita de cerrados, $F_1, \ldots, F_n$, $\bigcup_{i=1}^n F_i$ es cerrado.

Demostración. Probemos 1). Si $x \in \bigcup_{\alpha\in\Lambda} G_\alpha$ entonces existe α_0 tal que $x \in G_{\alpha_0}$. Como G_{α_0} es abierto existe $r > 0$ tal que $B_r(x) \subset G_{\alpha_0}$. En consecuencia $B_r(x) \subset \bigcup_{\alpha\in\Lambda} G_\alpha$. Podemos concluir que $\bigcup_{\alpha\in\Lambda} G_\alpha$ es abierto.
Para probar 2) notamos que $\big(\bigcap_{\alpha\in\Lambda} F_\alpha\big)^c = \bigcup_{\alpha\in\Lambda} F_\alpha^c$. Como cada F_α es cerrado sus complementos son abiertos y de 1) se sigue que $\big(\bigcap_{\alpha\in\Lambda} F_\alpha\big)^c$ es abierto. Por lo tanto $\bigcap_{\alpha\in\Lambda} F_\alpha$ es cerrado.
Ahora probaremos 3). Dado $x \in \bigcap_{i=1}^n G_i$, existe $r_i > 0$ tal que $B_{r_i}(x) \subset G_i$, $1 \leq i \leq n$. Con $r = \min\{r_1, \ldots, r_n\}$ está claro que $B_r(x) \subset \bigcap_{i=1}^n G_i$ y se sigue el resultado.
La demostración de 4) es análoga a la de 2). □

Retomemos el espacio métrico $\mathbb{R}^n$ (con la métrica euclidiana). Para $a = (a_1, \ldots, a_n)$ y $b = (b_1, \ldots, b_n)$ en $\mathbb{R}^n$ decimos que $a < b$ ($\leq$) si y sólo si $a_i < b_i$ ($\leq$), $1 \leq i \leq n$. Si $a < b$, definimos

$$(a, b) := \{x \in \mathbb{R}^n : a < x < b\} = (a_1, b_1) \times \cdots \times (a_n, b_n).$$

De manera análoga definimos a los conjuntos $(a, b]$, $[a, b)$ y $[a, b]$. Por otro lado

$$(a, \infty) := (a_1, \infty) \times \cdots \times (a_n, \infty)$$

y definimos de manera análoga a $(-\infty, a)$, $(-\infty, a]$ y $[a, \infty)$.

Proposición B.7. *Para $a < b$ en $\mathbb{R}^n$, (a, b), $(-\infty, a)$ y (a, ∞) son abiertos.*

Demostración. Sea $x \in (a, b)$ y $r = \text{mín}\{x_1 - a_1, b_1 - x_1, \ldots, x_n - a_n, b_n - x_n\} > 0$. Tomemos $y \in B_r(x)$; dado que $|x_i - y_i| \leq d(x, y)$ tenemos

$$a_i = x_i - (x_i - a_i) \leq x_i - r = y_i + (x_i - y_i) - r \leq y_i + d(x, y) - r < y_i$$

y

$$y_i < y_i - d(x, y) + r \leq y_i + (x_i - y_i) + r = x_i + r \leq x_i + (b_i - x_i) = b_i.$$

Por lo tanto $y \in (a, b)$ y se sigue que $B_r(x) \subset (a, b)$. Por lo tanto (a, b) es abierto.
La prueba de que $(-\infty, a)$ y (a, ∞) son abiertos es análoga. □

Definición B.8. X es **separable** si y sólo si existe un subconjunto denso y numerable.

Definición B.9. Sea $B = \{V_\alpha\}_{\alpha \in A}$ una familia de abiertos en X. Decimos que B es una **base** si para todo abierto G existe $C \subset A$ tal que $G = \bigcup_{\alpha \in C} V_\alpha$.

Nótese que B es base si y sólo si para todo abierto $G \subset X$ y $x \in G$ existe $V_x \in B$ tal que $x \in V_x \subset G$.

Proposición B.10. *Si X es separable entonces tiene una base numerable.*

Demostración. Sea D un subconjunto denso numerable de X y

$$B = \{B_r(x) : x \in D,\, r \in \mathbb{Q}^+\}.$$

Claramente B es numerable. Para ver que B es base basta tomar cualquier abierto G y probar que para todo $y \in G$ existe $x \in G \cap D$ y $r \in \mathbb{Q}^+$ tal que $y \in B_r(x) \subset G$. Dado que G es abierto, para todo $y \in G$ existe $r \in \mathbb{Q}^+$ tal que $B_r(y) \subset G$. Ya sea que $y \in D$ o no, tenemos que existe $x \in B_{r/2}(y) \cap D$. Luego, tenemos que $y \in B_{r/2}(x)$ y para todo $z \in B_{r/2}(x)$,

$$d(y, z) \leq d(y, x) + d(x, z) < r/2 + r/2 = r.$$

Por lo tanto $y \in B_{r/2}(x) \subset B_r(y) \subset G$. □

Sean $(X_1, d_1), \ldots, (X_n, d_n)$ espacios métricos. Tenemos una métrica en $X = \prod_{i=1}^n X_i$ dada por

$$d(x, y) := \left(\sum_{i=1}^n d_i(x_i, y_i)^2 \right)^{1/2}.$$

Es fácil ver que d en efecto es una métrica y es llamada la **métrica producto**. Es importante notar que para $1 \leq j \leq n$ y $x, y \in X$,

$$d_j(x_j, y_j) \leq d(x, y) \leq \sum_{i=1}^{n} d_i(x_i, y_i).$$

Proposición B.11. *Sea $G_i \subset X_i$ abiertos para $1 \leq i \leq n$. Si consideramos a X como espacio métrico equipado con la métrica producto, entonces $\prod_{i=1}^{n} G_i$ es abierto en X.*

Demostración. Tómese $x = (x_1, \ldots, x_n) \in \prod_{i=1}^{n} G_i$. Para cada x_i existe $r_i > 0$ tal que $B_{r_i}(x_i) \subset G_i$. Sea $r = \min\{r_1, \ldots, r_n\}$, entonces para todo $y \in B_r(x)$ y para todo $i \in \{1, \ldots, n\}$ tenemos

$$d_i(x_i, y_i) \leq d(x, y) < r \leq r_i.$$

Se sigue que $B_r(x) \subset \prod_{i=1}^{n} B_{r_i}(x_i) \subset \prod_{i=1}^{n} G_i$. □

Proposición B.12. *Si B_i es base de X_i para $1 \leq i \leq n$, entonces,*

$$B = \left\{ \prod_{i=1}^{n} G_i : G_i \in B_i, 1 \leq i \leq n \right\},$$

es una base para X equipado con la métrica producto.

Demostración. Sea $G \subset X$ abierto. Para ver que B es base, basta probar que para todo $x \in G$ existe $\prod_{i=1}^{n} G_i \in B$ tal que $x \in \prod_{i=1}^{n} G_i \subset G$. Sea $r > 0$ tal que $B_r(x) \subset G$. Podemos ver fácilmente que $\prod_{i=1}^{n} B_{r/n}(x_i) \subset B_r(x)$:
Si $y_i \in B_{r/n}(x_i)$ para $1 \leq i \leq n$, entonces para $y = (y_1, \ldots, y_n)$ tenemos

$$d(x, y) \leq \sum_{i=1}^{n} d_i(x_i, y_i) < \sum_{i=1}^{n} r/n = r.$$

Ahora, como cada B_i es base, existe $G_i \in B_i$ tal que $x_i \in G_i \subset B_{r/n}(x_i)$. Por lo tanto

$$x \in \prod_{i=1}^{n} G_i \subset \prod_{i=1}^{n} B_{r/n}(x_i) \subset B_r(x) \subset G.$$ □

Uno de los conceptos íntimamente ligados con los espacios métricos es el de *espacio completo*.

Definición B.13.

1) Una sucesión $(x_n)_{n \in \mathbb{N}}$ de elementos de un espacio métrico X **converge** a $x \in X$ si para cualquier $\epsilon > 0$ existe $N \in \mathbb{N}$ tal que $d(x_n, x) < \epsilon$ para todo $n \geq N$.

2) Una sucesión $(x_n)_{n \in \mathbb{N}}$ de elementos de un espacio métrico X es de **Cauchy** si para cualquier $\epsilon > 0$ existe $N \in \mathbb{N}$ tal que $d(x_n, x_m) < \epsilon$ para cualesquiera $n, m \geq N$.

3) Diremos que X es **completo** si cualquier sucesión de Cauchy es convergente.

Un ejemplo de espacio completo es el de los reales con la métrica euclidiana. En la práctica resulta conveniente trabajar con espacios que sean completos, pues nos asegura que sucesiones de elementos que 'estén cada vez más cerca' terminen afuera del espacio en el que estemos trabajando. Por fortuna siempre podremos extender un espacio métrico a un espacio métrico completo, al que llamaremos la *compleción* de X bajo d.

Proposición B.14. *Sea (X, d) un espacio métrico. Existe un espacio métrico (Y, D) que es completo y extiende a (X, d).*

Demostración. Sea $\tilde{X}$ el conjunto de todas las sucesiones de Cauchy en X. Para $x, y \in \tilde{X}$ diremos que $x \sim y$ si y sólo si $\lim_n d(x_n, y_n) = 0$. Es claro que $\sim$ es una relación de equivalencia.

Consideremos entonces Y como el conjunto de clases de equivalencia en $\tilde{X}$ bajo $\sim$. Por otra parte definamos $D([x], [y]) = \lim_n d(x_n, y_n)$ para $[x], [y] \in Y$. Primero veremos que esta función está bien definida y que no depende del representante de clase que consideremos. Es decir que el límite existe y que si $x \sim x'$ y $y \sim y'$ entonces $D([x], [y]) = D([x'], [y'])$. Es claro que

$$\begin{aligned} |d(x_n, y_n) - d(x_m, y_m)| &\leq |d(x_n, y_n) - d(x_n, y_m)| + |d(x_n, y_m) - d(x_m, y_m)| \\ &\leq d(y_m, y_n) + d(x_m, x_n), \end{aligned}$$

por lo que $(d(x_n, y_n))_{n\in\mathbb{N}}$ es una sucesión de Cauchy en los reales y por lo tanto convergente. Ahora notemos que

$$D([x], [y]) = \lim_{n\to\infty} d(x_n, y_n) \leq \lim_{n\to\infty} (d(x_n, x'_n) + d(x'_n, y'_n) + d(y'_n, y_n)) = D([x'], [y']),$$

donde la última igualdad se da por la definición de D. Análogamente se prueba que $D([x'], [y']) \leq D([x], [y])$, lo que prueba que D está bien definida. Se deja como ejercicio comprobar que D es una métrica en Y. Ahora se probará que (Y, D) es un espacio completo. Es decir que queremos probar que cualquier sucesión de Cauchy en Y es convergente.

Con este fin consideremos $h : X \to Y$ el mapeo $z \mapsto [(z, z, \ldots)]$ el cual es claramente inyectivo y además cumple que $d(z_1, z_2) = D(h(z_1), h(z_2))$. Por lo tanto h es una isometría entre X y $h(X) \subset Y$. Hacemos notar que $(h(z_n))_{n\in\mathbb{N}}$ es una sucesión de Cauchy en Y si y sólo si $(z_n)_{n\in\mathbb{N}}$ es una sucesión de Cauchy en X, por lo que esta última sucesión pertenecerá a una clase de equivalencia $[z^*] \in Y$. Notemos que en este caso para cualquier $\epsilon > 0$ existirá $N \in \mathbb{N}$ tal que $d(z_n, z_m) < \epsilon$ para toda $n, m \geq N$, de donde $D(h(z_n), [z^*]) < \varepsilon$ para toda $n \geq N$, lo que implica que $(h(z_n))_{n\in\mathbb{N}}$ converge a $[z^*]$.

Por lo tanto, si demostramos que $h(X)$ es denso en Y podremos concluir pues si fuese cierto y $([x_n])_{n\in\mathbb{N}}$ fuese de Cauchy en Y, existirá $(z_n)_{n\in\mathbb{N}}$ tal que $D([x_n], h(z_n)) < n^{-1}$, en cuyo caso $(h(z_n))_{n\in\mathbb{N}}$ será de Cauchy y como vimos, convergerá a $[z^*] \in Y$. Notando que

$$D([x_n], [z^*]) \leq D([x_n], h(z_n)) + D(h(z_n), [z^*])$$

se sigue que $([x_n])_{n\in\mathbb{N}}$ también converge a $[z^*] \in Y$.

Para mostrar que $h(X)$ es denso en Y consideremos un elemento arbitrario $[x] \in Y$ con representante $(x_n)_{n\in\mathbb{N}}$ y consideremos la sucesión $(h(x_n))_{n\in\mathbb{N}}$ en $h(X)$. Por hipótesis para cada $\epsilon > 0$ existe $M \in \mathbb{N}$ tal que $d(x_n, x_m) < \epsilon$ para $n, m \geq M$ o bien que $\lim_m d(x_n, x_m) < \epsilon$ para $n \geq M$. Esto es lo mismo que $D(h(x_n), [x]) < \epsilon$ para $n \geq M$. □

Apéndice C. **Espacios metrizables**

En teoría de la medida, más cuando nos restringimos al terreno de la probabilidad, comenzaremos con un espacio arbitrario X y un sistema de conjuntos abiertos sin que éste sea generado por una métrica. A diferencia de los espacios métricos, al comenzar con un conjunto de conjuntos abiertos, necesitaremos pedir que satisfagan algunas condiciones para que estos estén bien definidos.

Definición C.1. Sea X un conjunto no vacío. Diremos que $\tau \subset \mathcal{P}(X)$ es una **topología** si se cumplen:

1) $\emptyset, X \in \tau$,

2) si $\{U_\alpha\}_{\alpha \in A} \subset \tau$ entonces $\bigcup_{\alpha \in A} U_\alpha \in \tau$,

3) si $\{U_i\}_{1 \leq i \leq n} \subset \tau$ para $n \in \mathbb{N}$ entonces $\bigcap_{i=1}^{n} U_i \in \tau$.

Al par ordenado (X, τ) se le da el nombre de **espacio topológico** y a los conjuntos $U \subset X$ tales que $U \in \tau$ se les conoce como *conjuntos abiertos*. En este caso la definición de *base* sigue teniendo sentido y se hace notar que una topología puede tener más de una base, en cuyo caso diremos que las bases son *equivalentes*. Notemos que en el caso particular de los espacios métricos, las bases que considerábamos para generar la colección de conjuntos abiertos consistía en las bolas abiertas generadas por una métrica. Esto da pie a la siguiente definición.

Definición C.2. Diremos que el espacio topológico (X, τ) es **metrizable** si existe una métrica $d : X \times X \to \mathbb{R}^+$ tal que

$$B = \{B_r(x) : x \in X, r \in \mathbb{R}^+\}$$

es una base para τ. En este caso diremos que d *metriza* a τ.

Si es el caso que d_1 y d_2 metricen a τ entonces diremos que las métricas d_1 y d_2 son *equivalentes*.

En general decir si un espacio topológico es metrizable o no, no resulta ser sencillo. Resulta ser el caso que si sabemos de otro espacio que sí sea metrizable y exista un homeomorfismo entre ambos, entonces nuestro espacio de interés será metrizable.

Proposición C.3. *Un espacio topológico (X, τ) es metrizable si y sólo si existe un espacio métrico (Y, d) homeomorfo a (X, τ).*

Demostración. Si (X, τ) es metrizable entonces existe d_X que metriza a tau por lo que basta considerar $(Y, d) = (X, d_X)$. Para la suficiencia sea $f : X \to Y$ un homeomorfismo, es decir una función biyectiva que satisfaga que para cada $U \subset X$ abierto existe $V \subset Y$ abierto tal que $f(U) = V$ o equivalentemente $U = f^{-1}(V)$. Es fácil probar que $d_f : X \times X \to \mathbb{R}^+$ dada por $(x, y) \mapsto d(f(x), f(y))$ es una métrica y que $f^{-1}\big(B_r^Y(f(x_0))\big) = B_r^X(x_0)$, por lo que d_f metriza a τ. $\square$

Para ejemplificar la última proposición, veremos que $\overline{\mathbb{R}}$ es un espacio metrizable cuando consideramos como topología cuya base está dada por elementos de la forma

$$[-\infty, a), \quad (a, b), \quad (b, \infty].$$

Para esto notemos que $f : \mathbb{R} \to (-1,1)$ dado por $x \mapsto 2\arctan(x)/\pi$ es un homeomorfismo cuando en ambos casos consideramos la topología usual, es decir la generada por la métrica $d(x,y) = |x-y|$. Notemos que podemos definir una topología para $[-1,1]$ similar a la que definimos para $\overline{\mathbb{R}}$ por aquella cuya base está compuesta por conjuntos de la forma $[-1,a)$, (a,b) y $(b,1]$, la cual coincide con la topología metrizable que es heredada por $\mathbb{R}$ a $[-1,1]$, por lo que el homeomorfismo f entre $\mathbb{R}$ y $(-1,1)$ lo podemos extender a $\overline{\mathbb{R}}$ y $[-1,1]$. En este caso, es fácil ver que la topología en $[-1,1]$ está metrizada por $d(x,y) = |x-y|$, cuya base tendrá conjuntos de la forma $[-1,b)$, (a,b), $(a,1]$ y $[-1,1]$ para $a,b \in (-1,1)$. Por la proposición anterior la topología que usamos para $\overline{\mathbb{R}}$ estará metrizada por

$$\tilde{d}(x,y) = 2|\arctan(x) - \arctan(y)|/2$$

y las bolas abiertas en $\overline{\mathbb{R}}$ serán de la forma $\overline{\mathbb{R}}$, (a,b), $[-\infty,b)$ y $(a,\infty]$ para $a,b \in \mathbb{R}$.

Así, si denotamos por $\mathcal{C} = \{(a,b) : a < b, a,b \in \mathbb{R}\} \cup \{[-\infty,a) : a \in \mathbb{R}\} \cup \{(b,\infty] : b \in \mathbb{R}\}$, claramente tendremos que $\sigma(\mathcal{C}) \subset \mathcal{B}(\overline{\mathbb{R}})$. Por otra parte, $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{E})$ donde $\mathcal{E} = \{(a,\infty] : a \in \mathbb{R}\}$, de donde $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{C})$.

Apéndice D. **Stone–Weierstrass**

En este apéndice se presenta la prueba de uno de los teoremas más importantes de aproximación de funciones: el teorema de Stone–Weierstrass.

Diremos que un espacio topológico (X,τ) es un **espacio de Hausdorff** si para cualesquiera dos puntos distintos $x,y \in X$ existen $U_1, U_2 \in \tau$ tales que $x \in U_1$, $y \in U_2$ y $U_1 \cap U_2 = \emptyset$.

Diremos que $\mathcal{A} \subset C(X,\mathbb{R})$ o $\mathcal{A} \subset C(X)$ **separa puntos** si para cualesquiera $x,y \in X$ distintos existe $f \in \mathcal{A}$ tal que $f(x) \neq f(y)$. $\mathcal{A}$ es un **álgebra** si es un subespacio de $C(X,\mathbb{R})$ o $C(X)$ y $fg \in \mathcal{A}$ cuando $f,g \in \mathcal{A}$. Finalmente, $\mathcal{A} \subset C(X,\mathbb{R})$ es un **retículo** si $f \vee g \in \mathcal{A}$ y $f \wedge g \in \mathcal{A}$ cuando $f,g \in \mathcal{A}$.

Para probar el teorema de Stone–Weierstrass primero se probarán una serie de lemas, en los cuales supondremos que (X,τ) es un espacio de Hausdorff compacto.

Lema D.1. *Consideremos el álgebra dada por $\mathbb{R}^2$ con la suma y multiplicación entrada a entrada. Las únicas subálgebras de $\mathbb{R}^2$ son $\mathbb{R}^2$, $\{(0,0)\}$ y los subespacios generados por $(1,0)$, $(0,1)$ y $(1,1)$ de forma respectiva.*

Demostración. Consideremos $\mathfrak{A} \subset \mathbb{R}^2$ un álgebra distinta de $\{(0,0)\}$ y consideremos $(0,0) \neq (a,b) \in \mathfrak{A}$. Si (a,b) es tal que $a \neq b$, $a \neq 0$ y $b \neq 0$ entonces $(a^2,b^2) \in \mathfrak{A}$ es linealmente independiente de (a,b) y por tanto $\mathfrak{A} = \mathbb{R}^2$. Los demás casos se obtienen de forma análoga. □

Lema D.2. *Para cualquier $\epsilon > 0$ existe un polinomio, P, con valores en $\mathbb{R}$ tal que $P(0) = 0$ y $||x| - P(x)| < \epsilon$ para toda $x \in [-1,1]$.*

Demostración. Consideremos la serie de Maclaurin de $(1-t)^{1/2}$ dada por

$$g(t) = 1 + \sum_{n=1}^{\infty} \left(\prod_{i=1}^{n} \frac{2n-3}{2} \right) \left(-\frac{t^n}{n!} \right),$$

la cual converge cuando $t \in (-1,1)$ por el criterio de D'Lambert. Más aún, si $K \subset (-1,1)$ es compacto entonces la serie dada converge de forma uniforme sobre K, así como lo hace

$$h(t) = \sum_{n=1}^{\infty} \left(\prod_{i=1}^{n} \frac{2n-3}{2} \right) \left(-\frac{t^{n-1}}{(n-1)!} \right),$$

por lo que $h(t) = g'(t)$ para $t \in (-1,1)$. Además tenemos que $g(t) = -2(1-t)g'(t)$ de donde $(1-t)^{-1/2}g(t)$ es constante; esto implica que $g(t) = (1-t)^{1/2}$ para $t \in (-1,1)$. Por el teorema de convergencia monótona

$$\lim_{t \nearrow 1}(1-g(t)) = \lim_{t \nearrow 1} \left(1-(1-t)^{1/2}\right) = 1.$$

Esto junto al teorema de convergencia dominada implica que la serie de Maclaurin converge absolutamente y de forma uniforme a $(1-t)^{1/2}$ en $[-1,1]$.

Por tanto, para $\epsilon > 0$ dado existe $N \in \mathbb{N}$ tal que $g_N(t) := 1 + \sum_{n=1}^{N} \left(\prod_{i=1}^{n} \frac{2n-3}{2} \right) \left(-\frac{t^n}{n!} \right)$ es un polinomio tal que $\left|(1-t)^{1/2} - g_N(t)\right| < \epsilon/2$ cuando $t \in [-1,1]$. Definiendo $Q(x) = g_N(1-x^2)$ obtenemos un polinomio tal que $||x| - Q(x)| < \epsilon/2$ cuando $x \in [-1,1]$, por lo que $|Q(0)| < \epsilon/2$. El polinomio que concluye la prueba está dado por $P(x) := Q(x) - Q(0)$. $\square$

Lema D.3. *Si $\mathcal{A}$ es una subálgebra de $C(X,\mathbb{R})$ cerrada entonces $|f| \in \mathcal{A}$ cuando $f \in \mathcal{A}$. Además $\mathcal{A}$ es un retículo.*

Demostración. Si $f \in \mathcal{A}$ es idénticamente cero la primera parte es trivial. Así supongamos que $f \neq 0$ y sea $g = f/\|f\|_u$, por lo que g toma valores en $[-1,1]$. Para $\epsilon > 0$ arbitrario, sea P el polinomio dado en el lema anterior, por lo que $\||g| - P \circ g\|_u < \epsilon$. Dado que $P(0) = 0$ y $g \in \mathcal{A}$, se sigue que $P \circ g \in \mathcal{A}$. Como $\mathcal{A}$ es cerrada y $\epsilon > 0$ era arbitraria, entonces $|g| \in \mathcal{A}$ y por lo tanto $|f| = \|f\|_u |g| \in \mathcal{A}$.

La segunda parte del enunciado se sigue de la primera y que $\mathcal{A}$ es un álgebra, pues

$$f \vee g = \frac{1}{2}(f+g+|f-g|) \quad \text{y} \quad f \wedge g = \frac{1}{2}(f+g-|f-g|). \qquad \square$$

Lema D.4. *Sea $\mathcal{A} \subset C(X,\mathbb{R})$ un retículo cerrado y $f \in C(X,\mathbb{R})$. Si para cualesquiera $x, y \in X$ distintos existe $g_{xy} \in \mathcal{A}$ tal que $g_{xy}(x) = f(x)$ y $g_{xy}(y) = f(y)$, entonces $f \in \mathcal{A}$.*

Demostración. Para $\epsilon > 0$ dada, para $x, y \in X$ distintos definamos $U_{xy} = \{z \in X : f(z) < g_{xy}(z) + \epsilon\}$ y $V_{xy} = \{z \in X : f(z) > g_{xy}(z) - \epsilon\}$. Por continuidad U_{xy} y V_{xy} son abiertos y $x, y \in U_{xy} \cap V_{xy}$ por hipótesis. Si fijamos $y \in X$, entonces $\{U_{xy} : x \in X\}$ forma una cubierta abierta de X y por ser éste compacto existe una subcubierta finita de X, $\{U_{x_iy} : x_i \in X, i \in \{1,\ldots,n_1\}\}$. Sea $g_y = g_{x_1y} \vee \cdots \vee g_{x_{n_1}y}$; por construcción $f < g_y + \epsilon$ en X y $f > g_y - \epsilon$ en $V_y = \bigcap_{i=1}^{n_1} V_{x_iy}$. Es claro que $\{V_y : y \in X\}$ es otra cubierta abierta de X, por lo que existe una subcubierta finita $\{V_{y_i} : y_i \in X, i \in \{1,\ldots,n_2\}\}$. Así, al definir $g = g_{y_1} \wedge \cdots \wedge g_{y_{n_2}}$, notamos que $\|f-g\|_u < \epsilon$. Se concluye por ser $\mathcal{A}$ retículo y $\epsilon > 0$ arbitrario. $\square$

Teorema D.5 (Stone–Weierstrass real)**.** *Sea (X,τ) un espacio de Hausdorff compacto. Si $\mathcal{A} \subset C(X,\mathbb{R})$ es una subálgebra que separa puntos, entonces $\bar{\mathcal{A}} = C(X,\mathbb{R})$ o $\bar{\mathcal{A}} = \{f \in C(X,\mathbb{R}) : f(x_0) = 0\}$ para alguna $x_0 \in X$. La primera conclusión se satisface si y sólo si $\mathcal{A}$ contiene a las funciones constantes.*

Demostración. Por continuidad, podemos extender la propiedad de álgebra y retículo a $\bar{\mathcal{A}}$. Dados $x, y \in X$ distintos, definamos $\mathcal{A}_{xy} = \{(f(x), f(y)) : f \in \bar{\mathcal{A}}\}$ que es una subálgebra de $\mathbb{R}^2$. Si $\mathcal{A}_{xy} = \mathbb{R}^2$ para cualesquiera x y y, como $\bar{\mathcal{A}}$ es un retículo cerrado, por el lema anterior $\bar{\mathcal{A}} = C(X, \mathbb{R})$. En caso de que exista un $\mathcal{A}_{xy}$ que sea subálgebra propia de $\mathbb{R}^2$; como $\mathcal{A}$ separa puntos no puede ser el caso que la subálgebra sea $\{(0, 0)\}$ o el subespacio generado por $(1, 1)$. Por lo tanto A_{xy} será algunos de los subespacios generados por $(1, 0)$ o $(0, 1)$; en cualquier caso existirá $x_0 \in X$ tal que $f(x_0) = 0$ para toda $f \in \bar{\mathcal{A}}$. Como $\mathcal{A}$ separa puntos, podemos asegurar la unicidad de x_0, por lo que si $x, y \neq x_0$ entonces $\mathcal{A}_{xy} = \mathbb{R}^2$. Nuevamente por el lema anterior, por ser $\bar{\mathcal{A}}$ un retículo cerrado, $\bar{\mathcal{A}} = \{f \in C(X, \mathbb{R}) : f(x_0) = 0\}$.

Para concluir basta notar que si $\mathcal{A}$ contiene las funcionces constantes, entonces no habrá $x_0 \in X$ tal que $f(x_0) = 0$ para toda $f \in \mathcal{A} \subset \bar{\mathcal{A}}$. $\square$

Diremos que (X, τ) es **localmente compacto** si para cualquier $x \in X$ existen $U \in \tau$ y $K \subset X$ compacto tales que $x \in U \subset K$. Si tenemos un espacio (X, τ) que es localmente compacto entonces podremos definir una topología τ^* sobre $Y = X \cup \{\infty\}$ tal que (Y, τ^*) sea un espacio compacto, llamado la **compactificación de un punto** de X. En particular τ^* está dado por lo conjuntos tales que $\infty \notin U$ y $U \subset X$ es abierto o $\infty \in U$ y $U^c \subset X$ es cerrado y compacto. En este caso $f \in C(X)$ se puede extender continuamente a una función en Y si y sólo si $f = g + c$ con $g \in C_0(X)$ y c una constante, en cuyo caso $f(\infty) = c$.

Teorema D.6 (Stone–Weierstrass localmente compacto)**.** *Sea (X, τ) un espacio de Hausdorff localmente compacto. Si $\mathcal{A} \subset C_0(X, \mathbb{R})$ es una subálgebra que separa puntos, entonces $\bar{\mathcal{A}} = C_0(X, \mathbb{R})$ o $\bar{\mathcal{A}} = \{f \in C_0(X, \mathbb{R}) : f(x_0) = 0\}$ para alguna $x_0 \in X$.*

Demostración. Supongamos que existe $x_0 \in X$ tal que $f(x_0) = 0$ para toda $f \in \mathcal{A}$, entonces podemos aplicar la compactificación de un punto a $X - \{x_0\}$; en caso de que no exista tal x_0 entonces aplicamos la compactificación de un punto a X. En cualquier caso la extensión continua de $f \restriction_{X - \{x_0\}}$ o f, de forma respectiva, a la compactificación será tal que evaluada en ∞ dé cero, esto para toda $f \in \mathcal{A}$. Por Stone–Weierstrass real se concluye la demostración. $\square$

Apéndice E. **Lema de Urysohn**

Un resultado importante de la topología, que es usado regularmente en la teoría de la medida y probabilidad para aproximar ciertas funciones indicadoras en espacios con ciertas propiedad es aquel conocido como *lema de Urysohn*. Para introducirlo debemos de definir primero el concepto de *espacio normal*.

Definición E.1. Diremos que un espacio topólogico (X, τ) es **normal** si para $A, B \subset X$ cerrados y disjuntos, existen $U, V \in \tau$ disjuntos tales que $A \subset U$ y $B \subset V$.

Teorema E.2 (Lema de Urysohn)**.** *Un espacio (X, τ) es normal si y sólo si para cualesquiera $A, B \subset X$ cerrados existe $f \in C(X, [0, 1])$ tal que $f(A) = 0$ y $f(B) = 1$.*

Demostración. Para la suficiencia, notemos que $A \subset f^{-1}((-\infty, 1/2))$ y $B \subset f^{-1}((1/2, \infty))$. Por continuidad se sigue el resultado.

De forma recíproca, supongamos que (X, τ) es normal y $A, B \subset X$ son cerrados y ajenos. Por normalidad existe $U_{1/2} \in \tau$ tal que $A \subset U_{1/2}$ y $\overline{U}_{1/2} \cap B = \emptyset$. De esta manera tendremos que A y $X - U_{1/2}$ son cerrados y disjuntos, al igual que $\overline{U}_{1/2}$ y B, y por la normalidad de (X, τ) existirán $U_{1/4}, U_{3/4} \in \tau$ tales que

$$A \subset U_{1/4}, \quad \overline{U}_{1/4} \subset U_{1/2}, \quad \overline{U}_{1/2} \subset U_{3/4} \quad y \quad \overline{U}_{3/4} \cap B = \emptyset.$$

Procediendo de forma inductiva sobre $n \in \mathbb{N}$, supongamos que se han definido $\{U_{r2^{-n}}\}_{r=1}^{2^n-1}$ abiertos tales que $A \subset U_{r2^{-n}}$ y $\overline{U}_{r2^{-n}} \cap B = \emptyset$ para cada $r \in \{1, \ldots, 2^n - 1\}$ y $\overline{U}_{r2^{-n}} \subset \overline{U}_{s2^{-n}}$ cuando $1 \leq r < s \leq 2^{n-1}$. Nuevamente por normalidad existirán $U_{2^{-n-1}}$ y $U_{1-2^{-n-1}}$ abiertos tales que $A \subset U_{2^{-n-1}}$, $\overline{U}_{2^{-n-1}} \subset U_{2^{-n}}$, $\overline{U}_{1-2^{-n}} \subset U_{1-2^{-n-1}}$ y $\overline{U}_{1-2^{-n-1}} \cap B = \emptyset$. Por la misma normalidad, para $r \in \{2, \ldots, 2^n - 1\}$ podremos construir conjuntos tales que $\overline{U}_{\frac{r-1}{2^n}} \subset U_{\frac{2r-1}{2^{n+1}}}$ y $\overline{U}_{\frac{2r-1}{2^{n+1}}} \subset U_{r2^{-n}}$. De esta forma obtenemos una familia de abiertos $\{U_r\}_{r\in\Delta}$, donde $\Delta = \{r2^{-n} : n \in \mathbb{N}; r \in \{1, \ldots, 2^n\}\}$, tal que $A \subset U_r$ y $\overline{U}_r \cap B = \emptyset$ para cada $r \in \Delta$, además de que $\overline{U}_r \subset U_s$ cuando $r < s$.

Definamos ahora $f : X \to \mathbb{R}$ por

$$f(x) = \begin{cases} 1 & \textit{si } x \notin \bigcup_{r\in\Delta} U_r, \\ \inf\{r \in \Delta : x \in U_r\} & \textit{en otro caso}. \end{cases}$$

Por construcción $f(A) = 0$ y $f(B) = 1$ por lo que acabaremos la demostración al probar que f es continua. Notando que para $0 \leq \alpha < \beta \leq 1$, $f(x) > \alpha$ y $f(y) < \beta$ si y sólo si existen $r > \alpha$ y $s < \beta$ tales que $x \notin U_r$ y $y \in U_s$ si y sólo si $x \in \bigcup_{r>\alpha}(\overline{U}_r)^c \in \tau$ y $y \in \bigcup_{s<\alpha} U_s \in \tau$ obtendremos que

$$f^{-1}((-\infty, \alpha)) = \begin{cases} \emptyset & \textit{si } \alpha \leq 0, \\ \bigcup_{r<\alpha} U_r & \textit{si } 0 < \alpha \leq 1, \\ X & \textit{si } \alpha > 1, \end{cases} \quad y \quad f^{-1}((\alpha, \infty)) = \begin{cases} X & \textit{si } \alpha < 0, \\ \bigcup_{s>\alpha}(\overline{U}_s)^c & \textit{si } 0 \leq \alpha < 1, \\ \emptyset & \textit{si } \alpha \geq 1. \end{cases}$$

Esto último prueba que f es continua. $\square$

Bibliografía

[ADD00] Robert B. Ash and Catherine A. Doléans-Dade. *Probability and Measure Theory*. Harcourt/Academic Press, 2 edition, 2000.

[Bar95] Robert G. Bartle. *Elements of Integration*. Wiley-Interscience, 1 edition, 1995.

[Bil86] Patrick Billingsley. *Probability and Measure*. Wiley series in probability and mathematical statistics. Wiley, 2 edition, 1986.

[Dud02] R. M. Dudley. *Real Analysis and Probability*. Cambridge studies in advanced mathematics 74. Cambridge University Press, 2 edition, 2002.

[Dur10] Rick Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4 edition, 2010.

[Fol99] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs and Tracts. Wiley-Interscience, 2 edition, 1999.

[JP02] Jean Jacod and Philip Protter. *Probability Essentials*. Springer, 2 edition, 2002.

[RF10] Halsey Royden and Patrick Fitzpatrick. *Real Analysis*. Prentice Hall, 4 edition, 2010.

[Rud86] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, 3 edition, 1986.