Capítulo 3

Funciones Medibles e Integración

3.1. Introducción

Sea $X : \Omega \to \Omega'$ donde Ω' es el recorrido de X, es decir, para todo $\omega' \in \Omega'$ existe $\omega \in \Omega$ con $X(\omega) = \omega'$. X determina la función

$$X^{-1} : \mathcal{P}(\Omega') \to \mathcal{P}(\Omega)$$

definida por

$$X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\} = \{X \in A'\}$$

para $A' \subset \Omega'$.

Propiedades.

(i) $X^{-1}(\emptyset) = \emptyset,\ \ X^{-1}(\Omega') = \Omega$.

(ii) $X^{-1}(A'^c) = (X^{-1}(A'))^c$ de modo que $X^{-1}(\Omega' - A') = \Omega - X^{-1}(A')$.

(iii) $X^{-1}(\cup_{t\in T} A'_t) = \cup_{t\in T} X^{-1}(A'_t),\ \ X^{-1}(\cap_{t\in T} A'_t) = \cap_{t\in T} X^{-1}(A'_t)$.

Veamos la demostración de (ii)

$$\omega \in X^{-1}(A'^c) \Leftrightarrow X(\omega) \in A'^c \Leftrightarrow X(\omega) \notin A' \Leftrightarrow \omega \notin X^{-1}(A') \Leftrightarrow \omega \in (X^{-1}(A'))^c.$$

Si $\mathcal{C}' \subset \mathcal{P}(\Omega')$ es una colección de subconjuntos de Ω', definimos

$$X^{-1}(\mathcal{C}') = \{X^{-1}(C') : C' \in \mathcal{C}'\}.$$

Proposición 3.1 *Si $\mathcal{F}'$ es una σ-álgebra de subconjuntos de Ω' entonces $X^{-1}(\mathcal{F}')$ es una σ-álgebra de subconjuntos de Ω.*

Demostración. Usando las propiedades de X^{-1},

(i) $\Omega' \in \mathcal{F}' \Rightarrow X^{-1}(\Omega') = \Omega \in X^{-1}(\mathcal{F}')$.

(ii) $A' \in \mathcal{F}' \Rightarrow (A')^c \in \mathcal{F}'$. Por lo tanto si $X^{-1}(A') \in X^{-1}(\mathcal{F}')$ tenemos $(X^{-1}(A'))^c = X^{-1}((A')^c) \in X^{-1}(\mathcal{F}')$.

(iii) Si $X^{-1}(A'_n) \in X^{-1}(\mathcal{F}')$ entonces $\cup_n X^{-1}(A'_n) = X^{-1}(\cup_n A'_n) \in X^{-1}(\mathcal{F}')$ porque $\cup_n A'_n \in \mathcal{F}'$.

$\blacksquare$

Proposición 3.2 *Si $\mathcal{C}'$ es una colección de subconjuntos de Ω' entonces*

$$X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}')).$$

Demostración. Por la proposición 3.1, $X^{-1}(\sigma(\mathcal{C}'))$ es una σ-álgebra y $X^{-1}(\mathcal{C}') \subset X^{-1}(\sigma(\mathcal{C}'))$ porque $\mathcal{C}' \subset \sigma(\mathcal{C}')$ y por minimalidad

$$\sigma(X^{-1}(\mathcal{C}')) \subset X^{-1}(\sigma(\mathcal{C}')).$$

Para ver el recíproco definimos

$$\mathcal{F}' = \{B' \in \mathcal{P}(\Omega') : X^{-1}(B') \in \sigma(X^{-1}(\mathcal{C}'))\}.$$

Entonces $\mathcal{F}'$ es una σ-álgebra ya que

(i) $\Omega' \in \mathcal{F}'$ porque $X^{-1}(\Omega') = \Omega \in \sigma(X^{-1}(\mathcal{C}'))$.

(ii) Si $A' \in \mathcal{F}'$ entonces $(A')^c \in \mathcal{F}'$ ya que

$$X^{-1}(A'^c) = (X^{-1}(A'))^c \in \sigma(X^{-1}(\mathcal{C}))$$

si $X^{-1}(A') \in \sigma(X^{-1}(\mathcal{C}'))$.

(iii) Si $B'_n \in \mathcal{F}'$ entonces $X^{-1}(B'_n) \in \sigma(X^{-1}(\mathcal{C}'))$, por lo tanto

$$X^{-1}(\cup_n B'_n) = \cup_n X^{-1}(B'_n) \in \sigma(X^{-1}(\mathcal{C}'))$$

Por definición, $X^{-1}(\mathcal{F}') \subset \sigma(X^{-1}(\mathcal{C}'))$. Además $\mathcal{C}' \subset \mathcal{F}'$ porque $X^{-1}(\mathcal{C}') \subset \sigma(X^{-1}(\mathcal{C}'))$. Como $\mathcal{F}'$ es una σ-álgebra, $\sigma(\mathcal{C}') \subset \mathcal{F}'$ y por lo tanto

$$X^{-1}(\sigma(\mathcal{C}')) \subset X^{-1}(\mathcal{F}') \subset \sigma(X^{-1}(\mathcal{C}')).$$

■

3.2. Funciones Medibles

Definición 3.1 Sean $(\Omega, \mathcal{F})$ y $(\Omega', \mathcal{F}')$ dos espacios medibles. Una función $X : \Omega \to \Omega'$ es *medible* si $X^{-1}(\mathcal{F}') \subset \mathcal{F}$. En este caso escribimos $X \in \mathcal{F}/\mathcal{F}'$ o $X \in \mathcal{F}$ si esta claro cual es la σ-álgebra $\mathcal{F}'$.

Observación 3.1

1. En Teoría de Probabilidad se dice que X es una *variable aleatoria.*
2. La medibilidad de X no implica que $X(A) \in \mathcal{F}'$ para todo $A \in \mathcal{F}$. Por ejemplo, si $\mathcal{F}' = \{\emptyset, \Omega'\}$ entonces toda función $X : \Omega \to \Omega'$ es medible, cualquiera sea $\mathcal{F}$, pero si $A \in \mathcal{F}$ y $X(A)$ es un subconjunto no vacío y propio de Ω' entonces $X(A) \notin \mathcal{F}'$.
3. Si $(\Omega, \mathcal{F})$ es un espacio medible y $X : \Omega \to \mathbb{R}^n$, decimos que X es (Borel) medible si X es medible cuando tomamos la σ-álgebra de los borelianos $\mathcal{B}_n$ como σ-álgebra en $\mathbb{R}^n$. Si Ω es un subconjunto de $\mathbb{R}^k$ o es $\mathbb{R}^k$ y decimos que X es Borel medible, suponemos que $\mathcal{F} = \mathcal{B}_k$.

Ejemplo 3.1
Si $A \in \mathcal{B}$ la función $\mathbf{1}_A(x)$ es una función medible.

Proposición 3.3 *Sea $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ y supongamos que $\mathcal{F}' = \sigma(\mathcal{C}')$, entonces X es medible si y sólo si $X^{-1}(\mathcal{C}) \subset \mathcal{F}$.*

Demostración.

$$X^{-1}(\mathcal{C}) \subset \mathcal{F} \Rightarrow \sigma(X^{-1}(\mathcal{C}')) \subset \mathcal{F}$$

y en consecuencia

$$X^{-1}(\sigma(\mathcal{C}')) = X^{-1}(\mathcal{F}') = \sigma(X^{-1}(\mathcal{C}')) \subset \mathcal{F}.$$

■

Ejemplo 3.2
Una función continua X de $\mathbb{R}^k$ en $\mathbb{R}^n$ es medible porque si $\mathcal{O}$ es la clase de los conjuntos abiertos de $\mathbb{R}^n$, entonces para todo $A \in \mathcal{O}$ se tiene que $X^{-1}(A)$ es abierto y por lo tanto está en $\mathcal{B}_k$.

Ejemplo 3.3
Para mostrar que una función $X : \Omega \to \mathbb{R}$ es Borel medible, es suficiente mostrar que $\{\omega : X(\omega) > c\} \in \mathcal{F}$ para todo real c. Porque si $\mathcal{C}$ es la clase de conjuntos (c, ∞), $c \in \mathbb{R}$, entonces $\sigma(\mathcal{C}) = \mathcal{B}$. Los conjuntos $\{\omega : X(\omega) > c\}$ se pueden reemplazar por $\{\omega : X(\omega) \geq c\}$, $\{\omega : X(\omega) < c\}$ o $\{\omega : X(\omega) \leq c\}$.

Definición 3.2 Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ una variable aleatoria. Definimos la función $P_X = P \circ X^{-1}$ en $\mathcal{F}'$ por

$$P_X(A') = P \circ X^{-1}(A') = P(X^{-1}(A')).$$

P_X es una probabilidad sobre $(\Omega', \mathcal{F}')$ que se conoce como la *medida inducida* por, la *distribución* o la *ley* de X

Para verificar que P_X es una medida de probabilidad sobre $\mathcal{F}'$ observamos que

1. $P_X(\Omega') = P(\Omega) = 1$.
2. $P_X(A') \geq 0, \ \forall A' \in \mathcal{F}'$.
3. Si $(A'_n)_{n\geq 1}$ son disjuntos 2 a 2 entonces

$$P_X(\cup_n A'_n) = P(\cup_n X^{-1}(A'_n)) = \sum_n P(X^{-1}(A'_n)) = \sum_n P_X(A'_n)$$

porque $(X^{-1}(A'_n))_{n\geq 1}$ son disjuntos en $\mathcal{F}$.

Si X toma valores en $\mathbb{R}$ su distribución P_X es una probabilidad en $\mathbb{R}$ caracterizada por su función de distribución $F_X(x) = P_X((-\infty, x]) = P(X \leq x)$. Decimos que F_X es la función de distribución de X. Si F_X tiene una densidad f_X decimos que es la densidad de X.

Proposición 3.4 *La composición de funciones medibles es medible.*

Demostración. Sean $X : (E, \mathcal{E}) \to (F, \mathcal{F})$ y $Y : (F, \mathcal{F}) \to (G, \mathcal{G})$ dos funciones medible. Sea $A \in \mathcal{G}$ entonces

$$(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A)).$$

Como Y es medible, $B = Y^{-1}(A) \in \mathcal{F}$. Como X es medible, $X^{-1}(B) \in \mathcal{E}$. ■

Proposición 3.5 *Si X, Y son funciones medibles, también lo son $X + Y$, $X \cdot Y$, $X \vee Y$, $X \wedge Y$ y X/Y $(Y \neq 0)$.*

Demostración. Ejercicio.

Dada una función X definimos X^+ y X^- por

$$X^+(\omega) = \text{máx}\{0, X(\omega)\}, \qquad X^-(\omega) = -\text{mín}\{0, X(\omega)\}.$$

entonces $X(\omega) = X^+(\omega) - X^-(\omega)$ y $|X(\omega)| = X^+(\omega) + X^-(\omega)$ y ambas funciones X^+ y X^- son no-negativas.

Corolario 3.1 *Las funciones X^+, X^- y $|X|$ son medibles.*

Demostración. Ejercicio.

Corolario 3.2 *Si $\mathbf{X} = (X_1, \dots, X_k)$ es un vector aleatorio y $g : \mathbb{R}^k \to \mathbb{R}$ es medible, entonces $g(\mathbf{X})$ es medible. En particular, si g es continua entonces es medible y el resultado vale.*

Demostración. Ejercicio.

Ejemplos 3.4

$$\sum_{i=1}^k x_i, \qquad \frac{1}{k}\sum_{i=1}^k x_i, \qquad \sup_{1\leq i\leq k} x_i, \qquad \prod_{i=1}^k x_i, \qquad \sum_{i=1}^k x_i^2$$

son todas funciones continuas de $\mathbb{R}^k$ en $\mathbb{R}$.

Otro ejemplo interesante es la proyección $\pi_i : \mathbb{R}^k \to \mathbb{R}$ definida por

$$\pi_i(x_1, \dots, x_k) = x_i.$$

π_i es continua y si $\mathbf{X} = (X_1, \dots, X_k)$ es un vector aleatorio entonces $\pi_i(\mathbf{X}) = X_i$ es una variable aleatoria.

Proposición 3.6 *$\mathbb{X} = (X_1, \dots, X_k)$ es un vector aleatorio si y sólo si X_i es una variable aleatoria para $i = 1, \dots, k$.*

Demostración. El ejemplo anterior muestra que la condición es necesaria. Para el recíproco comenzamos con

$$\mathcal{B}_k = \sigma(\mathcal{S}_k)$$

donde $\mathcal{S}_k$ es la clase de los rectángulos abiertos en $\mathbb{R}^k$. Sean $X_1, \dots, X_k$ variables aleatorias y $B = I_1 \times \cdots \times I_k$ un rectángulo de lados $I_1, \dots, I_k$. Entonces

$$X^{-1}(B) = \bigcap_{i=1}^k X_i^{-1}(I_i).$$

Como cada X_i es una variable aleatoria, $X_i^{-1}(I_i) \in \mathcal{B}_k$, de modo que $X^{-1}(B) \in \mathcal{B}_k$ y $X^{-1}(\mathcal{S}_k) \subset \mathcal{B}_k$. En consecuencia X^{-1} es medible. ∎

Proposición 3.7 *$\mathbf{X} = (X_1, X_2, \dots)$ es una sucesión aleatoria si y sólo si para cada $i \geq 1$ la i-ésima componente X_i es una v.a. Además, $\mathbf{X}$ es una sucesión aleatoria si y sólo si $(X_1, \dots, X_k)$ es un vector aleatorio para todo k.*

Demostración. Ejercicio.

Proposición 3.8 *Sean $X_1, X_2, \dots$ funciones medibles definidas sobre $(\Omega, \mathcal{F})$, entonces* *(i)* $\sup X_n,\ \inf X_n,\ \limsup X_n,\ \liminf X_n$ *son funciones medibles.*

(ii) Si $X_n(\omega) \to X(\omega)$ *para todo* $\omega \in \Omega$, X *es una función medible.*

(iii) El conjunto en el cual la sucesión (X_n) *converge es medible.*

Demostración. (i) $\{X_n \leq \alpha\} \in \mathcal{F}$ porque X_n es medible. Por lo tanto $\{\sup_n X_n \leq \alpha\} = \cap_n \{X_n \leq \alpha\}$ está en $\mathcal{F}$, para todo α y en consecuencia $\sup_n X_n$ es medible. De manera similar $\{\inf_n X_n \leq \alpha\} = \cup_n \{X_n \leq \alpha\} \in \mathcal{F}$ e $\inf_n X_n$ es medible. Como consecuencia $\liminf X_n$ y $\limsup_n X_n$ son medibles.

(ii) Es consecuencia de (i) porque $\lim X_n = \limsup_n X_n = \liminf X_n$.

(iii) Tenemos

$$\begin{aligned}\{\omega : \lim_n X_n(\omega) \text{ existe }\}^c &= \{\omega : \liminf X_n(\omega) < \limsup X_n(\omega)\} \\ &= \bigcup_{r\in\mathbb{Q}} \{\liminf X_n < r < \limsup X_n\} \\ &= \bigcup_{r\in\mathbb{Q}} \{[\liminf X_n < r] \cap [\limsup X_n \leq r]^c\} \in \mathcal{B}.\end{aligned}$$

Observación 3.2 Observamos que la clase de las funciones medibles no es cerrada si la operaciones anteriores se realizan una cantidad no-numerable de veces. Es decir si A no es numerable y $f_\alpha : \Omega \to \mathbb{R}$ es medible para cada $\alpha \in A$, no es necesariamente cierto que

$$f(\omega) = \sup_{\alpha\in A} f_\alpha(\omega)$$

sea medible. Por ejemplo, si $A \subset [0,1]$ es un conjunto no medible y ponemos

$$f_\alpha(\omega) = \begin{cases} 1 & \text{si } \omega = \alpha \\ 0 & \text{si } \omega \neq \alpha \end{cases}$$

entonces para cada $\alpha \in A$, f_α es $\mathcal{M}$-medible pero

$$\sup_{\alpha\in A} f_\alpha(\omega) = \mathbf{1}_A(\omega)$$

no es medible.

3.3. σ-Álgebras Generadas por Funciones

Definición 3.3 Sea $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ una variable aleatoria. La σ-álgebra generada por X, $\sigma(X)$, se define como

$$\sigma(X) = X^{-1}(\mathcal{B}).$$

Equivalentemente,

$$\sigma(X) = \{\{X \in A\},\ A \in \mathcal{B}\}.$$

En general, si $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ definimos $\sigma(X) = X^{-1}(\mathcal{F})$.

Si $\mathcal{G} \subset \mathcal{F}$ es una sub-σ-álgebra de $\mathcal{F}$ decimos que X es medible respecto a $\mathcal{G}$ si $\sigma(X) \subset \mathcal{G}$.

Si para cada $t \in T$ tenemos $X_t : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ definimos

$$\sigma(X_t, t \in T) = \sigma(\cup_{t\in T} \sigma(X_t))$$

la menor σ-álgebra que contiene a todas las $\sigma(X_t)$.

Ejemplos 3.5
1. Si $X(\omega) = 32$ para todo ω entonces

$$\sigma(X) = \{\{X \in B\}, B \in \mathcal{B}\} = \sigma(\emptyset, \Omega\} = \{\emptyset, \Omega\}$$

2. Sea $X = \mathbf{1}_A$ para algún $A \in \mathcal{B}$. X toma valores en $\{0, 1\}$ y

$$X^{-1}(\{0\}) = A^c, \qquad X^{-1}(\{1\}) = A,$$

y por lo tanto $\sigma(X) = \{\emptyset, \Omega, A, A^c\}$.

Definición 3.4 Una función es *simple* si es medible y toma un número finito de valores.

Si X toma valores $a_1, \dots a_k$, definimos

$$A_i = X^{-1}(\{a_i\}) = \{X = a_i\}.$$

Entonces $\{A_i, i = 1, \dots, k\}$ es una partición de Ω y podemos representar a X como

$$X = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}$$

y

$$\sigma(X) = \sigma(A_1, \dots, A_k) = \{\cup_{i \in I} A_i : I \subset \{i, \dots. k\}\}.$$

Proposición 3.9 *Sea X una función medible y $\mathcal{C}$ una clase de subconjuntos de $\mathbb{R}$ tales que $\sigma(\mathcal{C}) = \mathcal{B}$. Entonces*

$$\sigma(X) = \sigma(\{X \in B\} : B \in \mathcal{C})$$

Demostración.

$$\begin{aligned}\sigma(\{X \in B\} : B \in \mathcal{C}\} &= \sigma(X^{-1}(B), B \in \mathcal{C}\} = \sigma(X^{-1}(\mathcal{C}))\\ &= X^{-1}(\sigma(\mathcal{C})) = X^{-1}(\mathcal{B}) = \sigma(X).\end{aligned}$$

■

El siguiente teorema es fundamental para la definición de la integral de Lebesgue.

Teorema 3.1 (de Aproximación) *Dada una función medible y no negativa $X : \Omega \to \mathbb{R}^+$, existe una sucesión creciente de funciones simples no negativas $X_n : \Omega \to \mathbb{R}^+$ tales que $X_n \uparrow X$.*

Demostración. Para todo entero positivo n definimos

$$\begin{aligned}Q_{p,n} &= \{\omega : \frac{p-1}{2^n} \le X(\omega) < \frac{p}{2^n}\}, \quad p = 1, \dots, 2^{2n}\\ Q_{0,n} &= \Omega - \cup_{p=1}^{2^{2n}} Q_{p,n} = \{\omega : X(\omega) \ge 2^n\}\end{aligned}$$

Entonces, como X es $\mathcal{F}$-medible, $Q_{p,n} \in \mathcal{F}$ y los conjuntos $Q_{p,n}, p = 0, 1, \dots, 2^{2n}$ forman una partición de Ω. Definimos

$$X_n(\omega) = \begin{cases} \frac{p-1}{2^n} & \text{para } \omega \in Q_{p,n},\ p = 1, 2, \dots, 2^{2n} \\ 2^n & \text{para } \omega \in Q_{0,n}. \end{cases}$$

Esta función es simple y tenemos que

$$0 \le X_n \le X.$$

Si $\omega \in Q_{p,n}$ entonces o bien $\omega \in Q_{2p-1,n+1}$ o $\omega \in Q_{2p,n+1}$ y en consecuencia

$$X_n(\omega) = X_{n+1}(\omega) \qquad \text{o} \qquad X_n(\omega) + \frac{1}{2^{n+1}} = X_{n+1}(\omega).$$

Además, si $\omega \in Q_{0,n}$ entonces $X_n(\omega) = 2^n \leq X(\omega)$, y en consecuencia $\omega \in Q_{0,n+1}$ o $\omega \in Q_{p,n+1}$ para algún $p \geq 2^{2n+1} + 1$. En cualquier caso $X_{n+1}(\omega) \geq X_n(\omega)$. Por lo tanto para todo entero n

$$X_n(\omega) \leq X_{n+1}(\omega) \quad \text{para todo } \omega \in \Omega,$$

y la sucesión de funciones simples (X_n) es creciente.

Si $X(\omega)$ es finita, entonces para $2^n > X(\omega)$ tenemos que

$$0 \leq X(\omega) - X_n(\omega) < 2^{-n}$$

y en consecuencia $X_n(\omega) \to X(\omega)$ cuando $n \to \infty$. Por otro lado, si $X(\omega) = \infty$ entonces para todo n, $X_n(\omega) = 2^n$, y de nuevo tenemos que $X_n(\omega) \to X(\omega)$ cuando $n \to \infty$. ■

3.4. Integrales

3.4.1. Funciones Simples No Negativas

Si $X(\omega) = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}(\omega)$ con $c_i \geq 0$, $i = 1, \dots, n$, definimos

$$\int X \, d\mu = \sum_{i=1}^{n} c_i \mu(A_i).$$

Esta suma siempre está definida, aunque su valor puede ser $+\infty$, ya que todos los términos son no-negativos, y se conoce como la *integral* de X respecto a μ, el *valor esperado* o la *esperanza* de X. Tomamos la convención de que si $x_i = 0$ y $\mu(A_i) = \infty$ entonces $c_i \mu(A_i) = 0$. Como la representación de una función simple no es única, debemos verificar que la definición es independiente de la representación que usemos. Supongamos que

$$X(\omega) = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}(\omega) = \sum_{j=1}^{m} d_j \mathbf{1}_{B_j}(\omega),$$

como ambas colecciones de conjuntos son particiones de Ω tenemos que

$$\mu(A_i) = \sum_{j=1}^{m} \mu(A_i \cap B_j) \qquad \text{y} \qquad \mu(B_j) = \sum_{i=1}^{n} \mu(A_i \cap B_j). \tag{3.1}$$

Además, si $A_i \cap B_j$ no es vacío, tiene al menos un elemento ω y $X(\omega) = c_i = d_j$. Por lo tanto

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

Si tenemos dos funciones simples no-negativas

$$X(\omega) = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}(\omega), \qquad Y(\omega) = \sum_{j=1}^{m} d_j \mathbf{1}_{B_j}(\omega),$$

podemos usar las representaciones

$$X = \sum_{i=1}^{n}\sum_{j=1}^{m} c_i \mathbf{1}_{A_i\cap B_j}(\omega), \qquad Y = \sum_{j=1}^{m}\sum_{i=1}^{n} d_j \mathbf{1}_{A_i\cap B_j}(\omega),$$

en términos de la partición $A_i \cap B_j$. Ahora la función simple $X+Y$ tiene la representación

$$X + Y = \sum_{i=1}^{n}\sum_{j=1}^{m}(c_i + d_j)\mathbf{1}_{A_i\cap B_j}$$

y

$$\begin{aligned}
\int (X+Y)\,d\mu &= \sum_{i=1}^{n}\sum_{j=1}^{m}(c_i + d_j)\mu(A_i\cap B_j)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} c_i\mu(A_i\cap B_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} d_j\mu(A_i\cap B_j)\\
&= \sum_{i=1}^{n} c_i\mu(A_i) + \sum_{j=1}^{m} d_j\mu(B_j)\\
&= \int X\,d\mu + \int Y\,d\mu
\end{aligned}$$

de modo que la integral de funciones simples no-negativas es aditiva. Es inmediato además que si $\alpha \geq 0$, $\beta \geq 0$ y X, Y son funciones simples no-negativas entonces

$$\int (\alpha X + \beta Y)\,d\mu = \alpha\int X\,d\mu + \beta\int Y\,d\mu$$

de modo que la integración es lineal en la clase de las funciones simples no-negativas. También es fácil ver que si X, Y son funciones simples y $X \geq Y$ entonces $\int X\,d\mu \geq \int Y\,d\mu$.

3.4.2. Funciones Medibles No-Negativas

Dada una función medible no-negativa $X : \Omega \to \mathbb{R}^+$, sabemos que existe una sucesión creciente de funciones simples no-negativas X_n tales que $X_n \uparrow X$. La integral $\int X_n\,d\mu$ está definida para todo n y es creciente, por lo tanto tiene un límite, que puede ser $+\infty$. Definimos

$$\int X\,d\mu = \lim_{n\to\infty}\int X_n d\mu.$$

Tenemos que verificar que esta definición no depende de la sucesión (X_n) de funciones simples aproximantes.

Proposición 3.10 *Sea (X_n) una sucesión creciente de funciones simples no-negativas y $X = \lim_n X_n \geq Y$ donde Y es simple y no-negativa. Entonces*

$$\lim_{n\to\infty}\int X_n d\mu \geq \int Y\,d\mu. \tag{3.2}$$

Demostración. Sea $Y = \sum_{i=1}^{k} c_i\mathbf{1}_{E_i}$. Si $\int Y\,d\mu = \infty$, para algún entero i, $1 \leq i \leq k$, tal que $c_i > 0$, $\mu(E_i) = +\infty$. Entonces para cualquier ε fijo con $0 < \varepsilon < c_i$, definimos los conjuntos

$$A_n = \{\omega : X_n(\omega) + \varepsilon > Y(\omega)\} \tag{3.3}$$

La sucesión de conjuntos $A_n \cap E_i,\ n \geq 1$, es monótona creciente y converge a E_i, y por lo tanto $\mu(A_n \cap E_i) \to \infty$ cuando $n \to \infty$. Por otro lado

$$\int X_n \, d\mu \geq \int X_n \mathbf{1}_{A_n \cap E_i} d\mu \geq (c_i - \varepsilon)\mu(A_n \cap E_i) \to \infty \quad (n \to \infty).$$

Por lo tanto (3.2) vale si $\int Y \, d\mu = \infty$. Supongamos ahora que esta integral es convergente y sea

$$A = \{\omega : Y(\omega) > 0\} = \cup_{i:c_i>0} E_i.$$

Como Y es simple, $c = \text{mín}_{c_i>0} c_i > 0$ y $\mu(A) < \infty$. Supongamos que $\varepsilon > 0$ y definimos A_n de nuevo por (3.3). Entonces

$$\begin{aligned} \int X_n d\mu &\geq \int X_n \mathbf{1}_{A_n \cap A} d\mu \geq \int (Y - \varepsilon) \mathbf{1}_{A_n \cap A} d\mu \\ &= \int Y \mathbf{1}_{A_n \cap A} d\mu - \varepsilon\mu(A_n \cap A) \geq \int Y \mathbf{1}_{A_n} d\mu - \varepsilon\mu(A). \end{aligned}$$

Como $\mu(A_n \cap E_i) \to \mu(E_i)$ para cada i, podemos evaluar las integrales como sumas finitas y encontrar un entero $n_0 = n_0(\varepsilon)$ tal que

$$\int X_n \, d\mu \geq \int Y d\mu - \varepsilon - \varepsilon\mu(A) \quad \text{para } n \geq n_0,$$

y hemos establecido (3.2) también en el caso $\int Y \, d\mu < \infty$. ■

Veamos ahora que la definición de la integral no depende de la sucesión aproximante. Supongamos que tenemos dos sucesiones crecientes de funciones simples (X_n) y (Y_m) que convergen a la función X. Para cada m fijo tenemos

$$X = \lim_n X_n \geq Y_m,$$

y por la proposición anterior,

$$\lim_n \int X_n d\mu \geq \int Y_m \, d\mu.$$

Haciendo ahora $m \to \infty$,

$$\lim_n \int X_n \, d\mu \geq \lim_m \int Y_m \, d\mu.$$

Un argumento similar demuestra que la desigualdad en sentido contrario también vale y por lo tanto

$$\lim_n \int X_n \, d\mu = \lim_m \int Y_m \, d\mu.$$

Por lo tanto la integral está bien definida para funciones medibles no-negativas. Como consecuencia de la linealidad de la integral para funciones simples, tenemos que si $\alpha \geq 0,\ \beta \geq 0$ entonces

$$\int (\alpha X + \beta Y) \, d\mu = \alpha \int X \, d\mu + \beta \int Y \, d\mu.$$

De acuerdo a nuestra definición, si $X \geq 0$ es medible, $\int X \, d\mu$ puede ser finita o $+\infty$. Decimos que una función medible $X \geq 0$ es integrable con respecto a la medida μ si $\int X \, d\mu < \infty$.

Hay dos razones por las cuales una función de este tipo puede no ser integrable. O bien existe una función simple $Y \leq X$ para la cual $\int Y \, d\mu = \infty$, lo que implica la existencia de un $c > 0$ para el cual $\mu\{\omega : X(\omega) > c\} = +\infty$, o bien es posible que $\int Y \, d\mu < \infty$ para todas las funciones simples tales que $Y \leq X$ (lo que implica que $\mu\{\omega : X(\omega) > c\} < \infty$, para todo $c > 0$) pero, para cualquier sucesión Y_n de funciones simples que converja a X, $\int Y_n \, d\mu \to +\infty$ cuando $n \to \infty$.

3.4.3. Funciones Medibles Integrables.

Sabemos que si $X : \Omega \to \mathbb{R}^*$ es medible, también lo son X^+, X^- y tenemos $X = X^+ - X^-$. Si ambas funciones X^+ y X^- son integrables, decimos que X es integrable y definimos

$$\int X\, d\mu = \int X^+\, d\mu - \int X^-\, d\mu.$$

Usamos la notación $L^1(\Omega, \mathcal{F}, \mu)$ o $L^1(\mu)$ para la clase de funciones integrables y la notación

$$\int X(\omega)\, \mu(d\omega) = \int X(\omega)\, d\mu(\omega) = \int X\, d\mu$$

para la integral.

Si $(\Omega, \mathcal{F}, P)$ es un espacio de probabilidad y $X : \Omega \to \mathbb{R}$ es una variable aleatoria integrable llamamos *esperanza* a su integral y escribimos

$$\mathrm{E}(X) = \int X(\omega)\, dP(\omega).$$

3.4.4. Varianza y Covarianza

Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria y supongamos que $X^2 \in L^1$, que escribimos como $X \in L^2$. Definimos la varianza de X como

$$\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2$$

y también usamos la notación σ_X^2 para ella. Por la linealidad de la esperanza tenemos

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$$

La raíz cuadrada de la varianza $\sigma_X = \sqrt{\mathrm{Var}(X)}$ se conoce como la *desviación típica* de la variable aleatoria.

Si $X, Y \in L^2$ definimos

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))],$$

y de nuevo por linealidad

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y).$$

Observamos que si $X = Y$, entonces $\mathrm{Cov}(X, Y) = \mathrm{Var}(X)$.

La covarianza es una medida de la dependencia lineal entre dos variables aleatorias, pero su valor depende de las unidades en las cuales expresemos las variables. Para obtener una medida normalizada de la dependencia lineal entre dos variables, dividimos la covarianza entre las desviaciones típicas de las variables. Este parámetro se conoce como la *correlación* entre X e Y:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Si $\mathrm{Cov}(X, Y) = 0$ decimos que X e Y no están correlacionadas. Demostraremos más adelante que $|\rho(X, Y)| \leq 1$.

La covarianza es una función bilineal: Si $X_1, \dots, X_k$ y $Y_1, \dots, Y_m$ son variables aleatorias en L^2, entonces para cualesquiera constantes $a_1, \dots, a_k$ y $b_1, \dots, b_m$

$$\mathrm{Cov}\Big(\sum_{i=1}^k a_i X_i, \sum_{j=1}^m b_j Y_j\Big) = \sum_{i=1}^k \sum_{j=1}^m a_i b_j\, \mathrm{Cov}(X_i, Y_j). \tag{3.4}$$

Para ver esto, podemos suponer, sin pérdida de generalidad, que $\mathrm{E}(X_i) = \mathrm{E}(Y_j) = 0$ para $i = 1, \dots, k;\ j = 1, \dots, m$. Entonces

$$\begin{aligned}
\mathrm{Cov}\Big(\sum_{i=1}^{k} a_i X_i, \sum_{j=1}^{m} b_j Y_j\Big) &= \mathrm{E}\Big(\sum_{i=1}^{k} a_i X_i \sum_{j=1}^{m} b_j Y_j\Big) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{m} a_i b_j \,\mathrm{E}(X_i Y_j) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{m} a_i b_j \,\mathrm{Cov}(X_i, Y_j)
\end{aligned}$$

Un caso especial de esta fórmula se usa para calcular la varianza de la suma de variables $X_1, \dots, X_n$ en L^2. Tenemos

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \mathrm{Cov}\Big(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\Big) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j).$$

Dividimos el conjunto de índices en los casos $i = j$ e $i \neq j$ y obtenemos

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) &= \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) + 2 \sum_{1\leq i<j\leq n}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1\leq i<j\leq n}^{n} \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

En particular si $\mathrm{Cov}(X_i, X_j) = 0$ para $i \neq j$, es decir, si las variables no están correlacionadas, entonces

$$\mathrm{Var}\Big(\sum_{i=1}^{n} (X_i)\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

3.4.5. Integral sobre un conjunto medible.

Si $A \in \mathcal{F}$ definimos

$$\int_A X\,d\mu = \int X\mathbf{1}_A\,d\mu$$

siempre que esta integral esté bien definida. Por lo tanto $\int_A X\,d\mu$ está definida si $X\mathbf{1}_A$ es medible y no-negativa o si $X\mathbf{1}_A$ es medible e integrable. Decimos que X es integrable en A si $X\mathbf{1}_A$ es integrable. Es inmediato que

$$\int_\Omega X\,d\mu = \int X\,d\mu.$$

Observamos que si $E \in \mathcal{F}$ con $\mu(E) = 0$, entonces cualquier función $X : \Omega \to \mathbb{R}^*$ es integrable sobre E y

$$\int_E X\,d\mu = 0.$$

3.5. Propiedades de la Integral

Teorema 3.2 *Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida, A, B son conjuntos en $\mathcal{F}$ y $X : \Omega \to \mathbb{R}^*$, $Y : \Omega \to \mathbb{R}^*$ son dos funciones integrables con respecto a μ. Entonces X es integrable sobre A, $X + Y$ y $|X|$ son integrables y*

1. *Si $A \cap B = \emptyset$, $\int_{A\cup B} X\,d\mu = \int_A X\,d\mu + \int_B X\,d\mu$.*

2. *X es finita c.s.*
3. $\int (X+Y)\, d\mu = \int X\, d\mu + \int Y\, d\mu$.
4. $\left| \int X\, d\mu \right| \leq \int |X|\, d\mu$.
5. *Para cualquier $c \in \mathbb{R}$, cX es integrable y $\int cX\, d\mu = c \int X\, d\mu$.*
6. $X \geq 0 \Rightarrow \int X\, d\mu \geq 0$; $X \geq Y \Rightarrow \int X\, d\mu \geq \int Y\, d\mu$.
7. *Si $X \geq 0$ y $A \subset B$, $\int_A X\, d\mu \leq \int_B X\, d\mu$.*
8. *Si $X \geq 0$ y $\int X\, d\mu = 0$, entonces $X = 0$ c.s.*
9. $X = Y$ *c.s.* $\Rightarrow \int X\, d\mu = \int Y\, d\mu$.
10. *Si $Z : \Omega \to \mathbb{R}^*$ es $\mathcal{F}$-medible y $|Z| \leq X$, entonces h es integrable.*

Demostración.
Si $X : \Omega \to \mathbb{R}^+$ es medible, no-negativa e integrable y $0 \leq Y \leq X$ con $Y : \Omega \to \mathbb{R}^+$ medible, sigue de la definición de la integral de funciones medibles no-negativas que Y es integrable. Como para cualquier $A \in \mathcal{F}$, $\mathbf{1}_A$ es medible,

$$0 \leq X^+ \mathbf{1}_A \leq X^+ \qquad \text{y} \qquad 0 \leq X^- \mathbf{1}_A \leq X^-$$

y en consecuencia si X es integrable sobre Ω también lo es sobre cualquier conjunto medible A.
(1) Si A, B son disjuntos, $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B$, y por lo tanto

$$X^+ \mathbf{1}_{A\cup B} = X^+ \mathbf{1}_A + X^+ \mathbf{1}_B, \qquad X^- \mathbf{1}_{A\cup B} = X^- \mathbf{1}_A + X^- \mathbf{1}_B$$

y como sabemos que la propiedad que queremos demostrar es válida para funciones medibles no-negativas tenemos

$$\begin{aligned}\int_{A\cup B} X\, d\mu &= \int X^+ \mathbf{1}_{A\cup B} d\mu - \int X^- \mathbf{1}_{A\cup B} d\mu \\ &= \int X^+ \mathbf{1}_A d\mu - \int X^- \mathbf{1}_A d\mu + \int X^+ \mathbf{1}_B d\mu - \int X^- \mathbf{1}_B d\mu \\ &= \int_A X\, d\mu + \int_B X\, d\mu\end{aligned}$$

ya que todos los términos son finitos.
(2) Si X no es finita c.s., entonces al menos uno de los conjuntos

$$A_1 = \{\omega : X(\omega) = +\infty\}, \quad A_2 = \{\omega : X(\omega) = -\infty\}$$

tiene medida positiva. Supongamos que $\mu(A_1) > 0$ entonces se tiene de la definición que $\int X^+\, d\mu = +\infty$ lo que implica que X no es integrable.
(3) Ya hemos visto que esta propiedad vale para funciones no-negativas. Si $X_1,\ X_2$ son no-negativas y $X = X_1 - X_2$ entonces $X_1 + X^- = X_2 + X^+$ y usando (3) para funciones no-negativas obtenemos

$$\int X_1\, d\mu + \int X^-\, d\mu = \int X_2\, d\mu + \int X^+\, d\mu$$

de modo que

$$\int X\, d\mu = \int X_1\, d\mu - \int X_2\, d\mu.$$

El resultado general se obtiene observando que para X, Y finitas,

$$X + Y = (X^+ + Y^+) - (X^- + Y^-)$$

y en consecuencia

$$\begin{aligned}\int (X+Y)\,d\mu &= \int (X^+ + Y^+)\,d\mu - \int (X^- + Y^-)\,d\mu\\ &= \int X^+\,d\mu - \int X^-\,d\mu + \int Y^+\,d\mu - \int Y^-\,d\mu\\ &= \int X\,d\mu + \int Y\,d\mu\end{aligned}$$

Finalmente usamos (3) con la función $|X| = X^+ - X^-$ para deducir que $|X|$ es integrable y

$$\int |X|\,d\mu = \int X^+\,d\mu + \int X^-\,d\mu.$$

(4) Tenemos

$$\Big|\int X\,d\mu\Big| = \Big|\int X^+\,d\mu - \int X^-\,d\mu\Big| \leq \int X^+\,d\mu + \int X^-\,d\mu = \int |X|\,d\mu.$$

(5) Si $c = 0$, $cX = 0$ y $\int cX\,d\mu = 0 = c\int X\,d\mu$. Si $c > 0$ entonces

$$(cX)^+ = cX^+, \qquad (cX)^- = cX^-,$$

y el resultado es válido porque ya lo demostramos para funciones medibles no-negativas. De manera similar, si $c < 0$

$$(cX)^+ = -cX^-, \qquad (cX)^- = -cX^+,$$

$$\int cX\,d\mu = \int (cX)^+\,d\mu - \int (cX)^-\,d\mu = (-c)\int X^-\,d\mu + c\int X^+\,d\mu = c\int X\,d\mu.$$

(6) La primera afirmación es cierta por la definición. Si $X \geq Y$, entonces $X = Y + (X-Y)$ y $(X-Y) \geq 0$. Por (3) tenemos

$$\int X\,d\mu = \int Y\,d\mu + \int (X-Y)\,d\mu \geq \int Y\,d\mu.$$

(7) es consecuencia de (6) ya que $X\mathbf{1}_A \leq X\mathbf{1}_B$.
(8) Si$\{\omega : X(\omega) > 0\}$ tiene medida positiva, por la continuidad de la medida μ existe un entero n tal que, si $A = \{\omega : X(\omega) > 1/n\}$, se tiene que $\mu(A) > 0$. Pero $n^{-1}\mathbf{1}_A \leq X\mathbf{1}_A \leq X$, de modo que

$$\int X\,d\mu \geq \frac{1}{n}\int \mathbf{1}_A\,d\mu = \frac{1}{n}\mu(A) > 0.$$

Por lo tanto, si $X \geq 0$ y $\int X\,d\mu = 0$, necesariamente $\mu\{\omega : X(\omega) > 0\} = 0$.
(9) Si $X = Y$ c.s. entonces $X^+ = Y^+$, $X^- = Y^-$ c.s. En la construcción de la sucesión aproximante en el teorema 3.1, los conjuntos $Q_{p,s}$ para las funciones X^+ y Y^+ tendrán todos la misma medida. Por lo tanto hay funciones simples $X_n \to X^+$, $Y_n \to Y^+$ tales que

$$\int X_n\,d\mu = \int Y_n\,d\mu, \qquad n = 1, 2, \ldots$$

y por lo tanto se tiene que $\int X^+\,d\mu = \int Y^+\,d\mu$. De manera similar $\int X^-\,d\mu = \int Y^-\,d\mu$.
(10) Si $|Z| \leq X$ entonces $0 \leq Z^+ \leq X$, $0 \leq Z^- \leq X$. Usando (6) tenemos que tanto $\int Z^+\,d\mu$ como $\int Z^-\,d\mu$ son finitos y por lo tanto Z es integrable. ■

Ejemplo 3.6
Sea $\Omega = [0,1]$ y λ la medida de Lebesgue. Sea $X(\omega) = \mathbf{1}_{\mathbb{Q}\cap[0,1]}$. Sabemos que $\lambda(\mathbb{Q}) = 0$ de modo que

$$\lambda(X = 0) = 1 = \lambda([0,1] - \mathbb{Q}).$$

Por lo tanto

$$\int X\, d\mu = \int 0\, d\mu = 0.$$

Corolario 3.3 *Si $X : \Omega \to \mathbb{R}^*$ es acotada, medible y se anula fuera de un conjunto $E \in \mathcal{F}$ con $\mu(E) < \infty$, entonces X es integrable con respecto a μ.*

Demostración. Si $|X| \leq K$, entonces la función simple $K\mathbf{1}_E$ es integrable y la integrabilidad de X sigue ahora de (9). ■

Observación 3.3 Si $\mathcal{F}$ es completa respecto a μ, entonces podemos modificar (9) de la siguiente manera: Si $X : \Omega \to \mathbb{R}^*$ es integrable y $Y : \Omega \to \mathbb{R}^*$ es tal que $X = Y$ c.s., entonces Y es integrable y $\int X\, d\mu = \int Y\, d\mu$.

Hay un recíproco de la afirmación anterior: Si X y Y son integrables y

$$\int_E X\, d\mu = \int_E Y\, d\mu \quad \text{para todo } E \in \mathcal{F},$$

entonces $X = Y$ c.s. Para ver esto supongamos que es falso, de modo que $\mu\{x : X(\omega) \neq Y(\omega)\} > 0$. Entonces al menos uno de los conjuntos $\{x : X(\omega) > Y(\omega)\}$, $\{x : X(\omega) < Y(\omega)\}$ tiene medida positiva. Supongamos que es el primero, por la continuidad de la medida existe un entero n tal que

$$E_n = \{\omega : X(\omega) \geq Y(\omega) + \frac{1}{n}\}, \quad \mu(E_n) > 0.$$

Pero entonces

$$\int_{E_n} X\, d\mu - \int_{E_n} Y\, d\mu > \frac{1}{n}\mu(E_n) > 0$$

lo que es una contradicción y por lo tanto demuestra el resultado.

3.5.1. Desigualdades de Markov y Chebychev

Proposición 3.11 (Desigualdad de Markov) *Sea $X \in L^1(\Omega, \mathcal{F}, P)$ una variable aleatoria. Para cualquier $\lambda > 0$,*

$$P(|X| \geq \lambda) \leq \mathrm{E}(|X|)/\lambda.$$

Demostración. Observamos que

$$\mathbf{1}_{\{\frac{|X|}{\lambda} \geq 1\}} \leq \frac{|X|}{\lambda}\mathbf{1}_{\{\frac{|X|}{\lambda} \geq 1\}} \leq \frac{|X|}{\lambda}.$$

Tomando esperanza se obtiene el resultado. ■

Corolario 3.4 (Desigualdad de Chebychev) *Si $X \in L^2(\Omega)$ es una variable aleatoria,*

$$P(|X - \mathrm{E}(X)| \geq \lambda) \leq \mathrm{Var}(X)/\lambda^2.$$

Demostración. Usando la desigualdad de Markov

$$P(|X - \mathrm{E}(X)| \geq \lambda) = P(|X - \mathrm{E}(X)|^2 \geq \lambda^2) \leq \mathrm{E}(X - \mathrm{E}(X))^2/\lambda^2 = \mathrm{Var}(X)/\lambda^2.$$

■

3.6. Las Integrales de Lebesgue y de Lebesgue-Stieltjes

Hemos definido la integral sobre un espacio de medida abstracto $(\Omega, \mathcal{F}, \mu)$ aunque históricamente se consideró primero el espacio $(\mathbb{R}, \mathcal{M}, \lambda)$ donde λ denota la medida de Lebesgue definida sobre los conjuntos Lebesgue-medibles $\mathcal{M}$.

Si $E \in \mathcal{M}$ y f es Lebesgue-medible es usual escribir $\int_E f(x)\,dx$ en lugar de $\int_E f\,d\lambda$. En particular si E es un intervalo con extremos a, b, escribimos $\int_a^b f(x)\,dx$. Observamos que como la medida de Lebesgue de un punto es cero, no importa si incluimos los extremos en el intervalo o no. En particular, a puede ser $-\infty$ y b puede ser $+\infty$ de modo que $\int_{-\infty}^{\infty} f(x)\,dx$ quiere decir $\int_{\mathbb{R}} f\,d\lambda = \int f\,d\lambda$.

Es importante observar que la integral sobre un intervalo infinito se define directamente, ya que un intervalo de este tipo es medible, y no como límite de integrales sobre intervalo finitos.

Usar la medida de Lebesgue-Stieltjes que se obtiene a partir de una función de Stieltjes F en lugar de usar la medida de Lebesgue es equivalente a trabajar en el espacio de medida $(\mathbb{R}, \mathcal{M}(F), \mu_F)$, donde μ_F es la medida de Lebesgue-Stieltjes que definimos en el capítulo anterior y $\mathcal{M}(F)$ es la σ-álgebra de los conjuntos medibles respecto a μ_F^*. Usamos las notaciones

$$\int_E f(x)\,dF(x) \qquad \text{o} \qquad \int_E f(x)\,d\mu_F(x).$$

En este caso la medida de un punto puede ser distinta de cero y por lo tanto al integrar sobre un intervalo es necesario especificar si se incluyen los extremos o no. Por esto no usamos la notación $\int_a^b f(x)dF(x)$ a menos que sepamos que F es continua.

Como las σ-álgebras de conjuntos medibles son completas con respecto a las medidas de Lebesgue o de Lebesgue-Stieltjes, vemos que si $f : \mathbb{R} \to \mathbb{R}^*$ es integrable y $f = g$ c.s., entonces $g : \mathbb{R} \to \mathbb{R}^*$ también es integrable.

Caso Particular

Llamamos δ_x a la medida de Dirac en el punto x:

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

δ es una medida de probabilidad definida en $\mathcal{P}(\mathbb{R})$. Si $x_n,\ n \geq 1$ es una sucesión de números reales y p_n, $n \geq 1$ es una sucesión de números reales positivos entonces

$$\mu(A) = \sum_{n=1}^{\infty} p_n\,\delta_{x_n}(A)$$

define una medida de Lebesgue-Stieltjes concentrada en el conjunto numerable $\{x_n,\ n \geq 1\}$. Si $\sum p_n < \infty$, la medida es finita y si $\sum p_n = 1$ es una medida de probabilidad.

Si $f : \mathbb{R} \to \mathbb{R}$ es una función cualquiera entonces

$$\int f\,d\mu = \sum_{n=1}^{\infty} p_n f(x_n).$$

En el caso particular $p_n = 1$ para todo n y $f(x) = x$, obtenemos la serie $\sum x_n$, de modo que la teoría de series absolutamente convergentes de números reales es un caso particular de la integral de Riemann-Stieltjes.

3.7. Integrales y Límites

Ahora podemos considerar los teoremas sobre la continuidad del operador de integración.

Teorema 3.3 (Teorema de Convergencia Monótona) *Sea $X_n : \Omega \to \mathbb{R}^+$, $n \geq 1$ una sucesión creciente de funciones medibles no-negativas con $X_n(\omega) \to X(\omega)$ para todo $\omega \in \Omega$, entonces*

$$\lim_{n\to\infty} \int X_n \, d\mu = \int X \, d\mu,$$

es decir, si X es integrable, las integrales $\int X_n \, d\mu$ converge a $\int X \, d\mu$, mientras que si X no es integrable o bien X_n es integrable para todo n y $\int X_n \, d\mu \to +\infty$ cuando $n \to \infty$, o existe un entero N tal que X_N no es integrable de modo que $\int X_n \, d\mu = +\infty$ para $n \geq N$.

Demostración. Para cada $n = 1, 2, \ldots$ escogemos una sucesión creciente $X_{n,k}$, $k \geq 1$ de funciones simples no-negativas que converge a X_n y definimos $Y_k = \max_{n\leq k} X_{n,k}$. Entonces (Y_k) es una sucesión no-decreciente de funciones simples no-negativas y $Y = \lim_k Y_k$ es una función medible no-negativa. Pero

$$X_{n,k} \leq Y_k \leq X_k \leq X \quad \text{para } n \leq k \tag{3.5}$$

de modo que $X_n \leq Y \leq X$, y si hacemos $n \to \infty$ vemos que $X = Y$. Usando la propiedad (6) del teorema 3.2 y (3.5) obtenemos

$$\int X_{n,k} d\mu \leq \int Y_k \, d\mu \leq \int X_k \, d\mu \quad \text{para } n \leq k.$$

Para n fijo, hacemos $k \to \infty$. Por la definición de la integral

$$\int X_n d\mu \leq \int Y \, d\mu \leq \lim_{k\to\infty} \int X_k \, d\mu.$$

Haciendo ahora $n \to \infty$ obtenemos

$$\lim_{n\to\infty} \int X_n \, d\mu \leq \int Y \, d\mu \leq \lim_{k\to\infty} \int X_k \, d\mu.$$

Como los dos extremos de la desigualdad son iguales, tenemos

$$\lim_{n\to\infty} \int X_n \, d\mu = \int Y \, d\mu = \int X \, d\mu.$$

■

Corolario 3.5 *Sea $X_n : \Omega \to \mathbb{R}$, $n \geq 1$ una sucesión de funciones medibles tales que $X_n \uparrow X$ y $\int X_1 \, d\mu > -\infty$. Entonces*

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

Demostración. Supongamos que $X \leq 0$. Sea $Y_n = -X_n \downarrow -X = Y$. Entonces para todo n,

$$0 \leq \int Y \, d\mu \leq \int Y_n \, d\mu < +\infty.$$

Pero $0 \leq Y_1 - Y_n \uparrow Y_1 - Y$, y por el teorema anterior

$$\int (Y_1 - Y_n) \, d\mu \uparrow \int (Y_1 - Y) \, d\mu < +\infty.$$

Como estas integrales son finitas, podemos restarlas de $\int Y_1 \, d\mu$ y obtenemos

$$\int Y_n \, d\mu \downarrow \int Y \, d\mu \quad \Rightarrow \quad \int X_n \, d\mu \uparrow \int X \, d\mu.$$

En el caso general tenemos $X_n^+ \uparrow X^+$ y $X_n^- \downarrow X^-$ con $\int X^- \, d\mu < +\infty$, y por lo anterior tenemos

$$\int X_n^+ \, d\mu \uparrow \int X^+ \, d\mu, \qquad +\infty > \int X_n^- \, d\mu \downarrow \int X^- \, d\mu \geq 0.$$

En consecuencia

$$\int X_n \, d\mu \uparrow \int X \, d\mu$$

■

Corolario 3.6 *Si $X_n \geq 0, \ n \geq 1$ son funciones medibles no-negativas, entonces*

$$\int \big(\sum_{n=1}^{\infty} X_n\big) \, d\mu = \sum_{n=1} \int X_n \, d\mu$$

Demostración. Tenemos

$$\begin{aligned} \int \big(\sum_{n=1}^{\infty} X_n\big) \, d\mu &= \int \big(\lim_{k\to\infty} \sum_{n=1}^{k} X_n\big) \, d\mu = \lim_{k\to\infty} \int \big(\sum_{n=1}^{k} X_n\big) \, d\mu \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \int X_n \, d\mu = \sum_{n=1}^{\infty} \int X_i \, d\mu \end{aligned}$$

■

Corolario 3.7 *Si X es integrable entonces, para $A \in \mathcal{F}$,*

$$\int_A X \, d\mu \to 0 \quad \textit{cuando } \mu(A) \to 0.$$

Demostración Definimos

$$X_n = \begin{cases} X & \text{si } |X| \leq n, \\ n & \text{si } |X| > n. \end{cases}$$

Entonces $|X_n|$ es creciente y converge a $|X_n|$ cuando $n \to \infty$. Por el teorema 3.2 $|X|$ es integrable y

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

cuando $n \to \infty$. Dado $\varepsilon > 0$ escogemos N tal que

$$\int |X| \, d\mu < \int |X_n| \, d\mu + \frac{1}{2}\varepsilon$$

para $n \geq N$. Entonces si $A \in \mathcal{F}$ es tal que $\mu(A) < \varepsilon/2N$, tenemos, por el teorema 3.2,

$$\begin{aligned} \Big| \int_A X \, d\mu \Big| &\leq \int_A |X| \, d\mu = \int_A |X_N| \, d\mu + \int_A (|X| - |X_N|) \, d\mu \\ &< \frac{1}{2}\varepsilon + \int (|X| - |X_N|) \, d\mu < \varepsilon. \end{aligned}$$

■

Teorema 3.4 (Fatou) *Si* (X_n) *es una sucesión de funciones medibles que está acotada inferiormente por una función integrable, entonces*

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

Demostración. Como (X_n) está acotada por debajo por una función integrable Y podemos suponer, sin pérdida de generalidad, que $X_n \geq 0$ para todo n. Para $Z_n = X_n - Y \geq 0$ c.s. y

$$\int Z_n \, d\mu = \int X_n \, d\mu - \int Y \, d\mu, \qquad \liminf Z_n = \liminf X_n - Y \quad \text{c.s.}$$

Ponemos $Y_n = \inf_{k\geq n} X_k$ entonces Y_n es una sucesión creciente de funciones medibles y

$$\lim_{n\to\infty} Y_n = \liminf_{n\to\infty} X_n.$$

Como $X_n \geq Y_n$ para todo n

$$\liminf_{n\to\infty} \int X_n \, d\mu \geq \lim_{n\to\infty} \int Y_n \, d\mu = \int \lim_{n\to\infty} Y_n \, d\mu = \int \liminf_{n\to\infty} X_n \, d\mu,$$

por el teorema 3.3. ■

Corolario 3.8 *Si* (X_n) *es una sucesión de funciones medibles que está acotada superiormente por una función integrable, entonces*

$$\int \limsup_{n\to\infty} X_n \, d\mu \geq \limsup_{n\to\infty} \int X_n \, d\mu.$$

Demostración. Esto se puede demostrar por un método similar al del teorema anterior, o puede deducirse del teorema anterior poniendo $Y_n = -X_n$. ■

Como consecuencia de los dos resultados anteriores, si (X_n) es una sucesión de funciones medibles acotada superior e inferiormente por funciones integrables tenemos

$$\int \liminf_{n\to\infty} X_n \, d\mu \leq \liminf_{n\to\infty} \int X_n \, d\mu \leq \limsup_{n\to\infty} \int X_n \, d\mu \leq \int \limsup_{n\to\infty} X_n \, d\mu.$$

Ejemplo 3.7
Consideremos el espacio de probabilidad $([0,1], \mathcal{B}, \lambda)$, donde λ es la medida de Lebesgue y definimos

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

Para todo $\omega \in [0,1]$ se tiene que $\mathbf{1}_{(0,1/n)}(\omega) \to 0$, de modo que $X_n \to 0$. Sin embargo

$$\mathrm{E}(X_n) = n^2 \frac{1}{n} = n \to \infty,$$

y por lo tanto

$$\mathrm{E}(\liminf_{n\to\infty} X_n) = 0 < \liminf_{n\to\infty} \mathrm{E}(X_n) = \infty$$

y

$$\mathrm{E}(\limsup_{n\to\infty} X_n) = 0 < \limsup_{n\to\infty} \mathrm{E}(X_n) = \infty.$$

Teorema 3.5 (de Convergencia Dominada, Lebesgue) *(i) Si $Y : \Omega \to \mathbb{R}^+$ es integrable, (X_n) es una sucesión de funciones medibles de Ω en $\mathbb{R}^*$ tal que $|X_n| \leq Y$ para $n \geq 1$, y $X_n \to X$ cuando $n \to \infty$, entonces X es integrable y*

$$\int X_n \, d\mu \to \int X \, d\mu \quad \text{cuando } n \to \infty.$$

(ii) Supongamos que $Y : \Omega \to \mathbb{R}^+$ es integrable, $-\infty \leq a < b \leq +\infty$, y para cada $t \in (a,b)$, X_t es medible de Ω en $\mathbb{R}^*$. Entonces si $|X_t| \leq Y$ para todo $t \in (a,b)$ y $X_t \to X$ cuando $t \to a^+$ o $t \to b^-$, entonces X es integrable y

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

Demostración. (i) Primero probamos un caso especial con $X_n \geq 0$ y $X_n \to 0$ cuando $n \to \infty$. En este caso podemos usar el teorema 3.4 y el corolario para obtener

$$\begin{aligned}\limsup \int X_n \, d\mu &\leq \int \limsup X_n \, d\mu = \int 0 \, d\mu = 0 \\ &= \int \liminf X_n \, d\mu \leq \liminf \int X_n \, d\mu \leq \limsup \int X_n \, d\mu.\end{aligned}$$

Por lo tanto el límite existe y vale 0.

En el caso general ponemos $Z_n = |X_n - X|$, entonces $0 \leq Z_n \leq 2Y$, $2Y$ es integrable y Z_n es medible con $Z_n \to 0$ cuando $n \to \infty$. Pero entonces

$$\Big| \int X_n \, d\mu - \int X \, d\mu \Big| \leq \int |X_n - X| \, d\mu \to 0 \quad \text{cuando } n \to \infty,$$

y X es integrable por el teorema 3.2.

(ii) Supongamos, por ejemplo, que $X_t \to X$ cuando $t \to a^+$, entonces podemos aplicar la parte anterior del teorema a $X_n = X_{t_n}$, donde (t_n) es una sucesión en (a,b) que converge a a. Como $X = \lim X_n$, tenemos

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

Pero el lado derecho es independiente de la sucesión (t_n), de modo que $\int X_t \, d\mu$ converge al límite $\int X \, d\mu$ cuando $t \to a, \ t \in (a,b)$. ■

Como consecuencia de los teoremas sobre límites tenemos las siguiente propiedades

Proposición 3.12 *1. Si $X \geq 0$ y $\{A_n, \ n \geq 1\}$ es una sucesión de eventos disjuntos*

$$\int_{\cup_n A_n} X \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} X \, d\mu. \tag{3.6}$$

2. Si $X \in L^1$ y $\{A_n, \ n \geq 1\}$ es una sucesión monótona de eventos con $A_n \to A$ entonces

$$\int_{A_n} X \, d\mu \to \int_A X \, d\mu \tag{3.7}$$

Demostración. Para ver (1) tenemos

$$\begin{aligned}\int_{\cup_n A_n} X \, d\mu &= \int X \mathbf{1}_{\cup_n A_n} \, d\mu = \int \sum_{n}^{\infty} X \mathbf{1}_{A_n} \, d\mu \\ &= \sum_{n}^{\infty} \int X \mathbf{1}_{A_n} \, d\mu = \sum_{n}^{\infty} \int_{A_n} X \, d\mu.\end{aligned}$$

(2) es consecuencia directa del Teorema de Convergencia Monótona. ■

Teorema 3.6 *a) Si $X, Y \in L^2(\Omega, \mathcal{F}, P)$ entonces $X \cdot Y \in L^1(\Omega, \mathcal{F}, P)$ y*

$$|\operatorname{E}(XY)| \le \big(\operatorname{E}(X^2)\operatorname{E}(Y^2)\big)^{1/2} \tag{3.8}$$

(b) $L^2 \subset L^1$ y si $X \in L^2$ entonces $(\operatorname{E}(X))^2 \le \operatorname{E}(X^2)$.
(c) L^2 es un espacio lineal: $X, Y \in L^2$, $\alpha, \beta \in \mathbb{R}$ entonces $\alpha X + \beta Y \in L^2$.

La desigualdad (3.8) se conoce como la desigualdad de Cauchy-Schwarz.

Demostración. (a) Tenemos $|XY| \le X^2 + Y^2$ de modo que $X, Y \in L^2 \Rightarrow X \cdot Y \in L^1$. Para $x \in \mathbb{R}$ tenemos

$$0 \le \operatorname{E}[(\alpha X + Y)^2] = \alpha^2 \operatorname{E}[X^2] + 2\alpha \operatorname{E}[XY] + \operatorname{E}[Y^2] \tag{3.9}$$

el discriminante de la ecuación cuadrática en α es

$$\big(4[(\operatorname{E}(XY))^2 - \operatorname{E}(X^2)\operatorname{E}(Y^2)]\big)^{1/2}$$

y como la ecuación siempre es no-negativa,

$$(\operatorname{E}(XY))^2 - \operatorname{E}(X^2)\operatorname{E}(Y^2) \le 0$$

(b) Sea $X \in L^2$, como $X = X \cdot 1$ y $1 \in L^2$ con $\operatorname{E}(1^2) = 1$ se obtiene el resultado por (a).
(c) Sean $X, Y \in L^2$. Para α, β constantes

$$(\alpha X + \beta Y)^2 \le 2\alpha^2 X^2 + 2\beta^2 Y^2 \in L^1$$

y por lo tanto $\alpha X + \beta Y \in L^2$ y L^2 es un espacio vectorial. ■

3.8. Cambio de Variables

Teorema 3.7 (Cambio de Variable) *Sea $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ una función medible y sea μ una medida en $\mathcal{F}$. Sea $\mu_X = \mu \circ X^{-1}$ la medida inducida por X en $(\Omega', \mathcal{F}')$, $f : (\Omega', \mathcal{F}') \to (\mathbb{R}^*, \mathcal{B}^*)$ y $A \in \mathcal{F}'$. Entonces*

$$\int_{X^{-1}(A)} f(X(\omega))\, d\mu(\omega) = \int_A f(\omega')\, d\mu_X(\omega'). \tag{3.10}$$

Esto quiere decir que si alguna de las integrales existe, también existe la otra y ambas son iguales.

Demostración. Si f es la función indicadora de un conjunto B, la ecuación (3.10) dice que

$$\mu(X^{-1}(A) \cap X^{-1}(B)) = \mu_X(A \cap B)$$

que es cierta por la definición de μ_X. Si f es una función simple no-negativa: $f = \sum_1^n c_i \mathbf{1}_{B_i}$, entonces

$$\begin{aligned}
\int_{X^{-1}(A)} f(X(\omega))\, d\mu(\omega) &= \sum_{i=1}^n c_i \int_{X^{-1}(A)} \mathbf{1}_{B_i}(X(\omega))\, d\mu(\omega) \\
&= \sum_{i=1}^n c_i \int_A \mathbf{1}_{B_i}(\omega')\, d\mu_X(\omega') \\
&= \int_A f(\omega')\, d\mu_X(\omega')
\end{aligned}$$

Si f es una función medible no-negativa, sea $f_1, f_2, \dots$ una sucesión creciente de funciones simples no-negativas que converge a f. Entonces, por lo que hemos probado,

$$\int_{X^{-1}(A)} f_n(X(\omega))\, d\mu(\omega) = \int_A f_n(\omega')\, d\mu_X(\omega')$$

y usando el Teorema de Convergencia Monótona obtenemos el resultado.

Finalmente, si $f = f^+ - f^-$ es cualquier función medible, hemos probado que el resultado es válido para f^+ y f^-. Si, por ejemplo, $\int_{X^{-1}(A)} f^+(X(\omega))\,d\mu(\omega) < \infty$ entonces $\int_A f(\omega')\,d\mu_X(\omega') < \infty$, y en consecuencia si una de las integrales existe, también existe la otra y ambas valen lo mismo. ■

Sea X una v.a. sobre el espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Recordemos que la distribución de X es la medida $P_X = P \circ X^{-1}$ en $(\mathbb{R}, \mathcal{B})$ definida por $P_X(A) = P \circ X^{-1}(A) = P(X \in A)$, para $A \in \mathcal{B}$. La función de distribución de X es $F_X(x) = P_X((-\infty, x]) = P(X \leq x)$. El teorema de cambio de variables nos permite calcular la integral abstracta

$$\mathrm{E}(X) = \int_\Omega X\,dP$$

como

$$\mathrm{E}(X) = \int_{\mathbb{R}} x\,dP_X(x).$$

que es una integral en $\mathbb{R}$.

Corolario 3.9 *(i) Si X es una v.a. integrable entonces*

$$\mathrm{E}(X) = \int_{\mathbb{R}} x\,dP_X(x)$$

(ii) Sea $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ una variable aleatoria con distribución P_X y sea $g : (\Omega', \mathcal{F}') \to (\mathbb{R}^+, (\mathbb{R}^+))$ una función medible no-negativa. La esperanza de $g(X)$ es

$$\mathrm{E}(g(X)) = \int_\Omega g(X(\omega))\,dP(\omega) = \int_{\Omega'} g(\omega')\,dP_X(\omega').$$

Demostración. Para ver (i) basta observar que $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B})$, $f(x) = x$, y $A = \Omega'$. (ii) es inmediato a partir del teorema de cambio de variable. ■

3.8.1. Densidades

Sea $\mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}_k)$ un vector aleatorio con distribución $P_{\mathbf{X}}$. Decimos que $\mathbf{X}$, $P_{\mathbf{X}}$ o $F_{\mathbf{X}}$ es absolutamente continua si existe una función $f : (\mathbb{R}^k, \mathcal{B}_k) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ tal que

$$P_{\mathbf{X}}(A) = \int_A f(\mathbf{x})\,d\mathbf{x}$$

donde $d\mathbf{x}$ es la medida de Lebesgue.

Proposición 3.13 *Sea $g : (\mathbb{R}^k, \mathcal{B}_k) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ una función medible no-negativa. Sea $\mathbf{X}$ un vector aleatorio con f.d. F. Si F es absolutamente continua con densidad f, la esperanza de $g(\mathbf{X})$ es*

$$\mathrm{E}(g(\mathbf{X})) = \int_{\mathbb{R}^k} g(\mathbf{x}) f(\mathbf{x})\,d\mathbf{x}.$$

Demostración. Ejercicio.

3.9. Comparación con la Integral de Riemann

Vimos en el ejemplo 3.6 que si f es el indicador del conjunto de racionales en $[0,1]$ entonces f no es integrable según Riemann pero si lo es según Lebesgue y su integral vale 0. En este caso existe otra función g que satisface $f = g$ c.s., que es integrable según Riemann y cuya integral vale lo mismo que la de f. Esta función es $g(x) = 0$ para todo $x \in [0,1]$. Veremos a continuación un ejemplo de una función que es integrable según Lebesgue pero no según Riemann, a pesar de ser acotada y estar definida en un intervalo acotado, y para la cual no existe ninguna función Riemann integrable g con $f = g$ c.s.

Ejemplo 3.8
Comenzamos por construir un conjunto boreliano $A \subset (0,1]$ tal que $0 < \lambda(A) < 1$ y tal que para cualquier subintervalo J de $(0,1]$ se tiene que $\lambda(A \cap J) > 0$. Para ello sea $\{r_1, r_2, \dots\}$ una enumeración de los racionales en $(0,1)$. Sea $\varepsilon > 0$ dado y para cada n escogemos un intervalo $I_n = (a_n, b_n)$ tal que $r_n \in I_n \subset (0,1)$ y $\lambda(I_n) = b_n - a_n < \varepsilon 2^{-n}$. Ponemos $A = \cup_{n\geq 1} I_n$, entonces $0 < \lambda(A) < \varepsilon$.

Como A contiene a los racionales de $(0,1)$, es denso en este conjunto. Por lo tanto A es un conjunto abierto y denso con medida menor que ε. Si J es un subintervalo abierto de $(0,1)$ entonces J intersecta a alguno de los I_n y en consecuencia $\lambda(A \cap I) > 0$.

Si $B = (0,1) - A$ entonces $1 - \varepsilon < \lambda(B) < 1$. El conjunto B no contiene ningún intervalo, de hecho es un conjunto nunca denso (Todo intervalo contiene un subintervalo que no contiene puntos de B). Sin embargo, su medida es casi 1.

Tomamos ahora $f = \mathbf{1}_A$ y supongamos que $g = f$ *c.s.* y que J_n es una descomposición de $(0,1]$ en subintervalos. Para ver que g no es integrable según Riemann basta demostrar que J_n contiene puntos $x_n,\ y_n$ tales que

$$\sum_n g(x_n)\lambda(J_n) \leq \lambda(A) < 1 = \sum_n g(y_n)\lambda(J_n). \tag{3.11}$$

Si $\lambda(J_n - A) = 0$ escogemos x_n en el conjunto $J_n \cap \{f = g\}$, que es un conjunto de medida $\lambda(J_n) > 0$. Entonces $g(x_n) = f(x_n) \leq 1$. En el caso contrario, escogemos x_n en el conjunto $(J_n - A) \cap \{f = g\}$, un conjunto de medida $\lambda(J_n - A) > 0$. Entonces $g(x_n) = f(x_n) = 0$. Si $\sum'$ denota la suma sobre los índices n para los cuales $\lambda(J_n - A) = 0$ entonces la suma de la izquierda en (3.11) es

$$\sum{}' g(x_n)\lambda(J_n) \leq \sum{}' \lambda(J_n) = \sum{}' \lambda(J_n \cap A) \leq \lambda(A).$$

Para hallar los y_n observamos que por la construcción de A,

$$\lambda(J_n \cap A \cap \{f = g\}) = \lambda(J_n \cap A) > 0.$$

Por lo tanto para algún $y_n \in J_n \cap A$ se tiene $g(y_n) = f(y_n) = 1$ y de aqui se obtiene la segunda desigualdad de (3.11).

Usaremos la notación $\int_a^b f(x)\,dx$ para la integral de Riemann de f en $[a,b]$ y $\int_{[a,b]} f(x)\,d\lambda(x)$ para la integral de Lebesgue.

Para la demostración del próximo teorema vamos a usar la siguiente definición para la integral de Riemann, que no es la usual, pero es sencillo demostrar que ambas son equivalentes. Para cualquier entero n dividimos $I_0 = (a,b]$ en 2^n intervalos semiabiertos

$$I_{n,i} = (a_{n,i-1}, a_{n,i}], \quad i = 1, 2, \dots, 2^n.$$

Definimos

$$m_{n,i} = \text{ínf}\{f(x) : x \in I_{n,i}\}; \qquad M_{n,i} = \sup\{f(x) : x \in I_{n,i}\}$$

$$g_n(x) = \begin{cases} m_{n,i} & \text{si } x \in I_{n,i}, \\ 0 & \text{si } x \notin I_0; \end{cases} \qquad h_n(x) = \begin{cases} M_{n,i} & \text{si } x \in I_{n,i}, \\ 0 & \text{si } x \notin I_0. \end{cases}$$

Entonces para todo entero n y $x \in I_0$,

$$g_n(x) \leq f(x) \leq h_n(x).$$

(g_n) es una sucesión creciente de funciones simples mientras que (h_n) es decreciente. Si definimos

$$g = \lim_n g_n, \qquad h = \lim_n h_n,$$

entonces $g \leq f \leq h$. Además tenemos

$$\int_{(a,b]} g(x)\, d\lambda(x) = \lim_n \int_{(a,b]} g_n(x)\, d\lambda(x) = \lim_n \frac{b-a}{2^n} \sum_{i=1}^{2^n} m_{n,i} := \lim_n s_n,$$

$$\int_{(a,b]} h(x)\, d\lambda(x) = \lim_n \int_{(a,b]} h_n(x)\, d\lambda(x) = \lim_n \frac{b-a}{2^n} \sum_{i=1}^{2^n} M_{n,i} := \lim_n S_n.$$

Decimos que f es integrable según Riemann en $[a,b]$ si y sólo si

$$\lim_n s_n = \lim_n S_n$$

y en este caso el límite común es $\int_a^b f(x)\, dx$.

Teorema 3.8 *Una función acotada $f : [a,b] \to \mathbb{R}$ es integrable según Riemann si y sólo si el conjunto de puntos $E \subset [a,b]$ en los cuales X es discontinua tiene medida 0: $\lambda(E) = 0$. Cualquier función $f : [a,b] \to \mathbb{R}$ que sea integrable según Riemann, es integrable según Lebesgue y su integral tiene el mismo valor.*

Demostración.

Si f es continua en $x \in (a,b)$ entonces $g(x) = h(x)$. Recíprocamente si $g(x) = h(x)$ y x no está en D, donde D es el conjunto numerable de extremos de los intervalos $I_{n,i}$, entonces f es continua en x.

Si $\int_a^b f(x)\, dx$ existe,

$$\int_{[a,b]} g(x)\, d\lambda(x) = \int_a^b f(x)\, dx = \int_{[a,b]} h(x)\, d\lambda(x).$$

Por el teorema 3.2 (8) y teniendo en cuenta que $g \leq f \leq h$, tenemos que $g = h$ *c.s.* Como el conjunto E de puntos donde f es discontinua está contenido en $D \cup \{x : g(x) \neq h(x)\}$, tenemos que $\lambda(E) = 0$. Además, como la medida de Lebesgue es completa, f es $\mathcal{M}$-medible y por las propiedades de la integral de Lebesgue,

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

Recíprocamente si el conjunto E satisface $\lambda(E) = 0$, esto implica que $g(x) = h(x)$ c.s., y por las propiedades de la integral obtenemos

$$\int_{[a,b]} g(x)\, d\lambda(x) = \int_{[a,b]} h(x)\, d\lambda(x)$$

de modo que f es integrable. ■

El teorema anterior nos muestra que una función integrable según Riemann tiene que ser continua c.s., en cambio tenemos ejemplos de funciones que son discontinuas en todo punto y sin embargo son integrables según Lebesgue. Sin embargo, en cierto sentido estas últimas pueden ser aproximadas por funciones muy regulares, es decir por funciones que pueden ser diferenciadas infinitas veces.

Teorema 3.9 *Dada cualquier función $f : \mathbb{R} \to \mathbb{R}^*$ integrable y cualquier $\varepsilon > 0$, existe un intervalo finito (a, b) y una función acotada $g : \mathbb{R} \to \mathbb{R}$ tal que $g(x)$ se anula fuera de (a, b), es infinitamente diferenciable para todo real x y*

$$\int |f(x) - g(x)|\, d\lambda(x) < \varepsilon.$$

Demostración. Hacemos la demostración en cuatro etapas.

(i) Primero hallamos un intervalo finito $[a, b]$ y una función medible y acotada f_1 que se anule fuera de $[a, b]$ tal que

$$\int |f(x) - f_1(x)|\, d\lambda(x) < \frac{1}{4}\varepsilon.$$

Para hacer esto consideramos la sucesión de funciones

$$g_n(x) = \begin{cases} f(x) & \text{si } |x| \leq n \text{ y } |f(x)| \leq n, \\ n & \text{si } |x| \leq n \text{ y } f(x) > n, \\ -n & \text{si } |x| \leq n \text{ y } f(x) < -n, \\ 0 & \text{si } |x| > n. \end{cases}$$

Entonces $g_n(x) \to f(x)$ para todo x y $|g_n| \leq |f|$. Por el Teorema de Convergencia Dominada tenemos que

$$\int |f(x) - g_n(x)|\, d\lambda(x) \to 0 \quad \text{cuando } n \to \infty$$

de modo que podemos fijar N suficientemente grande y poner $f_1(x) = g_N(x)$.

(ii) El siguiente paso es aproximar f_1 por una función simple f_2 que se anule fuera de $[a, b]$ y satisfaga

$$\int |f_2(x) - f_1(x)|\, d\lambda(x) < \frac{1}{4}\varepsilon.$$

Esto es posible porque hemos definido la integral como límite de funciones simples.

(iii) Una función simple es la suma finita de multiplos de funciones indicadoras. Si cada función indicadora puede aproximarse por la función indicadora de un número finito de intervalos disjuntos, tenemos que f_2 puede aproximarse por f_3, una función escalera de la forma

$$f_3(x) = \sum_{i=1}^{n} c_i \mathbf{1}_{J_i}(x),$$

donde cada J_i es un intervalo finito y

$$\int |f_2(x) - f_3(x)|\, d\lambda(x) < \frac{1}{4}\varepsilon.$$

Para ver que esto es posible comenzamos con un conjunto acotado Lebesgue-medible E y $\eta > 0$. Hallamos un conjunto abierto G con $E \subset G$ y tal que $\lambda(G - E) < \eta/2$. A partir de la unión numerable de intervalos abiertos disjuntos que forman a G escogemos una cantidad finita que forman el conjunto G_0 tal que $\lambda(G - G_0) < \eta/2$. Entonces tenemos que $\lambda(E \triangle G_0) < \eta$ de modo que

$$\int_{\mathbb{R}} |\mathbf{1}_E(x) - \mathbf{1}_{G_0}(x)|\, d\lambda(x) < \eta.$$

(iv) Finalmente para obtener la función infinitamente diferenciable g tal que

$$\int |g(x) - f_3(x)|\, d\lambda(x) < \frac{1}{4}\varepsilon.$$

es suficiente encontrar una función de este tipo para una de las componentes $\mathbf{1}_{J_i}(x)$ de f_3. Supongamos que $J = (a, b)$ y $0 < 2\eta < b - a$. Ponemos

$$\phi_{a,\eta}(x) = \begin{cases} \exp\{-1/[(x-a)^2 - \eta^2]\} & \text{para } |x-a| < \eta, \\ 0 & \text{para } |x-a| \geq \eta. \end{cases}$$

Si $c_\eta^{-1} = \int_{-\infty}^{\infty} \phi_{a,\eta}(x)\,dx$, sea $h(x) = c_\eta \int_{-\infty}^{x} [\phi_{a,\eta}(t) - \phi_{b,\eta}(t)]\,dt$. Es fácil verificar que h es infinitamente diferenciable y

$$\int |\mathbf{1}_J(x) - h(x)|\,d\lambda(x) < 4\eta,$$

ya que $0 \leq h(x) \leq 1$ para todo x y $\{x : \mathbf{1}_J(x) \neq h(x)\}$ está contenido en los dos intervalos $(a-\eta, a+\eta)$ y $(b-\eta, b+\eta)$. ■

Hemos enunciado el resultado de aproximación para funciones reales, pero un teorema similar es cierto para funciones en $\mathbb{R}^k$, en cuyo caso la función aproximante tiene derivadas parciales de todos los órdenes.

3.10. Funciones de Distribución

Sea F una función de distribución. Como F es monótona sabemos que tiene a lo sumo una cantidad numerable de discontinuidades. Sea $\{a_j\}$ la colección de los puntos donde ocurren los saltos de la función F y sea $\{b_j\}$ el tamaño de los saltos respectivos, es decir

$$F(a_j) - F(a_j^-) = b_j.$$

Consideremos la función

$$F_d(x) = \sum_j b_j \delta_{a_j}(x)$$

que representa la suma de todos los saltos de F en la semirecta $(-\infty, x]$. Esta función es creciente, continua por la derecha con

$$F_d(-\infty) = 0, \qquad F_d(+\infty) \leq 1.$$

Por lo tanto F_d es una función creciente y acotada.

Teorema 3.10 *Sea $F_c(x) = F(x) - F_d(x)$. Entonces F_c es positiva, creciente y continua.*

Demostración. Sea $x < x'$, entonces

$$\begin{aligned} F_d(x') - F_d(x) &= \sum_{x < a_j \leq x'} b_j = \sum_{x < a_j \leq x'} [F(a_j) - F(a_j^-)] \\ &\leq F(x') - F(x) \end{aligned}$$

Por lo tanto F_d y F_c son ambas crecientes, y si ponemos $x = -\infty$ vemos que $F_d \leq F$, de modo que F_c es positiva. Por otro lado F_d es continua por la derecha porque cada función δ_{a_j} lo es y la serie que define F_d converge uniformemente en x. Por la misma razón

$$F_d(x) - F_d(x^-) = \begin{cases} b_j & \text{si } x = a_j, \\ 0 & \text{en otro caso.} \end{cases}$$

Pero esta relación también es cierta si reemplazamos F_d por F, por la definición de a_j y b_j, de modo que para todo x:

$$F_c(x) - F_c(x^-) = F(x) - F(x^-) - F_d(x) - F_d(x^-) = 0.$$

Como ya sabemos que F_c es continua por la derecha, esto demuestra que es continua. ■

Teorema 3.11 *Sea F una f.d. Supongamos que existe una función continua G_c y otra función G_d de la forma*

$$G_d(x) = \sum_j b'_j \delta_{a'_j}(x)$$

donde $\{a'_j\}$ es una sucesión de números reales y $\sum_j |b'_j| < \infty$, tal que

$$F = G_c + G_d,$$

entonces

$$G_c = F_c, \qquad G_d = F_d,$$

donde F_c y F_d se definen como antes.

Demostración. Si $F_d \neq G_d$ entonces o bien los conjuntos $\{a_j\}$ y $\{a'_j\}$ no son idénticas o podemos cambiar las etiquetas en las sucesiones de modo que $a'_j = a_j$ para todo j pero $b'_j \neq b_j$ para algún j. En cualquier caso tenemos que al menos para un j y $\tilde{a} = a_j$ o a'_j,

$$[F_d(\tilde{a}) - F_d(\tilde{a}^-)] - [G_d(\tilde{a}) - G_d(\tilde{a}^-)] \neq 0.$$

Como $F_c - G_c = G_d - F_d$, esto implica que

$$[F_c(\tilde{a}) - G_c(\tilde{a})] - [F_c(\tilde{a}^-) - G_c(\tilde{a}^-)] \neq 0,$$

lo cual contradice el hecho de que $F_c - G_c$ es una función continua. En consecuencia $F_d = G_d$ y $F_c = G_c$. ■

Definición 3.5 Una f.d. F que se puede representar como

$$F = \sum_j b_j \delta_{a_j}$$

se llama una f.d. *discreta*. Una f.d. que es continua en todo punto se dice *continua*.

Supongamos que $F_c \neq 0,\ F_d \neq 0$ en la descomposición de la f.d. F, entonces podemos poner $\alpha = F_d(\infty)$ de modo que $0 < \alpha < 1$,

$$F_1 = \frac{1}{\alpha} F_d, \qquad F_2 = \frac{1}{1-\alpha} F_c,$$

y escribimos

$$F = \alpha F_1 + (1-\alpha) F_2. \tag{3.12}$$

Vemos que podemos descomponer a F como una combinación convexa de una f.d. discreta y una f.d. continua. Si $F_c \equiv 0$, entonces F es discreta y ponemos $\alpha = 1$, $F_1 \equiv F$, $F_2 \equiv 0$. Si $F_d \equiv 0$, entonces F es continua y ponemos $\alpha = 0$, $F_1 \equiv 0$, $F_2 \equiv F$, y en cualquier caso la ecuación 3.12 vale. Resumimos estos resultados en el siguiente teorema.

Teorema 3.12 *Toda f.d. puede escribirse como la combinación convexa de una f.d. continua y una f.d. discreta. Esta descomposición es única.*

3.10.1. Funciones Absolutamente Continuas y Singulares

Definición 3.6 Una función F es *absolutamente continua* si existe una función integrable f tal que para todo $a < b$

$$F(b) - F(a) = \int_a^b f(t)\,dt.$$

Es posible demostrar que en este caso F tiene derivada que es igual a f c.s. En particular, si F es una f.d., entonces

$$f \geq 0 \ c.s. \qquad \text{y} \qquad \int_{-\infty}^{\infty} f(t)\,dt = 1.$$

Recíprocamente, si f satisface las condiciones anteriores, la función F definida por

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

es una f.d. que es absolutamente continua.

Definición 3.7 Una función F es *singular* si no es idénticamente cero y F' existe y es igual a cero c.s.

En el siguiente teorema presentamos algunos resultados sobre funciones reales que no demostraremos. Las demostraciones se pueden hallar en el libro de Billingsley.

Teorema 3.13 *Sea F una función creciente y acotada con $F(-\infty) = 0$ y sea F' su derivada (cuando exista). Las siguientes proposiciones son válidas.*

a) Si S denota el conjunto de $x \in \mathbb{R}$ para los cuales $F'(x)$ existe y $0 \leq F'(x) < \infty$, entonces $\lambda(S^c) = 0$.

b) F' es integrable y para todo $a < b$

$$\int_a^b F'(t)\,dt \leq F(b) - F(a).$$

c) Si definimos

$$F_{ac}(x) = \int_{-\infty}^{x} F'(t)\,dt, \qquad F_s(x) = F(x) - F_{ac}(x),$$

entonces $F'_{ac} = F'$ c.s., de modo que $F'_s = F' - F'_{ac} = 0$ c.s. y en consecuencia F_s es singular si no es idénticamente cero.

Definición 3.8 Cualquier función positiva f que sea igual a F' c.s. es una *densidad*. F_{ac} es la parte *absolutamente continua* y F_s la parte *singular* de F.

Es claro que F_{ac} es creciente y $F_{ac} \leq F$. Por el teorema anterior tenemos que si $a < b$,

$$F_s(b) - F_s(a) = F(b) - F(a) - \int_a^b f(t)\,dt \geq 0.$$

En consecuencia F_s también es creciente y $F_s \leq F$. Por lo tanto tenemos el siguiente teorema.

Teorema 3.14 *Toda f.d. F se puede escribir como combinación convexa de tres f.d., una discreta, otra continua singular y la tercera absolutamente continua. Esta descomposición es única*