

# Teoría de la Probabilidad

Guillermo Ayala & Francisco Montes

---

Departament d'Estadística i Investigació Operativa  
Universitat de València



# Índice general

<b>1. Experimento, Probabilidad y Medida</b>	<b>1</b>
1.1. Experimento, espacio muestral y $\sigma$ álgebra de sucesos	1
1.2. Medida de probabilidad	3
1.3. Medida	6
1.4. Extensión de medidas	7
1.4.1. Medida exterior	9
1.4.2. Teorema de extensión de Caratheodory	12
1.4.3. Unicidad de la extensión: Teorema $\pi - \lambda$ de Dynkin	14
<b>2. Medidas de Probabilidad en Espacios Euclídeos. Funciones de Distribución</b>	<b>17</b>
2.1. Espacio de Probabilidad sobre $\mathcal{R}$	17
2.1.1. Función de distribución asociada a $\mathcal{P}$	17
2.1.2. El problema inverso	18
2.2. Tipos de funciones de distribución	20
2.2.1. El caso discreto	20
2.2.2. El caso continuo	21
2.2.3. Densidad de probabilidad	22
2.3. El caso k-dimensional	22
<b>3. Variable Aleatoria. Esperanza</b>	<b>25</b>
3.1. Aplicación medible	25
3.1.1. Aplicaciones medibles: Definición general	25
3.1.2. Función medible	26
3.1.3. Sucesiones de funciones medibles	26
3.2. Variable aleatoria y vector aleatorio	28
3.2.1. Variable aleatoria	28
3.2.2. Vector aleatorio	30
3.3. Integración de funciones medibles	32

3.3.1.	Funciones no negativas . . . . .	32
3.3.2.	Funciones cualesquiera . . . . .	36
3.3.3.	Teoremas de convergencia . . . . .	38
3.3.4.	Cambio de variable. Densidades . . . . .	40
3.3.5.	Esperanza de una variable aleatoria . . . . .	41
<b>4.</b>	<b>Independencia y medida producto</b>	<b>45</b>
4.1.	Independencia de sucesos y variables aleatorias . . . . .	45
4.1.1.	Dos sucesos independientes . . . . .	45
4.1.2.	Independencia de clases de sucesos . . . . .	47
4.1.3.	Independencia de variables y vectores aleatorios . . . . .	48
4.1.4.	Funciones de variables independientes . . . . .	50
4.2.	Espacio producto . . . . .	50
4.2.1.	Medida producto . . . . .	51
4.2.2.	Teorema de Fubini . . . . .	54
4.3.	Independencia entre variables y función de densidad . . . . .	56
4.4.	Algunas aplicaciones . . . . .	59
4.4.1.	Estimación del volumen de objetos tridimensionales . . . . .	59
4.4.2.	Covarianza de variables independientes . . . . .	59
4.4.3.	Una igualdad muy práctica . . . . .	60
4.4.4.	Problema de la aguja de Buffon . . . . .	60
<b>5.</b>	<b>Sucesiones de Variables Aleatorias. Leyes de los Grandes Números</b>	<b>63</b>
5.1.	Convergencia en probabilidad . . . . .	63
5.2.	Lemas de Borel-Cantelli . . . . .	65
5.3.	Ley Cero-Uno de Kolmogorov . . . . .	67
5.4.	Leyes de los Grandes Números . . . . .	69
5.5.	Aplicaciones . . . . .	72
5.5.1.	El teorema de Glivenko-Cantelli . . . . .	72
5.5.2.	Cálculo aproximado de integrales por el método de Monte-Carlo . . . . .	73
5.5.3.	Números normales de Borel . . . . .	73
<b>6.</b>	<b>Convergencia débil</b>	<b>75</b>
6.1.	Convergencia débil . . . . .	75
6.2.	Relación entre convergencia débil de medidas y convergencia de integrales . . . . .	79
6.3.	Un criterio de convergencia para sucesiones ajustadas . . . . .	80
<b>7.</b>	<b>Teorema Central del Límite</b>	<b>85</b>
7.1.	Función característica . . . . .	85
7.1.1.	Definición y propiedades . . . . .	86

7.1.2.	Función característica e independencia . . . . .	87
7.1.3.	Ejemplos de funciones características . . . . .	87
7.1.4.	Teorema de inversión. Unicidad . . . . .	88
7.1.5.	Teorema de continuidad . . . . .	90
7.2.	El Teorema de Lindeberg-Lévy . . . . .	90
7.3.	Las condiciones de Lindeberg y Lyapounov . . . . .	91
7.3.1.	La condición de Lyapounov . . . . .	91
7.3.2.	La condición de Lindeberg . . . . .	92
7.4.	El Teorema de Lindeberg-Feller . . . . .	93
7.5.	Una curiosa aplicación del TCL . . . . .	94
<b>8.</b>	<b>Teorema de Radon-Nikodym</b> . . . . .	<b>97</b>
8.1.	Funciones aditiva y completamente aditiva . . . . .	97
8.2.	Descomposiciones de Jordan y Hahn . . . . .	98
8.3.	Continuidad absoluta y singularidad . . . . .	101
8.4.	Descomposición de Lebesgue. Teorema de Radon-Nikodym . . . . .	103
8.5.	Aplicación a $(\mathcal{R}, \mathcal{B}, \lambda)$ . . . . .	105



# Capítulo 1

## Experimento, Probabilidad y Medida

### 1.1. Experimento, espacio muestral y $\sigma$ álgebra de sucesos

Supongamos que estamos llevando a cabo un experimento bajo ciertas condiciones que conducen a una serie de posibles resultados. Al conjunto formado por todos estos resultados posibles lo llamaremos **espacio muestral**,  $\Omega$ , y a un subconjunto de  $\Omega$ , que cumpla determinadas propiedades que más tarde detallaremos, **suceso**. Si el resultado elemental  $\omega$  del experimento pertenece al suceso A diremos que A se ha producido o ha tenido lugar.

**Ejemplo 1.1 (Número aleatorio uniforme en el intervalo unitario)** *Un experimento algo irreal que todos conocemos consiste en producir un valor en el intervalo unitario  $(0,1]$  de modo que sea uniforme sobre el mismo, esto es, que el punto no tenga ninguna preferencia de unos subintervalos frente a otros. Si repetimos un gran número de veces el experimento obtendremos un conjunto de puntos distribuidos homogéneamente sobre el intervalo. Es evidente que no estamos hablando de un experimento concreto. Existen muchos experimentos posibles que admiten una descripción como la dada.*

En lo que sigue pretendemos dar una definición formal de qué es un experimento y trataremos de obtener información sobre los mismos que no sea evidente a partir de la mera observación de una o varias repeticiones de los experimentos.

Dado un experimento raramente el observador tiene interés en todos los posibles sucesos. Es más habitual estar interesado en unos *ciertos* sucesos. En el ejemplo anterior parece natural plantearse si el número aleatorio pertenece a cada uno de los posibles subintervalos del intervalo unitario antes que, por ejemplo, si la parte entera

de veinte veces su valor es divisible por dos. Denotemos por  $\mathcal{T}$  a esa familia de *sucesos interesantes*.

También es razonable el siguiente hecho: si nos interesa conocer si se produce un cierto suceso  $A$ , es lógico plantearse el hecho de que no se produzca  $A$ , es decir,  $A^c$ . Si tenemos interés en dos sucesos  $A$  y  $B$  es razonable considerar si se produce  $A$  y  $B$ ,  $A \cap B$ , y si se produce  $A$  o  $B$ ,  $A \cup B$ .

Las operaciones lógicas entre sucesos nos obligan a ampliar nuestra familia de sucesos  $\mathcal{T}$  a una familia que sea estable frente a las operaciones conjuntistas de complementación, unión e intersección.

**Definición 1.1 ( $\sigma$  álgebra de sucesos)** Sea  $\mathcal{A}$  una familia de subconjuntos de  $\Omega$ .  $\mathcal{A}$  es una  $\sigma$  álgebra de sucesos si:

1.  $\Omega \in \mathcal{A}$ .
2. Si  $A \in \mathcal{A}$  entonces  $A^c \in \mathcal{A}$ .
3. Si  $\{A_n\}_{n \geq 1} \subset \mathcal{A}$  entonces  $\cup_{n \geq 1} A_n \in \mathcal{A}$ .

A partir de la definición es inmediato que si  $\{A_n\}_{n \geq 1} \subset \mathcal{A}$  entonces  $\cap_{n \geq 1} A_n = [\cup_{n \geq 1} A_n^c]^c \in \mathcal{A}$ . Además si se dan las dos primeras condiciones de la definición entonces el conjunto vacío está en  $\mathcal{A}$  por lo que para  $A_1, \dots, A_m \in \mathcal{A}$  :  $\cup_{k=1}^m A_k = \cup_{k=1}^m A_k \cup \emptyset \cup \emptyset \dots \in \mathcal{A}$  y  $\cap_{k=1}^m A_k = [\cup_{k=1}^m A_k^c]^c$ . Resumiendo: una  $\sigma$  álgebra es estable para la unión e intersección de una cantidad finita de elementos de la misma. Exigir la estabilidad frente a uniones e intersecciones de solo un número finito de elementos de  $\mathcal{A}$  nos conduce a una teoría excesivamente limitada.

En lo que sigue las familias de sucesos con las que vamos a trabajar supondremos que son  $\sigma$  álgebras. A los elementos de una  $\sigma$  álgebra los llamaremos **conjuntos medibles**.

Partiendo de nuestros sucesos interesantes  $\mathcal{T}$ , ¿cuál es la menor familia de subconjuntos de  $\Omega$  tal que es  $\sigma$  álgebra y contiene a  $\mathcal{T}$ ?

**Definición 1.2 ( $\sigma$  álgebra generada)** Dada  $\mathcal{T}$ , una familia de subconjuntos de  $\Omega$  definimos la  $\sigma$  álgebra generada por  $\mathcal{T}$ ,  $\sigma(\mathcal{T})$ , como la intersección de todas las  $\sigma$  álgebras que contienen a  $\mathcal{T}$ .

$\sigma(\mathcal{T})$  es  $\sigma$  álgebra, contiene a  $\mathcal{T}$  y cualquier otra  $\sigma$  álgebra que contenga a  $\mathcal{T}$  también contiene a  $\sigma(\mathcal{T})$ .

**Ejemplo 1.2 (Número aleatorio uniforme)** Si  $\mathcal{T} = \{(a, b] \text{ con } a, b \in (0, 1]\}$  entonces  $\sigma(\mathcal{T}) = \mathcal{B}((0, 1])$  recibe el nombre de  $\sigma$  álgebra de Borel o tribu de Borel en el intervalo unitario.



**Ejemplo 1.3 (Número aleatorio real y  $\sigma$  álgebra de Borel  $\mathcal{R}$ )** Si suponemos que nuestro experimento tiene como posibles resultados los números reales (no necesariamente en  $(0,1]$ ) y consideramos  $\mathcal{T} = \{(a,b] \text{ con } a,b \in \mathcal{R}\}$  entonces  $\sigma(\mathcal{T}) = \mathcal{B}(\mathcal{R})$  o simplemente  $\mathcal{B}$  si sobreentendemos el espacio muestral.  $\mathcal{B}$  es la  $\sigma$  álgebra de Borel en la recta real. Es fácil comprobar que:

$$\mathcal{B} = \sigma((-\infty, a] \text{ con } a \in \mathcal{R}) = \sigma((b, +\infty) \text{ con } b \in \mathcal{R}).$$

Además,  $\mathcal{B}((0,1]) = \mathcal{B} \cap (0,1] = \{B \subset (0,1] : B \in \mathcal{B} \text{ siendo } B \cap (0,1] = \{B \cap (0,1] \text{ con } B \in \mathcal{B}\}$ .

$(\Omega, \mathcal{A})$ , par formado por un espacio muestral,  $\Omega$ , y una  $\sigma$  álgebra de sucesos en  $\Omega$ ,  $\mathcal{A}$ , lo denominaremos **espacio medible**.

## 1.2. Medida de probabilidad

Volvamos sobre el experimento consistente en obtener un número aleatorio uniforme en el intervalo unitario. Hasta ahora conocemos:

- Los resultados que puede producir: el espacio muestral  $(0,1]$ .
- Hemos decidido qué cosas queremos conocer del experimento: si cada suceso de la  $\sigma$  álgebra tiene lugar o no.

Responder de una forma categórica a lo anterior es demasiado ambicioso. Es imposible predecir en cada realización del experimento si el resultado va a estar o no en cada conjunto medible. En Probabilidad la pregunta se formula del siguiente modo: ¿qué posibilidad hay de que tenga lugar cada uno de los conjuntos medibles? Nos falta pues un tercer elemento que nos proporcione esa información: Una función de conjunto  $\mathcal{P}$ , es decir, una función definida sobre la  $\sigma$  álgebra y que a cada suceso le asocie un valor numérico que exprese la mayor o menor probabilidad o posibilidad de producirse cuando se realiza el experimento. Esta función de conjunto se conoce como medida de probabilidad.

Fue A.N.Kolmogorov en 1933 quien propuso la siguiente definición generalmente aceptada.

**Definición 1.3 (Medida de probabilidad)**  $\mathcal{P}$  función de conjunto definida sobre la  $\sigma$  álgebra  $\mathcal{A}$  es una medida de probabilidad si:

1. (No negativa)  $\mathcal{P}(A) \geq 0$  para todo  $A \in \mathcal{A}$ .
2. (La probabilidad del espacio muestral es uno)  $\mathcal{P}(\Omega) = 1$ .

3. (Numerablemente aditiva o  $\sigma$  aditiva) Si  $\{A_n\}_{n \geq 1}$  es una sucesión de sucesos disjuntos de  $\mathcal{A}$  entonces

$$\mathcal{P}(\cup_{n \geq 1} A_n) = \sum_{n \geq 1} \mathcal{P}(A_n).$$

A la terna  $(\Omega, \mathcal{A}, \mathcal{P})$  la llamaremos **espacio de probabilidad**. Un espacio de probabilidad es una descripción del experimento que nos informa de cuál es el conjunto de posibles resultados del experimento,  $\Omega$ , de los sucesos cuya realización nos interesa, los conjuntos medibles  $\mathcal{A}$ , y nos permite cuantificar la incertidumbre de que se produzcan los sucesos que nos interesan, la medida de probabilidad  $\mathcal{P}$ .

**Ejemplo 1.4 (Espacio de probabilidad discreto)** Supongamos un espacio muestral,  $\Omega$ , numerable y como  $\sigma$  álgebra la familia formada por todos los posibles subconjuntos de  $\Omega$ . Sea  $f$  una función no negativa definida sobre  $\Omega$  y verificando:  $\sum_{\omega \in \Omega} f(\omega) = 1$ ; entonces  $\mathcal{P}(A) = \sum_{\omega \in A} f(\omega)$  nos define una medida de probabilidad. Todas las propiedades dadas en la anterior definición son inmediatas salvo la tercera que es una consecuencia del teorema de Dirichlet (Billingsley, página 571).

Dado un entero positivo  $n$  y  $p \in [0, 1]$  un espacio de probabilidad binomial con estos parámetros es aquel con  $\Omega = \{0, \dots, n\}$  y  $p(\omega) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega}$ .

Dado un real estrictamente positivo,  $\lambda$ , un espacio de probabilidad Poisson es aquel con  $\Omega = \mathbb{Z}_+$  (enteros no negativos) y  $p(\omega) = e^{-\lambda} \frac{\lambda^\omega}{\omega!}$ .

**Ejemplo 1.5 (Medida de probabilidad discreta)** Si el espacio medible,  $(\Omega, \mathcal{A})$ , es arbitrario pero la medida de probabilidad  $\mathcal{P}$  verifica que existe un subconjunto numerable de  $\Omega$ ,  $\{\omega_k\}_{k \geq 1}$ , y una sucesión de valores no negativos,  $\{p_k\}_{k \geq 1}$ , tal que:  $\mathcal{P}(A) = \sum_{\omega_k \in A} p_k$  entonces se dice que nuestra medida de probabilidad es discreta. Sea  $(\mathcal{R}, \mathcal{B})$  y consideremos  $\omega_k = k$  para  $k = 0, \dots, n$  y  $p_k = \binom{n}{k} p^k (1-p)^{n-k}$ . Tenemos una medida de probabilidad discreta binomial. Análogamente lo haríamos en el caso de Poisson.

De la definición de medida de probabilidad se deducen algunas propiedades de la misma que son muy útiles.

**La probabilidad del vacío es cero:**  $\Omega = \Omega \cup \emptyset \cup \emptyset \cup \dots$  y por  $\sigma$  aditividad,  $\mathcal{P}(\Omega) = \mathcal{P}(\Omega) + \sum_{k \geq 1} \mathcal{P}(\emptyset)$  de modo que:  $\mathcal{P}(\emptyset) = 0$ .

**Es finitamente aditiva:** Si  $A_1, \dots, A_n$  son elementos disjuntos de  $\mathcal{A}$  aplicando la  $\sigma$  aditividad y la propiedad anterior tendremos:

$$\mathcal{P}(\cup_{i=1}^n A_i) = \mathcal{P}(\cup_{i=1}^n A_i \cup \emptyset \cup \dots) = \sum_{i=1}^n \mathcal{P}(A_i).$$

En particular para  $A \in \mathcal{A}$ :  $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$ .

**Es monótona** Si  $A, B \in \mathcal{A}$ ,  $A \subset B$  entonces  $\mathcal{P}(A) \leq \mathcal{P}(B)$ . Notemos que  $\mathcal{P}(B) = \mathcal{P}(A) + \mathcal{P}(B - A)$ .

**Probabilidad de una unión cualquiera de conjuntos medibles:** Si  $A_1, \dots, A_n \in \mathcal{A}$  entonces

$$\mathcal{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathcal{P}(A_i) - \sum_{i < j} \mathcal{P}(A_i \cap A_j) + \dots + (-1)^{n+1} \mathcal{P}(A_1 \cap \dots \cap A_n).$$

Para  $n = 2$  es cierto pues por ser la probabilidad finitamente aditiva:

$$\mathcal{P}(A_1 \cup A_2) + \mathcal{P}(A_1 \cap A_2) = \mathcal{P}(A_1) + \mathcal{P}(A_2)$$

El caso general se prueba por inducción.

**Subaditividad:** Dados  $A_1, \dots, A_n$  conjuntos medibles no es inmediato qué relación existe entre la probabilidad de la unión de los  $A_i$  y la probabilidad de cada uno de ellos. Es la siguiente:

$$\mathcal{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathcal{P}(A_i).$$

Veamos su prueba: sean  $B_1 = A_1$  y  $B_i = A_i - (\cup_{j=1}^{i-1} A_j)$  para  $i = 2, \dots, n$ . Los  $B_i$  son disjuntos y  $\cup_{i=1}^n B_i = \cup_{i=1}^n A_i$ . Por la aditividad finita y la monotonía de  $\mathcal{P}$  se tiene

$$\mathcal{P}(\cup_{i=1}^n A_i) = \mathcal{P}(\cup_{i=1}^n B_i) = \sum_{i=1}^n \mathcal{P}(B_i) \leq \sum_{i=1}^n \mathcal{P}(A_i).$$

Dada una sucesión de conjuntos medibles,  $\{A_n\}_{n \geq 1}$ , con una prueba análoga se sigue:

$$\mathcal{P}(\cup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mathcal{P}(A_n)$$

**Continuidad de la medida de probabilidad:** Sea  $\{A_n\}_{n \geq 1}$  una sucesión de conjuntos medibles y  $A$  otro conjunto medible tal que:  $A_n \subset A_{n+1}$  para todo  $n$  y  $\cup_{n \geq 1} A_n = A$ ; en lo que sigue esto lo denotaremos  $A_n \uparrow A$  ( $A_n$  es una sucesión creciente de conjuntos convergente a  $A$ ). Es cierto que si  $A_n \uparrow A$  entonces  $\mathcal{P}(A_n)$  es una sucesión creciente convergente a  $\mathcal{P}(A)$ :  $\mathcal{P}(A_n) \uparrow \mathcal{P}(A)$  (*continuidad desde abajo*); ya que si definimos la sucesión de conjuntos medibles:  $B_n = A_n - \cup_{i=1}^{n-1} A_i = A_n - A_{n-1}$  para  $n > 1$  y  $B_1 = A_1$  entonces:

$$\begin{aligned} \mathcal{P}(A) &= \mathcal{P}(\cup_{n \geq 1} A_n) = \mathcal{P}(\cup_{n \geq 1} B_n) = \sum_{n \geq 1} \mathcal{P}(B_n) = \\ &= \lim_{n \rightarrow +\infty} \sum_{j=1}^n \mathcal{P}(B_j) = \lim_{n \rightarrow +\infty} \mathcal{P}(\cup_{j=1}^n B_j) = \lim_{n \rightarrow +\infty} \mathcal{P}(A_n). \end{aligned}$$

Diremos que  $\{A_n\}_{n \geq 1}$  es una sucesión decreciente de sucesos al suceso  $A$ ,  $A_n \downarrow A$ , si  $A_n^c \uparrow A^c$ . También es cierta la siguiente afirmación: si  $A_n \downarrow A$  entonces  $\mathcal{P}(A_n) \downarrow \mathcal{P}(A)$  (*continuidad desde arriba*). Su prueba es consecuencia de la continuidad desde abajo y de que  $\mathcal{P}(B^c) = 1 - \mathcal{P}(B)$  para todo  $B$  de  $\mathcal{A}$ .

### 1.3. Medida

Supongamos un conjunto  $\Omega$  formado por  $N$  elementos ( $N > 1$ ),  $\Omega = \{\omega_1, \dots, \omega_N\}$ , y consideremos como conjunto medible cualquier subconjunto de  $\Omega$ . Sea  $\mu(A)$  el cardinal de  $A$ , es decir, el número de elementos de  $A$ . Es inmediato que la función de conjunto  $\mu$  verifica que es no negativa y  $\sigma$  aditiva así como que  $\mu(\emptyset) = 0$ . Sin embargo, no es una medida de probabilidad puesto que  $\mu(\Omega) = N$ .

**Definición 1.4 (Medida)** *Dado un espacio medible,  $(\Omega, \mathcal{A})$ , una función de conjunto  $\mu$  sobre  $\mathcal{A}$  es una medida si:*

1. (No negativa)  $\mu(A) \geq 0$  para todo  $A \in \mathcal{A}$ .
2. (El vacío mide cero)  $\mu(\emptyset) = 0$ .
3. (Numerablemente aditiva o  $\sigma$  aditiva) Si  $\{A_n\}_{n \geq 1}$  es una sucesión de sucesos disjuntos de  $\mathcal{A}$  entonces:

$$\mu(\cup_{n \geq 1} A_n) = \sum_{n \geq 1} \mu(A_n).$$

La terna  $(\Omega, \mathcal{A}, \mu)$  recibe el nombre de **espacio de medida**. Si  $\mu(\Omega)$  es un número finito diremos que la medida es **finita** (las de probabilidad lo son). La medida anterior, conocida como medida de conteo, es finita porque  $\Omega$  lo es.

Supongamos que  $\Omega$  es un conjunto infinito numerable:  $\Omega = \{\omega_n\}_{n \geq 1}$ ; como conjuntos medibles seguimos considerando cualquier posible subconjunto de  $\Omega$  y  $\mu$  la medida de conteo. En este caso la medida de  $\Omega$  es infinita y por lo tanto la medida de conteo no es finita. Observemos que las propiedades de la medida dependen tanto de su definición como del espacio medible sobre el que está definida. No obstante,  $\Omega = \cup_{n \geq 1} \{\omega_n\}$  y  $\mu(\{\omega_n\}) = 1$  que es un valor finito. En definitiva que nuestro espacio de medida es tal que podemos recubrir  $\Omega$  con una colección numerable de conjuntos medibles cada uno con medida finita. Esta propiedad es de una gran trascendencia cuando estudiamos medidas que no son de probabilidad.

**Definición 1.5 (Medida  $\sigma$  finita)** *El espacio de medida,  $(\Omega, \mathcal{A}, \mu)$ , se dice  $\sigma$  finito o que  $\mu$  es una medida  $\sigma$  finita sobre  $(\Omega, \mathcal{A})$  si existe una sucesión  $\{A_n\}_{n \geq 1}$  de elementos de  $\mathcal{A}$  tal que  $\mu(A_n) < +\infty$  y  $\cup_{n \geq 1} A_n = \Omega$ .*

Las mismas pruebas que hemos utilizado para medidas de probabilidad son válidas para probar que una medida cualquiera,  $\mu$ , es monótona, numerablemente subaditiva y continua desde abajo. También es casi continua desde arriba, concretamente, si  $A_n$  decrece a  $A$  entonces  $\mu(A_n)$  decrece a  $\mu(A)$  siempre que exista un  $n_0$  tal que  $\mu(A_{n_0}) < \infty$ .

## 1.4. Extensión de medidas

Volvamos al experimento consistente en generar un número aleatorio uniforme en el intervalo unitario  $(0, 1]$ . Los posibles resultados los conocemos. También hemos fijado cuáles son nuestros sucesos interesantes, saber si el punto aleatorio está o no en cada uno de los subintervalos de  $(0, 1]$ . Por razones de conveniencia matemática trabajar con estos sucesos es excesivamente limitado y hemos de considerar la  $\sigma$  álgebra de Borel. Falta el tercer elemento del espacio de probabilidad: cuando hablamos de número aleatorio uniforme, ¿a qué medida de probabilidad  $\mathcal{P}$  nos estamos refiriendo? Obviamente sabemos que  $\mathcal{P}((0, 1]) = 1$  y para un intervalo  $(a, b]$  parece razonable que su probabilidad dependa exclusivamente de su tamaño y no de su posición dentro de  $(0, 1]$ . Concretamente,  $\mathcal{P}((a, b]) = \mathcal{P}((a + c, b + c])$  siempre que ambos intervalos estén contenidos en el intervalo unitario. Más adelante veremos formalmente algo que es intuitivo:  $\mathcal{P}((a, b]) = k(b - a)$  con  $k$  una constante positiva, es decir, la probabilidad de que el punto aleatorio esté en un intervalo es proporcional a su longitud. Pero  $\mathcal{P}((0, 1]) = 1$  de donde  $k = 1$ . En definitiva, el modelo uniforme razonable sería aquel en el que la probabilidad de que el punto aleatorio pertenezca a un subintervalo cualquiera viene dada por la longitud del mismo. No todo está hecho. Hemos de definir la medida de probabilidad sobre toda la tribu de Borel. ¿Cuál es la probabilidad de que el punto aleatorio esté en un boreliano  $A$  cualquiera? Parece natural definirla como la longitud de  $A$ ,  $\lambda(A)$ . No es trivial la cuestión. Medir la longitud de un intervalo  $(a, b]$  es sencillo,  $b - a$ , la diferencia de sus extremos. Si consideramos la unión de un número finito de intervalos disjuntos,  $\bigcup_{i=1}^n (a_i, b_i]$ , puesto que  $\lambda$  ha de ser finitamente aditiva hemos de definir

$$\lambda\left(\bigcup_{i=1}^n (a_i, b_i]\right) = \sum_{i=1}^n \lambda((a_i, b_i]) = \sum_{i=1}^n (b_i - a_i);$$

además es fácil comprobar que esta segunda definición es correcta (si  $\bigcup_{i=1}^n (a_i, b_i] = \bigcup_{j=1}^m (c_j, d_j]$  entonces  $\sum_{i=1}^n (b_i - a_i) = \sum_{j=1}^m (d_j - c_j)$ ). Pero en la tribu de Borel existen conjuntos que no son intervalos o uniones de intervalos, ¿cómo definir la longitud en estos casos?. El objeto de este tema es responder las dos preguntas siguientes:

1. ¿Existe una medida de probabilidad sobre  $((0, 1], \mathcal{B}(0, 1])$  que a los intervalos les asigne la diferencia entre sus extremos?

## 2. ¿Es única?

Una respuesta afirmativa nos permitirá hablar de la longitud de los subconjuntos de Borel de  $(0,1]$  o con mayor precisión de su medida de Lebesgue.

El intervalo unitario es un subconjunto de la recta real. Parece bastante artificioso definir la longitud solamente sobre dicho intervalo. También veremos en este tema que los mismos resultados que responden afirmativamente las preguntas anteriores nos van a permitir afirmar la existencia de una única función de conjunto sobre  $(\mathcal{R}, \mathcal{B})$ ,  $\lambda'$ , tal que  $\lambda'(a, b] = b - a$  para  $a$  y  $b$  en  $\mathcal{R}$ : la medida de Lebesgue sobre la recta real. La siguiente pregunta es obvia: ¿coinciden ambas funciones de conjunto sobre los borelianos del intervalo unitario?

Los resultados que vamos a tratar en este tema podríamos enunciarlos y probarlos para medidas de probabilidad. Es una limitación innecesaria. Salvo algún detalle menor las pruebas para medidas son las mismas y la generalidad que obtenemos la utilizaremos más adelante.

En los ejemplos antes comentados tenemos dos elementos comunes:

1. Una familia sencilla de conjuntos medibles que genera toda la  $\sigma$  álgebra de interés y sobre la cual es fácil definir la medida (*los intervalos*).
2. Una función de conjunto sobre esa familia verificando unas *ciertas* propiedades que la hacen candidata a poderla extender a una medida sobre la  $\sigma$  álgebra.

La familia de conjuntos no puede ser una cualquiera.

**Definición 1.6 (Semianillo)** Una clase de subconjuntos de  $\mathcal{A}$  es un semianillo si verifica:

1.  $\emptyset \in \mathcal{A}$ ;
2. Si  $A, B \in \mathcal{A}$  entonces  $A \cap B \in \mathcal{A}$ ;
3.  $A, B \in \mathcal{A}$  y  $A \subset B$ , entonces existen  $C_1, \dots, C_n$  en  $\mathcal{A}$  y disjuntos tales que  $B - A = \bigcup_{k=1}^n C_k$ .

Es fácil comprobar que los subintervalos del intervalo unitario o de  $\mathcal{R}$  son semianillos. Otro ejemplo importante de semianillo es el formado por los conjuntos de la forma:

$$\prod_{i=1}^d (a_i, b_i] \text{ con } a_i, b_i \in \mathcal{R},$$

a los que llamaremos **rectángulos** de  $\mathcal{R}^d$ .

### 1.4.1. Medida exterior

Partimos de una función de conjunto,  $\mu$ , definida sobre un semianillo  $\mathcal{A}_0$ , verificando las siguientes condiciones: toma valores no negativos, en el vacío vale cero, es finitamente aditiva y numerablemente subaditiva. Una medida verifica también estas propiedades.

¿Cómo podemos extender la función  $\mu$  a la  $\sigma$  álgebra engendrada a partir de  $\mathcal{A}_0$  de modo que dicha extensión sea una medida?

Empecemos definiendo la posible extensión sobre cualquier posible subconjunto del espacio muestral  $\Omega$ .

$$\mu^*(A) = \inf \sum_n \mu(A_n) \quad (1.1)$$

donde el ínfimo anterior se toma sobre todas las sucesiones finitas o numerables de conjuntos  $A_n$  de  $\mathcal{A}_0$  tales que  $A \subset \bigcup_n A_n$ . Cuando no exista ningún recubrimiento como el descrito el ínfimo anterior suponemos que vale  $+\infty$ .

La idea es simple: para medir el conjunto  $A$  elegimos *buenos* recubrimientos  $\{A_n\}$ , es decir, recubrimientos tales que no hay mucho solapamiento entre los conjuntos y que recubran prácticamente solo al conjunto  $A$ . Asignamos al conjunto  $A$  la suma de los valores  $\mu(A_n)$ . Estamos midiendo *desde el exterior* al conjunto  $A$ .  $\mu^*$  definida como ecuación 1.1 recibe el nombre de **medida exterior** de  $A$ .

Supongamos por un momento que  $\mu$  es una medida finita sobre  $(\Omega, \mathcal{A})$ . Entonces verificaría la siguiente igualdad:  $\mu(A^c) = \mu(\Omega) - \mu(A)$ . Si pretendemos medir *desde el interior* el conjunto  $A$ , teniendo en cuenta este último comentario parece natural definir la **medida interior** de  $A$  como:

$$\mu_*(A) = \mu^*(\Omega) - \mu^*(A^c). \quad (1.2)$$

En principio las medidas interior y exterior de un conjunto cualquiera no tienen porqué coincidir. Parece razonable considerar aquellos conjuntos donde ambos valores coinciden. Si  $A$  es uno de esos conjuntos verificará la igualdad:

$$\mu^*(A) + \mu^*(A^c) = \mu^*(\Omega). \quad (1.3)$$

Vamos a exigir la condición anterior no solamente para todo el espacio muestral sino para cualquier subconjunto de dicho espacio.

**Definición 1.7** Un subconjunto  $A$  diremos que es medible respecto  $\mu^*$  si

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E) \quad (1.4)$$

para todo subconjunto  $E$  de  $\Omega$ . Denotamos por  $\mathcal{M}(\mu^*)$  la familia de los conjuntos medibles respecto de  $\mu^*$ .

$\mathcal{M}(\mu^*)$  son aquellos conjuntos suficientemente suaves sobre los cuales la función  $\mu^*$  tiene algunas propiedades de una medida.

**Teorema 1.1** Si  $\mu^*$  la definimos como en ecuación 1.1 entonces  $\mathcal{M}(\mu^*)$  es una  $\sigma$  álgebra y  $\mu^*$  definida sobre  $\mathcal{M}(\mu^*)$  es una medida.

Previo a la prueba del teorema necesitamos el siguiente resultado.

**Proposición 1.1**  $\mu^*$  verifica:

1.  $\mu^*(\emptyset) = 0$ .
2. (No negativa)  $\mu^*(A) \geq 0$  para cualquier  $A \subset \Omega$ .
3. (Monótona) Si  $A \subset B$  implica  $\mu^*(A) \leq \mu^*(B)$ .
4. (Numerablemente subaditiva)  $\mu^*(\bigcup_n A_n) \leq \sum_n \mu^*(A_n)$ .

**Demostración.-** Solo necesita prueba la última propiedad: dado  $\epsilon > 0$ , para cada  $n$  existe un cubrimiento  $\{A_{nm}\}_{m \geq 1}$  de  $A_n$  verificando

$$\mu^*(A_n) + \frac{\epsilon}{2^n} \geq \sum_{m \geq 1} \mu(A_{nm}).$$

$\{A_{nm}\}_{n,m \geq 1}$  constituyen un cubrimiento numerable de  $\bigcup_{n \geq 1} A_n$  y por lo tanto:

$$\mu^*\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \sum_{m \geq 1} \mu(A_{nm}) \leq \sum_{n \geq 1} \left(\mu^*(A_n) + \frac{\epsilon}{2^n}\right) = \epsilon + \sum_{n \geq 1} \mu^*(A_n)$$

Lo anterior es cierto para cualquier  $\epsilon$  positivo de modo que  $\mu^*\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \mu^*(A_n)$ .  
□

En lo que sigue, siempre que utilicemos  $\mu^*$ , solamente necesitaremos que verifique las propiedades que aparecen recogidas en la anterior proposición. Notemos que por ser  $\mu^*$  numerablemente subaditiva un conjunto  $A$  pertenece a  $\mathcal{M}(\mu^*)$  si y solo si:

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

para cualquier subconjunto  $E$  de  $\Omega$ . Ya podemos probar el teorema.

**Demostración del teorema 1.1**

$\mathcal{M}(\mu^*)$  es una  $\sigma$  álgebra.

- Es fácil comprobar que  $\Omega \in \mathcal{M}(\mu^*)$  y la estabilidad de  $\mathcal{M}(\mu^*)$  frente a la complementación.



- $\mathcal{M}(\mu^*)$  es estable frente a uniones finitas: si  $A, B \in \mathcal{M}(\mu^*)$  y  $E \subset \Omega$  entonces

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

por lo que  $A \cup B \in \mathcal{M}(\mu^*)$ . Por inducción se completa la prueba.

- $\mathcal{M}(\mu^*)$  es estable bajo uniones numerables disjuntas: sea  $\{A_n\}_{n \geq 1} \subset \mathcal{M}(\mu^*)$  donde los  $A_n$  son disjuntos. Veamos que  $\bigcup_{n \geq 1} A_n \in \mathcal{M}(\mu^*)$ .

Empezamos probando la siguiente igualdad:

$$\text{si } E \subset \Omega, \quad \mu^*(E \cap \bigcup_{n=1}^m A_n) = \sum_{n=1}^m \mu^*(E \cap A_n). \quad (1.5)$$

La igualdad es inmediata para  $m = 1$ . La suponemos cierta para  $m$  y vamos a probarlo para  $m + 1$ . Notemos que  $\bigcup_{n=1}^m A_n \in \mathcal{M}(\mu^*)$  por lo que

$$\begin{aligned}\mu^*(E \cap \bigcup_{n=1}^{m+1} A_n) &= \\ \mu^*(E \cap \bigcup_{n=1}^{m+1} A_n \cap \bigcup_{n=1}^m A_n) + \mu^*(E \cap \bigcup_{n=1}^{m+1} A_n \cap (\bigcup_{n=1}^m A_n)^c) &= \\ \mu^*(E \cap \bigcup_{n=1}^m A_n) + \mu^*(E \cap A_{m+1}) &= \\ \sum_{n=1}^m \mu^*(E \cap A_n) + \mu^*(E \cap A_{m+1})\end{aligned}$$

Tenemos probada la ecuación 1.5. Vamos a particionar  $E$  utilizando  $\bigcup_{n=1}^m A_n$ ,

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

Tomando  $m \rightarrow +\infty$  se sigue:

$$\begin{aligned}\mu^*(E) &\geq \sum_{n \geq 1} (E \cap A_n) + \mu^*(E \cap (\bigcup_{n \geq 1} A_n)^c) \\ &\geq \mu^*(E \cap (\bigcup_{n \geq 1} A_n)) + \mu^*(E \cap (\bigcup_{n \geq 1} A_n)^c)\end{aligned}$$

Es decir,  $\bigcup_{n \geq 1} A_n \in \mathcal{M}(\mu^*)$ .

- Es estable bajo uniones numerables cualesquiera.

Sea  $\{A_n\}_{n \geq 1} \subset \mathcal{M}(\mu^*)$ . Definimos  $B_1 = A_1$  y  $B_n = A_n - (\bigcup_{i=1}^{n-1} A_i)$  para  $n > 1$ . Pero  $\mathcal{M}(\mu^*)$  es estable bajo la complementación y uniones finitas por lo que  $B_n \in \mathcal{M}(\mu^*)$  para todo  $n$ . Además  $\bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} B_n \in \mathcal{M}(\mu^*)$  ya que los  $B_n$  son disjuntos.

**$\mu^*$  definida sobre  $\mathcal{M}(\mu^*)$  es una medida.**

Si tenemos en cuenta los dos primeros apartados de la proposición anterior solo tenemos que ocuparnos de la  $\sigma$  aditividad de  $\mu^*$ .

- $\mu^*$  es finitamente aditiva: si  $A, B \in \mathcal{M}(\mu^*)$  disjuntos entonces:

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

y por inducción completamos la prueba.

- $\mu^*$  es  $\sigma$  aditiva: si  $\{A_n\}_{n \geq 1} \subset \mathcal{M}(\mu^*)$  disjuntos entonces

$$\sum_{n=1}^m \mu^*(A_n) = \mu^*\left(\bigcup_{n=1}^m A_n\right) \leq \mu^*\left(\bigcup_{n \geq 1} A_n\right) \text{ para todo } m \geq 1.$$

En consecuencia:  $\sum_{n \geq 1} \mu^*(A_n) \leq \mu^*(\bigcup_{n \geq 1} A_n)$  y por subaditividad de  $\mu^*$  se completa la prueba.  $\square$

#### 1.4.2. Teorema de extensión de Caratheodory

Nuestro resultado de extensión básico es el siguiente teorema.

**Teorema 1.2 (Un teorema de extensión)** *Sea  $\mu$  una función de conjunto definida sobre el semianillo  $\mathcal{A}_0$  verificando:*

1.  $\mu(A) \in [0, +\infty]$  para  $A \in \mathcal{A}_0$ ;
2.  $\mu(\emptyset) = 0$ ;

3.  $\mu$  es finitamente aditiva;

4.  $\mu$  es numerablemente subaditiva

entonces existe una medida sobre  $\sigma(\mathcal{A}_0)$  que coincide con  $\mu$  sobre  $\mathcal{A}_0$ .

**Demostración.-** Necesitamos, como resultado previo, comprobar que  $\mu$  es monótona. Si  $A, B \in \mathcal{A}_0$ ,  $A \subset B$ , entonces existen  $C_1, \dots, C_n$  disjuntos tales que  $B - A = \bigcup_{i=1}^n C_i$ . Por ser  $\mu$  finitamente aditiva  $\mu(B) = \mu(A) + \sum_{i=1}^n \mu(C_i) \geq \mu(A)$  ya que  $\mu$  es no negativa.

Definamos  $\mu^*$  sobre los subconjuntos de  $\Omega$  como:

$$\mu^*(A) = \inf \sum_{n \geq 1} \mu(A_n)$$

siendo  $\{A_n\}_{n \geq 1}$  un cubrimiento de  $A$  mediante elementos de  $\mathcal{A}_0$ . Por el teorema anterior  $(\Omega, \mathcal{M}(\mu^*), \mu^*)$  es un espacio de medida.

$\mathcal{A}_0$  está contenido en  $\mathcal{M}(\mu^*)$ : sea  $A \in \mathcal{A}_0$  y  $E \subset \Omega$ ; queremos probar que  $\mu^*(E) \geq \mu^*(A \cap E) + \mu^*(A^c \cap E)$ . Si  $\mu^*(E) = +\infty$  es inmediato. Para  $E$  tal que  $\mu^*(E)$  es finito y  $\epsilon > 0$  tomamos un cubrimiento  $\{A_n\}_{n \geq 1} \subset \mathcal{A}_0$  de  $E$  tal que:  $\sum_{n \geq 1} \mu(A_n) < \mu^*(E) + \epsilon$ .  $\mathcal{A}_0$  es semianillo por lo que  $B_n = A \cap A_n \in \mathcal{A}_0$  y  $A^c \cap A_n = A_n - B_n = \bigcup_{k=1}^{m_n} C_{nk}$  siendo  $C_{nk}$  con  $k = 1, \dots, m_n$  elementos disjuntos de  $\mathcal{A}_0$ . Además  $A \cap E \subset \bigcup_{n \geq 1} B_n$  y  $A^c \cap E \subset \bigcup_{n \geq 1} \bigcup_{k=1}^{m_n} C_{nk}$ . Se tiene:

$$\begin{aligned} \mu^*(A \cap E) + \mu^*(A^c \cap E) &\leq \sum_{n \geq 1} \mu(B_n) + \sum_{n \geq 1} \sum_{k=1}^{m_n} \mu(C_{nk}) = \\ &= \sum_{n \geq 1} (\mu(B_n) + \sum_{k=1}^{m_n} \mu(C_{nk})) = \sum_{n \geq 1} \mu(A_n) < \mu^*(E) + \epsilon \end{aligned}$$

y esto es cierto para cualquier  $\epsilon$  positivo. En definitiva  $A \in \mathcal{M}(\mu^*)$ .

$\mu^*$  coincide con  $\mu$  sobre  $\mathcal{A}_0$ : si  $A, A_n \in \mathcal{A}_0$  con  $A \subset \bigcup_{n \geq 1} A_n$  ya que suponemos  $\mu$  numerablemente subaditiva y hemos probado que  $\mu$  es monótona se sigue que

$$\mu(A) \leq \sum_{n \geq 1} \mu(A_n \cap A) \leq \sum_{n \geq 1} \mu(A_n)$$

de donde  $\mu(A) \leq \mu^*(A)$ . La desigualdad contraria es inmediata.

*El remate final:*  $\mathcal{A}_0 \subset \mathcal{M}(\mu^*)$  pero  $\mathcal{M}(\mu^*)$  por el teorema anterior es una  $\sigma$  álgebra de modo que  $\sigma(\mathcal{A}_0) \subset \mathcal{M}(\mu^*)$ .  $\mu^*$  es una medida sobre  $\mathcal{M}(\mu^*)$  por lo que seguirá siendo medida sobre  $\sigma(\mathcal{A}_0)$ ,  $\sigma$  álgebra incluida en  $\mathcal{M}(\mu^*)$ . Además esta medida  $\mu^*$  coincide con  $\mu$  sobre  $\mathcal{A}_0$ .  $\square$

**Ejemplo 1.6 (En la recta existe al menos una longitud)** *Estamos en  $\mathcal{R}$  con su  $\sigma$  álgebra de Borel. Nada más fácil que definir  $\lambda'(a, b] = b - a$ .  $\mathcal{A}_0 = \{(a, b] \text{ con } a, b \in \mathcal{R}\}$  es un semianillo. ¿ $\lambda'$  verifica las condiciones del teorema? Comprobar las tres primeras condiciones es trivial. La cuarta lo parece pero no lo es. En el tema siguiente lo veremos. Una prueba simple puede verse en Billingsley, página 24. Por el teorema anterior sabemos que existe al menos una medida sobre  $(\mathcal{R}, \mathcal{B})$  verificando que coincide con  $\lambda'$  sobre  $\mathcal{A}_0$ .*

*Si suponemos que estamos en  $((0, 1], \mathcal{B}(0, 1])$  lo dicho sigue siendo válido y al menos existe una medida en este espacio medible que sobre los intervalos nos da su longitud.*

*¿Son únicas estas medidas?*

### 1.4.3. Unicidad de la extensión: Teorema $\pi - \lambda$ de Dynkin

Enunciado de un modo simple en esta sección nos ocupamos del siguiente problema: si tenemos dos medidas,  $\mu$  y  $\nu$ , sobre un mismo espacio medible,  $(\Omega, \mathcal{A})$  pretendemos determinar qué condiciones han de verificar estas medidas y sobre qué subfamilia de conjuntos de  $\mathcal{A}$  (lo más sencilla posible en el sentido de que contenga *pocos* sucesos y las condiciones que intervengan en su definición sean fáciles de verificar) han de coincidir  $\mu$  y  $\nu$  para tener asegurada la igualdad de  $\mu$  y  $\nu$  sobre toda la  $\sigma$  álgebra  $\mathcal{A}$ . Por el teorema de extensión que hemos propuesto en la sección anterior parece que esa familia en donde coinciden  $\mu$  y  $\nu$  debiera de ser al menos un semianillo. Vamos a ver que es suficiente con que sea un  $\pi$  sistema.

**Definición 1.8**  $\mathcal{T}$  es un  $\pi$  sistema si:

$$A, B \in \mathcal{T} \text{ implica } A \cap B \in \mathcal{T}$$

Una respuesta suficiente al problema planteado es la siguiente:

**Teorema 1.3 (Teorema de unicidad)** *Sean  $\mu$  y  $\nu$  medidas sobre  $\sigma(\mathcal{T})$  donde  $\mathcal{T}$  es un  $\pi$  sistema verificando que:*

1.  $\mu$  y  $\nu$  coinciden sobre  $\mathcal{T}$ ;
2.  $\mu$  y  $\nu$  son  $\sigma$  finitas sobre  $\mathcal{T}$ , es decir, existen  $\{A_n\}_{n \geq 1} \subset \mathcal{T}$  con  $\bigcup_{n \geq 1} A_n = \Omega$  y  $\mu(A_n) = \nu(A_n) < +\infty$  para todo  $n$

*entonces  $\mu$  y  $\nu$  coinciden sobre  $\sigma(\mathcal{T})$ .*

Para probar el teorema necesitamos definir un nuevo concepto,  $\lambda$  sistema, y un resultado conocido como teorema  $\pi - \lambda$  de Dynkin.

**Definición 1.9 ( $\lambda$  sistema)** *Una clase de conjuntos  $\mathcal{L}$  se dice que es un  $\lambda$  sistema si:*

1.  $\Omega \in \mathcal{L}$ .
2. Si  $A \in \mathcal{L}$ , entonces  $A^c \in \mathcal{L}$ .
3. Si  $\{A_n\}_{n \geq 1} \subset \mathcal{L}$  disjuntos, entonces  $\bigcup_{n \geq 1} A_n \in \mathcal{L}$ .

El siguiente resultado es esencial en la prueba de muchos resultados de probabilidad incluyendo el anterior teorema.

**Teorema 1.4 (Teorema  $\pi - \lambda$  de Dynkin)** Si  $\mathcal{T}$  es un  $\pi$  sistema y  $\mathcal{L}$  es un  $\lambda$  sistema tales que  $\mathcal{T} \subset \mathcal{L}$  entonces  $\sigma(\mathcal{T}) \subset \mathcal{L}$ .

La demostración del resultado puede encontrarse en Billingsley, página 37.

#### Demostración del teorema de unicidad

Sea  $B \in \mathcal{T}$  verificando  $\mu(B) = \nu(B) < +\infty$  y consideremos:

$$\mathcal{L}(B) = \{A \in \sigma(\mathcal{T}), \mu(A \cap B) = \nu(A \cap B)\}.$$

Sin dificultad se comprueba que  $\mathcal{L}(B)$  es un  $\lambda$  sistema; pero  $\mathcal{T} \subset \mathcal{L}(B)$  lo que implica  $\sigma(\mathcal{T}) \subset \mathcal{L}(B)$ . En resumen: si  $B \in \mathcal{T}$  con  $\mu(B) = \nu(B) < +\infty$  entonces  $\mu(A \cap B) = \nu(A \cap B)$  para cualquier  $A$  de  $\sigma(\mathcal{T})$ .

$\mu$  y  $\nu$  son  $\sigma$  finitas sobre  $\mathcal{T}$  de modo que existirá una sucesión  $\{B_n\}_{n \geq 1}$  de elementos de  $\mathcal{T}$  tales que  $\bigcup_{n \geq 1} B_n = \Omega$  y  $\mu(B_n) = \nu(B_n) < +\infty$ . Si  $A \in \sigma(\mathcal{T})$  entonces

$$\begin{aligned} \mu\left(\left(\bigcup_{i=1}^n B_i\right) \cap A\right) &= \mu\left(\bigcup_{i=1}^n (B_i \cap A)\right) = \\ &= \sum_{i=1}^n \mu(B_i \cap A) - \sum_{1 \leq i < j \leq n} \mu(B_i \cap B_j \cap A) + \dots \end{aligned}$$

La última igualdad la probamos para medidas de probabilidad pero es igualmente válida para medidas cualesquiera siempre que la medida de los conjuntos sea finita. La misma cadena de igualdades la podemos escribir cambiando  $\mu$  por  $\nu$ .

$\mathcal{T}$  es un  $\pi$  sistema de modo que  $B_i \cap B_j, B_i \cap B_j \cap B_k, \dots$  son elementos de  $\mathcal{T}$  de medida finita y, por lo visto anteriormente:  $\mu(B_i \cap B_j \cap A) = \nu(B_i \cap B_j \cap A)$ ,  $\mu(B_i \cap B_j \cap B_k \cap A) = \nu(B_i \cap B_j \cap B_k \cap A), \dots$  Haciendo tender  $n$  a  $+\infty$  se sigue  $(\bigcup_{i=1}^n B_i) \cap A \uparrow A$  de donde  $\mu((\bigcup_{i=1}^n B_i) \cap A) \uparrow \mu(A)$  (lo mismo para  $\nu$ ) y en consecuencia  $\mu(A) = \nu(A)$ .  $\square$

**Ejemplo 1.7** Estamos en el espacio medible  $(\mathcal{R}, \mathcal{B})$ . Sabíamos de la existencia de al menos una función de conjunto  $\lambda'$  tal que  $\lambda'((a, b]) = b - a$ . ¿Cuántas  $\lambda'$  existen?

$\mathcal{A}_0 = \{(a, b] \text{ con } a, b \in \mathcal{R}\}$  es un  $\pi$  sistema que engendra  $\mathcal{B}$ . Si  $\lambda''$  es una medida sobre  $(\mathcal{R}, \mathcal{B})$  tal que  $\lambda''((a, b]) = b - a = \lambda'((a, b])$  entonces por el teorema de unicidad  $\lambda'$  y  $\lambda''$  han de ser iguales.

Consideremos el espacio medible  $((0, 1], \mathcal{B}((0, 1]))$ . En este espacio también podemos probar la existencia de al menos una medida que a los intervalos asocie su longitud. Por el mismo razonamiento que acabamos de utilizar esta medida es única. Pero  $\mathcal{B}((0, 1]) = \{A \in \mathcal{B}, A \subset (0, 1]\}$  de modo que  $\lambda'$  está definida sobre  $\mathcal{B}((0, 1])$ . ¿Coinciden  $\lambda$  y  $\lambda'$  sobre  $\mathcal{B}((0, 1])$ ? Es inmediato que sí. Esta única medida es la formalización matemática del concepto intuitivo de longitud y se la denomina medida de Lebesgue en  $\mathcal{R}$  o unidimensional y la denotaremos por  $\lambda$ .

## Capítulo 2

# Medidas de Probabilidad en Espacios Euclídeos. Funciones de Distribución

### 2.1. Espacio de Probabilidad sobre $\mathcal{R}$

En el capítulo anterior hemos trabajado con espacios de probabilidad abstractos  $(\Omega, \mathcal{A}, \mathcal{P})$ . En este capítulo suponemos que nuestra medida de probabilidad está definida sobre  $\mathcal{R}$  dotado con su correspondiente  $\sigma$  álgebra de Borel,  $\mathcal{B}$ . Nuestro espacio de probabilidad será pues  $(\mathcal{R}, \mathcal{B}, \mathcal{P})$ . El interés de trabajar en  $\mathcal{R}$  viene justificado porque en muchas ocasiones el resultado de nuestro experimento es un valor real y además porque, como veremos en el próximo capítulo, una variable aleatoria nos permitirá siempre trasladarnos de un espacio de probabilidad abstracto a  $(\mathcal{R}, \mathcal{B}, \mathcal{P})$  en el que resulta mucho más cómodo trabajar.

Una medida de probabilidad es un objeto matemático de una gran complejidad. Por esta razón una de las cuestiones básicas es definir objetos más simples que caractericen la medida de probabilidad: la función de distribución de la medida de probabilidad es el primer ejemplo.

#### 2.1.1. Función de distribución asociada a $\mathcal{P}$

A partir de la medida de probabilidad podemos definir sobre  $\mathcal{R}$  una función puntual mediante

$$F(x) = \mathcal{P}((-\infty, x]), \quad \forall x \in \mathcal{R}. \quad (2.1)$$

Así definida esta función tiene las siguientes propiedades:

**Monotonía:** De la monotonía de la medida se deduce fácilmente que  $F(x_1) \leq F(x_2)$  si  $x_1 \leq x_2$ .

**Continuidad por la derecha:** Consideremos una sucesión  $x_n \downarrow x$ , la correspondiente sucesión de intervalos verifica  $\bigcap_n (-\infty, x_n] = (-\infty, x]$ , y por la continuidad desde arriba de  $\mathcal{P}$  respecto del paso al límite tendremos  $\lim_{x_n \downarrow x} F(x_n) = F(x)$ .

Observemos por otra parte que  $(-\infty, x] = \{x\} \cup \lim_{n \rightarrow +\infty} (-\infty, x - \frac{1}{n}]$ , lo que, al tomar probabilidades, conduce a

$$F(x) = \mathcal{P}(\{x\}) + \lim_{n \rightarrow +\infty} F(x - \frac{1}{n}) = \mathcal{P}(\{x\}) + F(x-), \quad (2.2)$$

A partir de 2.2 se sigue que  $F(x)$  es continua en  $x$  si y solo si  $\mathcal{P}(\{x\}) = 0$ .

**Valores límites:** Si  $x_n \uparrow +\infty$  o  $x_n \downarrow -\infty$  entonces  $(-\infty, x_n] \uparrow \mathcal{R}$  y  $(-\infty, x_n] \downarrow \emptyset$  y por lo tanto

$$F(+\infty) = \lim_{x_n \uparrow +\infty} F(x_n) = 1, \quad F(-\infty) = \lim_{x_n \downarrow -\infty} F(x_n) = 0.$$

Una función  $F$  que verifica las tres propiedades anteriores recibe el nombre de *función de distribución de probabilidad* (en lo que sigue simplemente función de distribución).

La función de distribución asociada a una medida de probabilidad  $\mathcal{P}$  constituye una descripción de la misma. ¿Es una descripción completa? ¿Dos medidas de probabilidad  $\mathcal{P}_1$  y  $\mathcal{P}_2$  diferentes pueden tener la misma función de distribución asociada? No. Estas medidas verificarían la siguiente igualdad:

$$\mathcal{P}_1((-\infty, x]) = \mathcal{P}_2((-\infty, x]),$$

para cualquier  $x$  real. Pero los intervalos  $\{(-\infty, x] : x \in \mathcal{R}\}$  constituyen un  $\pi$  sistema que engendra la  $\sigma$  álgebra de Borel y aplicando el teorema de unicidad del capítulo anterior se sigue que  $\mathcal{P}_1$  y  $\mathcal{P}_2$  son la misma medida. En resumen: si  $F$  es una función real de variable real para la cual *sabemos* que existe una medida de probabilidad que la tiene por función de distribución entonces dicha medida viene caracterizada por  $F$ . Queda una pregunta por responder: dada  $F$ , ¿cómo podemos saber si existe una medida de probabilidad para la cual  $F$  es su función de distribución? El siguiente apartado se ocupa de dar la respuesta a esta pregunta.

### 2.1.2. El problema inverso

A partir de una función de distribución  $F$ , definimos sobre el semianillo  $\mathcal{T}$  de los intervalos acotados semiabiertos de  $\mathcal{R}$  una función de conjunto mediante

$$\mu((a, b]) = F(b) - F(a).$$



El teorema 1.2 del capítulo anterior sería de aplicación a la hora de extender  $\mu$  a la  $\sigma$  álgebra de Borel engendrada por  $\mathcal{T}$ , extensión que, dadas las características de  $F$ , será una medida de probabilidad. La aplicación del mencionado teorema requiere que nuestra  $\mu$  verifique sus cuatro hipótesis, siendo las tres primeras de fácil comprobación y requiriendo la cuarta, *la subaditividad numerable*, mayor atención. Los dos teoremas que siguen demuestran que  $\mu$  es no solamente subaditiva sino  $\sigma$  aditiva.

**Teorema 2.1** Si  $I = (a, b]$  y  $I_i = (a_i, b_i]$ ,  $i = 1, \dots, n$ , son elementos de  $\mathcal{T}$ , entonces

1.  $\sum_{i=1}^n \mu(I_i) \leq \mu(I)$  si  $\bigcup_{i=1}^n I_i \subset I$  y los  $I_i$  son disjuntos,
2.  $\mu(I) \leq \sum_{i=1}^n \mu(I_i)$  si  $\bigcup_{i=1}^n I_i \supset I$ , y
3.  $\sum_{i=1}^n \mu(I_i) = \mu(I)$  si  $I = \bigcup_{i=1}^n I_i$  y los  $I_i$  son disjuntos.

**Demostración.-** Como el caso  $n = 1$  es obvio supondremos en lo que sigue que  $n > 1$ .

1. Suponiendo que  $a \leq a_1 \leq b_1 \leq a_2 \leq \dots \leq a_n \leq b_n \leq b$ , podemos escribir

$$\begin{aligned} \mu(I) &= F(b) - F(a) \geq F(b_n) - F(a_1) = \\ &= \sum_{i=1}^n [F(b_i) - F(a_i)] + \sum_{i=1}^{n-1} [F(a_{i+1}) - F(b_i)] \geq \\ &\geq \sum_{i=1}^n [F(b_i) - F(a_i)] = \sum_{i=1}^n \mu(I_i). \end{aligned}$$

2. Sin pérdida de generalidad puede suponerse que los  $I_i$  son un cubrimiento minimal de  $I$  verificándose, si  $a_1 \leq a_2 \leq \dots \leq a_n$ , que  $a_1 \leq a \leq b_1$ ,  $a_n \leq b \leq b_n$  y  $a_{i+1} \leq b_i \leq b_{i+1}$  para  $i = 1, 2, \dots, n-1$ . Tendremos entonces

$$\begin{aligned} \mu(I) &= F(b) - F(a) \leq \\ &\leq \sum_{i=1}^{n-1} [F(a_{i+1}) - F(a_i)] + F(b_n) - F(a_n) \leq \\ &\leq \sum_{i=1}^n [F(b_i) - F(a_i)] = \sum_{i=1}^n \mu(I_i). \end{aligned}$$

3. Es una consecuencia inmediata de los dos resultados anteriores.  $\square$

El segundo teorema permite extender la aditividad a familias numerables de intervalos disjuntos.

**Teorema 2.2** Con la notación anterior, si los  $I_i$ ,  $i \geq 1$  son disjuntos y se verifica  $I = \bigcup_{i \geq 1} I_i$ , entonces  $\mu(I) = \sum_{i \geq 1} \mu(I_i)$ .

**Demostración.-** Una primera desigualdad es inmediata a partir del teorema anterior pues  $\bigcup_{i=1}^n I_i \subset I$ ,  $\forall n$ , y de aquí  $\sum_{i=1}^n \mu(I_i) \leq \mu(I)$ .

Para establecer la desigualdad contraria, sea  $\epsilon > 0$  tal que  $a + \epsilon < b$ , entonces  $[a + \epsilon, b] \subset [a, b]$ . La continuidad por la derecha de  $F$  nos permite encontrar para cada  $i$  un  $b'_i$  tal que  $b_i < b'_i$  y  $\mu(I'_i) \leq \mu(I_i) + \epsilon/2^i$ , con  $I'_i = (a_i, b'_i]$ . Como  $[a + \epsilon, b] \subset \bigcup_{i \geq 1} I'_i$ , por el teorema de Heine-Borel la familia finita  $I'_{i_1}, \dots, I'_{i_m}$  cubrirá a  $[a + \epsilon, b]$ , y por el apartado 2 del anterior teorema

$$\mu([a + \epsilon, b]) \leq \sum_{j=1}^m \mu(I'_{i_j}) \leq \sum_{i \geq 1} \mu(I_i) + \epsilon.$$

La continuidad por la derecha de  $F$  hace el resto para  $\epsilon$  cada vez más pequeño.  $\square$

Es importante notar que en la demostración de los dos teoremas anteriores no hemos utilizado los valores límites de una función de distribución en  $-\infty$  y  $+\infty$ . Todo se mantiene válido quitando esa propiedad salvo que la extensión que obtenemos sea una medida de probabilidad. En particular, si tomamos  $F(x) = x$  se tiene  $\mu((a, b]) = b - a$ , es decir, la medida de Lebesgue de  $(a, b]$  y los dos teoremas anteriores constituyen la prueba de que esta función  $\mu$  verifica las condiciones del teorema 1.2 y tiene pues sentido hablar de medida de Lebesgue en  $\mathcal{R}$ .

Por otra parte la unicidad de la extensión nos permite establecer una correspondencia biunívoca entre la medida de probabilidad  $\mathcal{P}$ , definida sobre  $\mathcal{R}$ , y su función de distribución asociada.

## 2.2. Tipos de funciones de distribución

Las peculiaridades de la medida de probabilidad quedan reflejadas en la forma de la función de distribución asociada. Hay dos situaciones de especial interés que describimos a continuación.

### 2.2.1. El caso discreto

Cuando  $\mathcal{P}$  es discreta existe un conjunto  $D$ , llamado *soporte* de  $\mathcal{P}$ ,  $D = \{a_i : i \geq 1\}$  con  $\mathcal{P}(\{a_i\}) > 0$  y  $\mathcal{P}(D) = 1$ . En este caso su  $F$  asociada viene dada por

$$F(x) = \sum_{a_i \leq x} \mathcal{P}(\{a_i\}). \quad (2.3)$$

De acuerdo con esto, si  $a_{(i)}$  y  $a_{(i+1)}$  son dos valores consecutivos, cualquier  $x$  tal que  $a_{(i)} \leq x < a_{(i+1)}$ ,  $F(x)$  es constante e igual a  $F(a_{(i)})$  y por lo tanto continua. Por

otra parte, como cada  $a_i$  acumula masa de probabilidad, la función es discontinua en cada uno de los puntos de  $D$  con un salto cuyo valor es, de acuerdo con 2.2,  $F(a_{(i)}) - F(a_{(i-1)}) = \mathcal{P}(\{a_i\})$ .

Se trata por lo tanto de una función escalonada, cuyos saltos se producen en los puntos de  $D$ , y que denominaremos *función de distribución de probabilidad discreta*.

A la medida de probabilidad discreta podemos asociarle una nueva función puntual que nos será de gran utilidad. La definimos para cada  $x \in \mathcal{R}$  mediante

$$f(x) = \begin{cases} \mathcal{P}(\{a_i\}), & \text{si } x = a_i \\ 0, & \text{en el resto,} \end{cases}$$

función que es conocida como *función de cuantía o de probabilidad* y cuya relación con la función de distribución es evidente a través de 2.2 y 2.3.

**Ejemplo 2.1** *La distribución de Poisson de parámetro  $\alpha$  es una de las distribuciones de probabilidad discretas más conocida. Se caracteriza por que a los elementos de  $D = \{0, 1, 2, 3, \dots\}$  les asigna la probabilidad*

$$\mathcal{P}(\{n\}) = \frac{e^{-\alpha} \alpha^n}{n!},$$

de manera que su función de distribución viene dada, como es bien sabido, por

$$F(x) = \sum_{n \leq x} \frac{e^{-\alpha} \alpha^n}{n!}.$$

### 2.2.2. El caso continuo

Si la medida de probabilidad definida sobre los borelianos de  $\mathcal{R}$  es tal que  $\mathcal{P}(\{x\}) = 0, \forall x$ , de acuerdo con 2.2 la función de distribución asociada es *continua* y así la designaremos.

**Ejemplo 2.2** *La distribución uniforme en el intervalo  $[a, b]$  viene dada para cada conjunto de Borel,  $A$ , por*

$$\mathcal{P}(A) = \frac{\lambda(A \cap [a, b])}{b - a}, \text{ donde } \lambda \text{ es la medida de Lebesgue.}$$

Su función de distribución es de la forma,

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b, \end{cases}$$

que es efectivamente continua en todo  $x$ .

### 2.2.3. Densidad de probabilidad

En ocasiones existe una función  $f$ , no negativa e integrable Riemman, verificando  $\int_{-\infty}^{\infty} f(x)dx = 1$  y tal que

$$\mathcal{P}((a, b]) = \int_a^b f(x)dx, \quad (2.4)$$

para  $a$  y  $b$  reales cualesquiera. A esta función se la conoce como *la densidad de  $\mathcal{P}$* . La medida de probabilidad  $\mathcal{P}$  sobre toda la tribu de Borel está determinada por los valores de la ecuación 2.4, es decir, por la probabilidad de los intervalos y consecuentemente la medida de probabilidad está determinada por la función de densidad. Sin embargo, por estética y por facilidad de manejo de las medidas de probabilidad que verifican relaciones como la dada en la ecuación 2.4 es conveniente obtener alguna expresión directa que nos de la probabilidad de cualquier boreliano  $A$  en términos de la función de densidad. Observando 2.4, es de esperar que la probabilidad sea la integral sobre cada conjunto medible de la densidad. Sin embargo, la integral de Riemann solamente está definida para intervalos con extremos finitos o infinitos. Necesitamos un concepto de integral más amplio. En el capítulo siguiente trataremos esta cuestión.

¿Cómo es la función de distribución de una medida de probabilidad cuando existe una función de densidad? A partir de 2.4 y por las propiedades de la integral de Riemann,

$$F(x) = \int_{-\infty}^x f(t)dt; \quad (2.5)$$

de donde  $F'(x) = f(x)$  si  $x$  es un punto de continuidad de  $f$ . Con esta definición de  $F$  a partir de una densidad,  $F$  es continua (es fácil comprobar que  $\mathcal{P}(\{x\}) = 0, \forall x$ ). En realidad es más que continua como se deriva del siguiente resultado, cuya demostración veremos en el capítulo 8.

**Teorema 2.3** *Si  $F$  viene definida por 2.5,  $F$  es absolutamente continua, es decir, para cada  $\epsilon$  existe un  $\delta$  tal que para cada familia finita disjunta de intervalos  $[a_i, b_i], i = 1, \dots, k$*

$$\sum_{i=1}^k |F(b_i) - F(a_i)| < \epsilon \text{ si } \sum_{i=1}^k (b_i - a_i) < \delta.$$

## 2.3. El caso k-dimensional

La generalización inmediata de  $(\mathcal{R}, \mathcal{B}, \mathcal{P})$  es la probabilización de  $\mathcal{R}^k$  mediante el espacio de probabilidad  $(\mathcal{R}^k, \mathcal{B}^k, \mathcal{P})$ , donde  $\mathcal{B}^k$  es la correspondiente  $\sigma$  álgebra de Borel que puede venir engendrada, entre otros, por los rectángulos  $(a, b] = \prod_{i=1}^k (a_i, b_i]$ , con  $a = (a_1, \dots, a_k)$ ,  $b = (b_1, \dots, b_k)$  y  $-\infty \leq a_i < b_i < +\infty$ .

- La función de distribución asociada a la medida de probabilidad en  $\mathcal{R}^k$  se define para cada punto  $x = (x_1, \dots, x_k)$  mediante

$$F(x) = F(x_1, \dots, x_k) = \mathcal{P}(S_x) = \mathcal{P}\left(\prod_{i=1}^k (-\infty, x_i]\right),$$

donde  $S_x$ , denominado *región suroeste*, es igual a  $\prod_{i=1}^k (-\infty, x_i]$ . Esta función posee propiedades análogas a las del caso unidimensional.

1. **Es monótona creciente para cada componente.**
2. **Continuidad conjunta por la derecha.** Si  $x^{(n)} \downarrow x$ , entonces  $F(x^{(n)}) \downarrow F(x)$ ,
3.  $\lim_{\forall x_i \uparrow +\infty} F(x_1, \dots, x_k) = 1$  y  $\lim_{\exists x_i \downarrow -\infty} F(x_1, \dots, x_k) = 0$ .
4. Si para  $a_i \leq b_i$ ,  $\Delta_{a_i, b_i}$  representa el operador diferencia, actuando de la forma

$$\begin{aligned} \Delta_{a_i, b_i} F(x_1, \dots, x_k) &= \\ &= F(x_1, \dots, x_{i-1}, b_i, \dots, x_k) - F(x_1, \dots, x_{i-1}, a_i, \dots, x_k), \end{aligned}$$

sucesivas aplicaciones del mismo conducen a

$$\Delta_{a, b} F(x) = \Delta_{a_1, b_1}, \dots, \Delta_{a_k, b_k} F(x_1, \dots, x_k) = \mathcal{P}((a, b]),$$

y siendo la probabilidad mayor o igual que cero, la siguiente desigualdad es siempre cierta,

$$\Delta_{a, b} F(x) \geq 0,$$

$$\forall a, b \in \mathcal{R}^k, \text{ con } a_i \leq b_i.$$

- Podemos, como hicimos antes en  $\mathcal{R}$ , proceder a la inversa. Si tenemos una función  $F$  verificando las propiedades anteriores, la definición

$$\mathcal{P}((a, b]) = \Delta_{a, b} F(x),$$

permite obtener sobre la  $\sigma$  álgebra  $\mathcal{B}^k$  una medida de probabilidad que es además única. La demostración de este resultado puede ser objeto de un ejercicio complementario de prácticas. En cualquier caso, buenas exposiciones del mismo pueden consultarse en los textos de Billingsley y Shiryaev.

- Por último, como en el caso unidimensional, una segunda función puede asociarse a la medida de probabilidad,

- a)** En el caso *discreto* se trata de la *función de cuantía o probabilidad* y su valor en cada punto de  $\mathcal{R}^k$  viene dado por

$$f(x_1, \dots, x_k) = \begin{cases} \mathcal{P}(\{a^{(i)}\}), & \text{si } x = (x_1, \dots, x_k) = a^{(i)} \\ 0, & \text{en el resto,} \end{cases}$$

donde los  $a^{(i)}$ ,  $i \in N$ , son los elementos del *sopo*te de  $\mathcal{P}$ .

- b)** En ocasiones existe una función  $f$ , la función de densidad de  $\mathcal{P}$ , tal que

$$\mathcal{P}((a, b]) = \int_{a_k}^{b_k} \dots \int_{a_1}^{b_1} f(x_1, \dots, x_k) dx_1 \dots dx_k.$$

En particular,

$$F(x) = \mathcal{P}(S_x) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f(t_1, \dots, t_k) dt_1 \dots dt_k.$$

## Capítulo 3

# Variable Aleatoria. Esperanza

### 3.1. Aplicación medible

#### 3.1.1. Aplicaciones medibles: Definición general

**Definición 3.1** Sean  $(\Omega_1, \mathcal{A}_1)$  y  $(\Omega_2, \mathcal{A}_2)$  dos espacios medibles. Una aplicación  $f : \Omega_1 \rightarrow \Omega_2$  se dice que es medible, si para todo elemento  $A_2$  de  $\mathcal{A}_2$  se verifica que  $f^{-1}(A_2) \in \mathcal{A}_1$ .

Una consecuencia inmediata de esta definición es que la composición de aplicaciones medibles es medible, tal como se enuncia en la proposición siguiente cuya demostración omitimos por sencilla.

**Proposición 3.1** Sean  $(\Omega_1, \mathcal{A}_1)$ ,  $(\Omega_2, \mathcal{A}_2)$  y  $(\Omega_3, \mathcal{A}_3)$  tres espacios medibles y  $f_1 : \Omega_1 \rightarrow \Omega_2$  y  $f_2 : \Omega_2 \rightarrow \Omega_3$  sendas aplicaciones medibles. La composición de  $f_1$  y  $f_2$ ,  $f_2 \circ f_1 : \Omega_1 \rightarrow \Omega_3$  es también una aplicación medible.

Conocer si una aplicación es medible exige por tanto saber que las antimágenes de todos los elementos de la  $\sigma$  álgebra del espacio imagen están en la  $\sigma$  álgebra del espacio de partida. La siguiente proposición nos da una caracterización de la medibilidad en términos de una familia más reducida.

**Proposición 3.2** Sean  $(\Omega_1, \mathcal{A}_1)$  y  $(\Omega_2, \mathcal{A}_2)$  dos espacios medibles y sea  $\mathcal{C}$  una familia de partes de  $\Omega_2$  tal que  $\mathcal{A}_2$  es la  $\sigma$  álgebra engendrada por  $\mathcal{C}$ . Una aplicación  $f : \Omega_1 \rightarrow \Omega_2$  es medible si y solo si  $\forall C \in \mathcal{C}$ ,  $f^{-1}(C) \in \mathcal{A}_1$ .

**Demostración.-** Obviamente si  $f$  es medible  $f^{-1}(C) \in \mathcal{A}_1$ ,  $\forall C \in \mathcal{C}$ . A la inversa, si  $\forall C \in \mathcal{C}$ ,  $f^{-1}(C) \in \mathcal{A}_1$ , podemos escribir que  $f^{-1}(\mathcal{C}) \subset \mathcal{A}_1$  y también lo estará la  $\sigma$  álgebra que engendra. Pero siendo la antiimagen de  $\mathcal{A}_2$  una  $\sigma$  álgebra que coincide

con la engendrada por  $f^{-1}(\mathcal{C})$ , como fácilmente se comprueba, la medibilidad de  $f$  está demostrada.  $\square$

Si  $f$  es una aplicación de un espacio medible arbitrario,  $(\Omega, \mathcal{A})$  en  $(\mathcal{R}^k, \mathcal{B}^k)$  entonces tendremos asegurada su medibilidad simplemente comprobando que  $f^{-1}((a, b]) \in \mathcal{A}$  para cualquier rectángulo  $(a, b]$ .

### 3.1.2. Función medible

Un caso de especial interés es aquel en el que espacio medible imagen es la recta real o la recta real ampliada. Por recta real ampliada,  $\mathcal{R}^*$ , entendemos  $\mathcal{R} \cup \{-\infty\} \cup \{+\infty\}$  y la  $\sigma$  álgebra de Borel en  $\mathcal{R}^*$  la definimos como la engendrada por los conjuntos de Borel en  $\mathcal{R}$ ,  $\{-\infty\}$  y  $\{+\infty\}$ . Entonces,  $f : \Omega \rightarrow \mathcal{R}$  o  $g : \Omega \rightarrow \mathcal{R}^*$ , cuando son medibles, reciben el nombre de *función medible*. La medibilidad de las funciones, en el caso más general, puede caracterizarse mediante el siguiente resultado:

**Teorema 3.1** *Sea  $f$  una función de  $(\Omega, \mathcal{A})$  en  $(\mathcal{R}^*, \mathcal{B}^*)$ . Entonces  $f$  es medible si y solo si  $f^{-1}([-\infty, a]) \in \mathcal{A}$ ,  $\forall a \in \mathcal{R}^*$ .*

Fácilmente se comprueba que  $\mathcal{B}^*$  puede ser engendrada a partir de la familia de conjuntos  $[-\infty, a]$  con  $a \in \mathcal{R}^*$ .

**Observación 3.1** *La condición  $f^{-1}([-\infty, a]) \in \mathcal{A}$  puede también expresarse escribiendo que  $\{\omega : f(\omega) \leq a\} \in \mathcal{A}$ ,  $\forall a \in \mathcal{R}^*$ , forma en la que será también utilizada con frecuencia.*

Esta nueva forma de escribir la condición nos conduce a un resultado interesante del que también haremos uso posteriormente.

**Teorema 3.2** *Si  $f$  y  $g$  son dos funciones medibles  $\{\omega : f(\omega) \leq g(\omega)\} \in \mathcal{A}$ .*

**Demostración.-** Si  $\{r_n\}_{n \geq 1}$  representa los racionales, el complementario del conjunto puede escribirse como

$$\{\omega : f(\omega) > g(\omega)\} = \bigcup_{n \geq 1} [\{\omega : f(\omega) > r_n\} \cap \{\omega : g(\omega) < r_n\}],$$

y por la observación anterior el conjunto está en  $\mathcal{A}$ .  $\square$

### 3.1.3. Sucesiones de funciones medibles

En este apartado estudiaremos la medibilidad de las distintas funciones límite asociadas a una sucesión de funciones medibles. En todo cuanto sigue supondremos que las funciones medibles toman valores en la recta real extendida.



**Teorema 3.3** Sea  $f_n$ ,  $n \geq 1$ , una sucesión de funciones medibles, entonces:

1. las funciones  $f = \inf_n f_n$  y  $g = \sup_n f_n$  son medibles,
2. las funciones  $\liminf f_n$  y  $\limsup f_n$  son medibles y, en particular,  $\lim f_n$ , cuando existe, es medible,
3. el conjunto  $\{\omega : \exists \lim f_n(\omega)\} \in \mathcal{A}$ ,
4. si  $f$  es medible, el conjunto  $\{\omega : \lim f_n(\omega) = f(\omega)\} \in \mathcal{A}$ .

**Demostración.-** 1) Como fácilmente se comprueba,

$$\{\omega : \sup_n f_n(\omega) \leq a\} = \cap_{n \geq 1} \{\omega : f_n(\omega) \leq a\},$$

y siendo cada miembro de la intersección un elemento de  $\mathcal{A}$ ,  $\forall a$ , la función  $f = \sup_n f_n$  será medible.

Para el ínfimo podemos proceder de igual manera, teniendo en cuenta que

$$\{\omega : \inf_n f_n(\omega) \leq a\} = (\{\omega : \inf_n f_n(\omega) > a\})^c,$$

y que este último conjunto puede escribirse

$$\{\omega : \inf_n f_n(\omega) > a\} = \cap_{n \geq 1} \{\omega : f_n(\omega) > a\}.$$

2) Es una consecuencia inmediata de los resultados anteriores puesto que

$$\limsup f_n = \inf_k \sup_{n \geq k} f_n \quad y \quad \liminf f_n = \sup_k \inf_{n \geq k} f_n.$$

Por otra parte  $\lim f_n = \limsup f_n = \liminf f_n$  cuando existe.

3) El conjunto en cuestión se puede escribir de la forma

$$\{\omega : \exists \lim f_n(\omega)\} = \{\omega : \liminf f_n(\omega) = \limsup f_n(\omega)\},$$

y siendo  $\liminf f_n$  y  $\limsup f_n$  medibles, del teorema 3.2 se deduce el resultado.

4) Como en el caso anterior basta escribir el conjunto como

$$\begin{aligned} \{\omega : \lim f_n(\omega) = f(\omega)\} &= \\ &= \{\omega : \liminf f_n(\omega) = f(\omega)\} \cap \{\omega : \limsup f_n(\omega) = f(\omega)\}. \square \end{aligned}$$

Finalizaremos el apartado con un teorema de caracterización para las funciones medibles. Este resultado junto con el teorema de la convergencia monótona de Lebesgue, que veremos más tarde en este mismo capítulo, configuran una metodología que nos ha de ser muy útil en gran número de demostraciones posteriores.

Necesitamos primero introducir el concepto de *función simple*.

**Definición 3.2** Una función simple en una función medible con rango finito.

Si  $s$  es una función simple con rango  $\{a_1, \dots, a_n\}$  entonces admite la siguiente expresión.

$$s = \sum_{i=1}^n a_i \mathbf{1}_{A_i},$$

donde  $\{A_i, i = 1, \dots, n\}$  es una partición de  $\Omega$  mediante elementos de su  $\sigma$  álgebra  $\mathcal{A}$ , y  $\mathbf{1}_{A_i}$  la función indicatriz de  $A_i$ . A la inversa, toda función que pueda ser representada de la forma anterior es una función simple, por lo que en ocasiones la expresión anterior se toma como definición.

Nos ocupamos ahora del teorema de caracterización:

**Teorema 3.4** La función  $f$  es medible si y solo si existe una sucesión  $\{s_n\}$  de funciones simples que la tienen por límite. En concreto se tiene

$$\begin{aligned} 0 &\leq s_n(\omega) \uparrow f(\omega), & \text{si } f(\omega) \geq 0 \\ 0 &\geq s_n(\omega) \downarrow f(\omega), & \text{si } f(\omega) \leq 0. \end{aligned}$$

**Demostración.-** Para cada entero positivo  $n$  definimos sobre  $\Omega$  la siguiente función,

$$s_n(\omega) = \begin{cases} -n, & \text{si } -\infty \leq f(\omega) \leq -n, \\ \frac{-(m-1)}{2^n}, & \text{si } \frac{-m}{2^n} < f(\omega) \leq \frac{-(m-1)}{2^n}, \quad 1 \leq m \leq n2^n \\ \frac{(m-1)}{2^n}, & \text{si } \frac{m-1}{2^n} \leq f(\omega) < \frac{m}{2^n}, \quad 1 \leq m \leq n2^n \\ n, & \text{si } n \leq f(\omega) \leq \infty. \end{cases}$$

La sucesión que se obtiene al variar  $n$  es la buscada como fácilmente pero de forma laboriosa se puede comprobar.  $\square$

## 3.2. Variable aleatoria y vector aleatorio

### 3.2.1. Variable aleatoria

**Definición 3.3** Una variable aleatoria  $X$  es una función medible definida sobre un espacio de probabilidad.

El interés de una variable aleatoria reside en permitirnos obviar el espacio de probabilidad abstracto y trabajar sobre la recta real conservando las características probabilísticas originales. En otras palabras, una variable aleatoria nos permite trasladar la

información probabilística relevante de  $\Omega$  a  $\mathcal{R}$  a través de una medida de probabilidad inducida mediante  $X$  y que se conoce como la *distribución de probabilidad de  $X$* . Esta probabilidad inducida actúa sobre los borelianos de  $\mathcal{R}$  de la siguiente forma,

$$\mathcal{P}_X(A) = \mathcal{P}(X^{-1}(A)), \quad \forall A \in \mathcal{B}.$$

Es fácil comprobar que  $\mathcal{P}_X$  es una medida de probabilidad sobre la  $\sigma$  álgebra de Borel, de manera que  $(\mathcal{R}, \mathcal{B}, \mathcal{P}_X)$  es un espacio de probabilidad al que podemos aplicarle todo cuanto se dijo en el capítulo anterior. En particular podemos hablar de  $F_X$ , función de distribución asociada a  $\mathcal{P}_X$ , a la que denominaremos *función de distribución de  $X$*  y de acuerdo con su definición,

$$F_X(x) = \mathcal{P}_X([-\infty, x]) = \mathcal{P}(X \leq x).$$

Si la probabilidad inducida por  $X$  es *discreta*,  $f_X$ , *función de cuantía o de probabilidad de  $X$* , representará  $f_X(x) = \mathcal{P}(X = x)$ , mientras que si dicha probabilidad es absolutamente continua respecto de la medida de Lebesgue en  $\mathcal{R}$ ,  $f_X$ , *función de densidad de probabilidad de  $X$* , nos permitirá obtener

$$\mathcal{P}_X(A) = \mathcal{P}(X \in A) = \int_A f_X d\lambda, \quad A \in \mathcal{B}.$$

Las relaciones entre  $F_X$  y  $f_X$  son las que se indicaron en el capítulo anterior.

Partiendo de una variable aleatoria hemos probabilizado  $\mathcal{R}$ , pero es también posible proceder de manera inversa como nos muestra el siguiente teorema de existencia.

**Teorema 3.5** *Si sobre los borelianos de  $\mathcal{R}$  tenemos definida una medida de probabilidad  $\mathcal{P}$ , existe sobre algún espacio de probabilidad una variable aleatoria  $X$ , tal que  $\mathcal{P}_X = \mathcal{P}$ .*

**Demostración.-** Basta considerar  $X = I$ , la aplicación idéntica, y comprobar que sobre el  $\pi$  sistema de intervalos de la forma  $[a, b]$ , que engendra la  $\sigma$  álgebra de Borel,  $\mathcal{P}_X$  y  $\mathcal{P}$  coinciden. El teorema de extensión hace el resto.  $\square$

**Observación 3.2** *Inducir probabilidades, como se habrá ya intuido, no es virtud exclusiva de las variables aleatorias. A partir de cualquier aplicación medible  $f$ , entre  $(\Omega_1, \mathcal{A}_1, \mu_1)$  y  $(\Omega_2, \mathcal{A}_2)$ , podemos inducir una medida en el espacio imagen mediante una definición análoga a la anterior,*

$$\mu_f(A_2) = \mu_1(f^{-1}(A_2)), \quad \forall A_2 \in \mathcal{A}_2. \quad (3.1)$$

**Observación 3.3** *Cuando no exista riesgo de confusión prescindiremos del subíndice  $X$  en todo cuanto haga referencia a la variable aleatoria  $X$ .*

### 3.2.2. Vector aleatorio

**Definición 3.4** Un vector aleatorio  $X = (X_1, \dots, X_k)$  es una aplicación medible, a valores en  $\mathcal{R}^k$ , definida sobre un espacio de probabilidad.

Es posible caracterizar la medibilidad del vector en términos de las de sus componentes. En efecto,

**Teorema 3.6**  $X = (X_1, \dots, X_k)$  es un vector aleatorio si y solo si cada una de sus componentes es una variable aleatoria.

**Demostración.-** Los conjuntos  $S_x = \prod_{i=1}^k ]-\infty, x_i]$  del capítulo anterior forman un  $\pi$  sistema que engendra la  $\sigma$  álgebra de Borel en  $\mathcal{R}^k$ . Por otra parte podemos escribir

$$\{X \in S_x\} = \cap_{i=1}^k \{X_i \leq x_i\}, \quad (3.2)$$

y la medibilidad de cada  $X_i$  implicará la de  $X$ . Para demostrar el inverso bastará fijar un  $x_j$  y haciendo  $x_i = n$ ,  $i \neq j$  al tender  $n$  a  $\infty$  los conjuntos 3.2 tienden a  $\{X_j \leq x_j\}$  y por tanto  $X_j$  es medible.  $\square$

### Distribuciones conjuntas

Al igual que en el caso unidimensional induciremos sobre los borelianos de  $\mathcal{R}^k$ , a partir del vector  $X$ , una medida de probabilidad que denominaremos *distribución de probabilidad conjunta del vector  $X$* . Lo haremos mediante

$$\mathcal{P}_X(A) = \mathcal{P}(X^{-1}(A)) = \mathcal{P}((X_1, \dots, X_k) \in A), \quad \forall A \in \mathcal{B}^k.$$

Conseguimos que  $(\mathcal{R}^k, \mathcal{B}^k, \mathcal{P}_X)$  sea un espacio de probabilidad al que podemos aplicarle todo cuanto se dijo en el capítulo anterior referente al caso k-dimensional. La función de distribución  $F_X$  asociada a  $\mathcal{P}_X$  representa ahora,

$$\begin{aligned} F_X(x) &= F_X(x_1, \dots, x_k) = \mathcal{P}_X(S_x) = \\ &= \mathcal{P}(X \in S_x) = \mathcal{P}(X_i \leq x_i, \quad i = 1, \dots, k), \end{aligned}$$

y se denomina *función de distribución conjunta del vector  $X$*

Si la probabilidad inducida por el vector  $X$  es *discreta*,  $f_X$ , *función de cuantía o de probabilidad conjunta del vector  $X$* , representará  $f_X(x) = \mathcal{P}(X = x) = \mathcal{P}(X_i = x_i, \quad i = 1, \dots, k)$ , mientras que si dicha probabilidad es absolutamente continua respecto de la medida de Lebesgue k-dimensional,  $f_X$ , *función de densidad de probabilidad conjunta del vector  $X$* , nos permitirá obtener

$$\mathcal{P}_X(A) = \mathcal{P}(X \in A) = \int_A f_X d\lambda_k, \quad A \in \mathcal{B}^k.$$

Las relaciones entre  $F_X$  y  $f_X$  son las que se indicaron en el capítulo anterior.

### Distribuciones marginales

Cada componente  $X_i$  del vector es una variable aleatoria, tendrá por lo tanto su distribución de probabilidad a la que denominaremos *distribución marginal de  $X_i$*  para distinguirla de la conjunta del vector. De la misma forma, las funciones asociadas  $F_{X_i}$  y  $f_{X_i}$  se las conoce como las correspondientes *marginales*.

El concepto de *marginalidad* se aplica en realidad a cualquier subvector del vector original. Así, para  $l \leq k$ , si  $X^l = (X_{i_1}, \dots, X_{i_l})$  es un subvector de  $X$ , podemos hablar de la *distribución conjunta marginal de  $X^l$* . Por otra parte, si en la relación 3.2 fijamos  $x_{i_1}, \dots, x_{i_l}$ , igualamos las restantes componentes a  $n$  y hacemos tender  $n$  a  $\infty$  en todas ellas, obtendremos

$$\{X^l \in S_{x^l}\} = \lim_{\substack{i \neq i_1, \dots, i_l \\ x_i \rightarrow \infty}} \{X \in S_x\}. \quad (3.3)$$

Relación que nos permite obtener la función de distribución marginal a partir de la conjunta sin más que tomar probabilidades,

$$F_{X^l}(x_{i_1}, \dots, x_{i_l}) = \lim_{\substack{i \neq i_1, \dots, i_l \\ x_i \rightarrow \infty}} F_X(x_1, \dots, x_k). \quad (3.4)$$

En particular la marginal de cada componente se expresa

$$F_{X_i}(x_i) = \lim_{\substack{j \neq i \\ x_j \rightarrow \infty}} F_X(x_1, \dots, x_k). \quad (3.5)$$

Si la distribución del vector  $X$  es discreta también lo serán las marginales de cada componente. Si por  $D$  y  $D_i$  designamos los correspondientes soportes, tendremos la siguiente expresión para la función de cuantía marginal de  $X_i$  en términos de la conjunta,

$$f_{X_i}(x_i) = \sum_{x_j \in D_j, j \neq i} f_X(x_1, \dots, x_k), \quad (3.6)$$

cuya obtención es sencilla si recordamos el significado de  $f_{X_i}$ . Para la función de densidad marginal de  $X_i$ , la correspondiente expresión es

$$\begin{aligned} f_{X_i}(x_i) &= \\ &= \int_{\mathcal{R}^{k-1}} f_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_k) d\lambda_{k-1}, \end{aligned}$$

para cuya justificación necesitamos un resultado del próximo capítulo: el teorema de Fubini.

### 3.3. Integración de funciones medibles

#### 3.3.1. Funciones no negativas

**Definición 3.5** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $f : \Omega \rightarrow \mathcal{R}^*$  una función medible no negativa. Sea  $A \in \mathcal{A}$  y para cada partición finita  $\{E_j, j = 1, \dots, n\} \subset \mathcal{A}$  de  $A$  definimos la suma

$$\sum_{j=1}^n \inf_{\omega \in E_j} f(\omega) \mu(E_j), \quad (3.7)$$

siempre bien definida si convenimos que  $0 \times \infty = 0$ . Al supremo de las sumas obtenidas sobre  $\mathcal{C}_A$ , familia de todas las posibles particiones medibles finitas de  $A$ , le llamaremos integral de  $f$  sobre  $A$  respecto de  $\mu$ . Es decir,

$$\int_A f d\mu = \sup_{\mathcal{C}_A} \left\{ \sum_{j=1}^n \inf_{\omega \in E_j} f(\omega) \mu(E_j) \right\}. \quad (3.8)$$

**Ejemplo 3.1** Sea  $s = \sum_{i=1}^m a_i \mathbf{1}_{A_i}$  una función simple no negativa y vamos a calcular su integral sobre  $A$ . Sea  $\{E_j, j = 1, \dots, n\}$  una partición finita cualquiera de  $A$  y  $\beta_j = \inf_{\omega \in E_j} s(\omega)$ . La suma correspondiente a esta partición es

$$\sum_j \beta_j \mu(E_j) = \sum_{ij} \beta_j \mu(E_j \cap A_i).$$

Si  $E_j \cap A_i \neq \emptyset$  entonces  $\beta_j \leq a_i$  y podremos escribir

$$\sum_j \beta_j \mu(E_j) \leq \sum_{ij} a_i \mu(E_j \cap A_i) = \sum_i a_i \mu(A \cap A_i).$$

La integral, como supremo de estas sumas, está mayorada por el sumatorio de la derecha, pero dicho valor se alcanza si tomamos como partición de  $A$ ,  $\{A \cap A_i, i = 1, \dots, m\}$  y el valor de la integral será

$$\int_A s d\mu = \sum_i a_i \mu(A \cap A_i).$$

Un caso especialmente interesante es cuando  $s = \mathbf{1}_B$ , es la función característica del conjunto  $B \in \mathcal{A}$ . De acuerdo con el anterior resultado  $\int_A \mathbf{1}_B d\mu = \mu(A \cap B)$ .

### Propiedades de la integral de una función medible no negativa

Mientras no indiquemos lo contrario las funciones son medibles y no negativas.

**Teorema 3.7** *Se verifica:*

1. Si  $A_1 \subset A_2$  entonces  $\int_{A_1} f d\mu \leq \int_{A_2} f d\mu$ .
2. Si  $f \leq g$  entonces  $\int_A f d\mu \leq \int_A g d\mu$ .

#### Demostración

1. Si  $E^1 \in \mathcal{C}_{A_1}$  es una partición de  $A_1$  bastará añadirle el conjunto  $A_2 - A_1$  para obtener  $E^2 \in \mathcal{C}_{A_2}$ . La suma correspondiente a  $E^1$  será menor o igual que la de  $E^2$  y al tomar supremos tendremos la desigualdad buscada.
2. Siendo  $f \leq g$  tendremos  $\inf_{\omega \in E_j} f(\omega) \leq \inf_{\omega \in E_j} g(\omega)$ , desigualdad que se mantendrá para la suma de cada partición. Al tomar supremos aparecerá la desigualdad buscada.  $\square$

Si en  $\int_A f d\mu$  fijamos  $f$  y variamos  $A$  a través de la  $\sigma$  álgebra  $\mathcal{A}$  tenemos definida una función sobre la  $\sigma$  álgebra que como acabamos de ver en el teorema precedente es *no negativa* y *monótona*. En realidad se trata de una medida sobre  $\mathcal{A}$  como vemos a continuación.

**Teorema 3.8** *La función  $\nu_f(A) = \int_A f d\mu$ , definida sobre los conjuntos de  $\mathcal{A}$ , en una medida.*

**Demostración.-** Como de la definición de integral se deduce fácilmente que  $\nu_f(\emptyset) = 0$ , solo nos queda por demostrar la *aditividad numerable* para comprobar que se trata de una medida. Para ello, sea  $\{A_m, m \geq 1\}$  una familia disjunta de elementos de  $\mathcal{A}$  y sea  $A = \bigcup_m A_m$ . Consideremos una partición medible finita de  $A$ ,  $\{E_j, j = 1, \dots, n\}$ , que generará por intersección particiones para cada  $A_m$ ,  $\{E_j \cap A_m, j = 1, \dots, n\}$ . Se verifica

$$\int_{A_m} f d\mu \geq \sum_{j=1}^n \inf_{\omega \in E_j \cap A_m} f(\omega) \mu(E_j \cap A_m), \quad m = 1, 2, \dots,$$

que al sumar sobre  $m$  conduce a

$$\sum_{m \geq 1} \int_{A_m} f d\mu \geq \sum_{j=1}^n \inf_{\omega \in E_j} f(\omega) \mu(E_j).$$

Desigualdad que es válida para cualquier partición de  $A$ , por lo que al tomar supremos

$$\sum_{m \geq 1} \int_{A_m} f d\mu \geq \int_A f d\mu. \quad (3.9)$$

Para obtener la desigualdad contraria elegimos, dado  $\epsilon > 0$ , particiones para cada  $A_m$ ,  $\{E_j^m, j = 1, \dots, n_m\}$ , verificando

$$\int_{A_m} f d\mu - \frac{\epsilon}{2^m} \leq \sum_{j=1}^{n_m} \inf_{\omega \in E_j^m} f(\omega) \mu(E_j^m),$$

que al sumar sobre  $m$  conduce a

$$\sum_{m \geq 1} \int_{A_m} f d\mu - \epsilon \leq \sum_{m \geq 1} \sum_{j=1}^{n_m} \inf_{\omega \in E_j^m} f(\omega) \mu(E_j^m).$$

Por otra parte, la familia  $\{\{E_j^m\}_{j=1}^{n_m}\}_{m=1}^k$  en una partición finita del conjunto  $\bigcup_{m=1}^k A_m$  que está incluido en  $A$ ,  $\forall k$ . La monotonía de la integral y operaciones sencillas nos proporcionan la desigualdad

$$\sum_{m \geq 1} \int_{A_m} f d\mu - \epsilon \leq \int_A f d\mu,$$

que junto con 3.9 nos permite finalmente escribir

$$\nu_f\left(\bigcup_m A_m\right) = \sum_{m \geq 1} \nu_f(A_m). \square$$

**Ejemplo 3.2** Una situación interesante, que nos será útil posteriormente, es la integración de funciones respecto de medidas discretas. Si  $\mu$  es discreta su soporte  $D = \{\omega_1, \omega_2, \dots\}$  acumula toda la medida, de manera que  $\mu(D^c) = 0$ . En el supuesto que cada  $\omega_i$  sea medible, es decir,  $\{\omega_i\} \in \mathcal{A}$ ,  $\forall i$ , para  $f$  no negativa tenemos  $\int_A f d\mu = \int_{A \cap D} f d\mu$ , y si tenemos en cuenta que  $D$  es numerable y que la integral es una medida,

$$\int_A f d\mu = \sum_{\omega_i \in A \cap D} \int_{\{\omega_i\}} f d\mu = \sum_{\omega_i \in A \cap D} f(\omega_i) \mu(\{\omega_i\}). \quad (3.10)$$

Las dos propiedades que siguen demuestran que la integral actúa como un *operador lineal*.

**Teorema 3.9** Se verifica:

1. Para  $\alpha$  constante no negativa,  $\int_A \alpha f d\mu = \alpha \int_A f d\mu$ .
2.  $\int_A (f + g) d\mu = \int_A f d\mu + \int_A g d\mu$ .



**Demostración**

1. Es evidente sin más que tener en cuenta que  $\inf\{\alpha f(\omega)\} = \alpha \inf f(\omega)$ .
2. Para  $\epsilon > 0$  los conjuntos  $\{\omega : (1+\epsilon)^{n-1} < f(\omega) \leq (1+\epsilon)^n\}$ ,  $n = 0, \pm 1, \pm 2, \dots$ ,  $\{\omega : f(\omega) = 0\}$  y  $\{\omega : f(\omega) = +\infty\}$ , constituyen una partición del espacio  $\Omega$ . La intersección de cada uno de ellos con  $A$  dará lugar a una partición  $\{E_n^f\}_{|n| \geq 1}$ . Sustituyendo  $f$  por  $g$  obtendremos una nueva partición de  $A$ ,  $\{E_n^g\}_{|n| \geq 1}$ . A su vez, de la intersección de ambas se obtendrá  $\{E_k\}_{k \geq 1}$ . Como  $\nu_f(A) = \int_A f d\mu$  es una medida,

$$\begin{aligned}
\int_A (f + g) d\mu &= \\
&= \int_{\bigcup_{k \geq 1} E_k} (f + g) d\mu = \sum_{k \geq 1} \int_{E_k} (f + g) d\mu \leq \\
&\leq \sum_{k \geq 1} \left( \sup_{\omega \in E_k} f(\omega) + \sup_{\omega \in E_k} g(\omega) \right) \mu(E_k). \tag{3.11}
\end{aligned}$$

Por construcción de las particiones iniciales, se tiene en cada  $\omega \in \{E_n^f\}$ ,

$$\begin{aligned}
(1 + \epsilon)^{n-1} &\leq \inf f(\omega) \leq \sup f(\omega) \leq (1 + \epsilon)^n = \\
&= (1 + \epsilon)(1 + \epsilon)^{n-1} \leq (1 + \epsilon) \inf f(\omega),
\end{aligned}$$

y recordando la definición de  $E_k$ , para cada  $\omega \in E_k$  se cumple

$$\sup f(\omega) \leq (1 + \epsilon) \inf f(\omega), \quad \sup g(\omega) \leq (1 + \epsilon) \inf g(\omega).$$

Sustituyendo en 3.11, como la sucesión  $\{B_n = \bigcup_{k=1}^n E_k\}$  es expansiva, obtendremos

$$\int_A (f + g) d\mu \leq (1 + \epsilon) \left\{ \int_A f d\mu + \int_A g d\mu \right\}, \tag{3.12}$$

Para obtener la desigualdad contraria, a partir del  $\epsilon$  inicial podemos encontrar una partición de  $A$ ,  $\{E_j, j = 1, \dots, n\}$  verificando

$$\sum_{j=1}^n \inf_{\omega \in E_j} f(\omega) \mu(E_j) \geq \int_A f d\mu - \frac{\epsilon}{2},$$

y también

$$\sum_{j=1}^n \inf_{\omega \in E_j} g(\omega) \mu(E_j) \geq \int_A g d\mu - \frac{\epsilon}{2}.$$

Ambas desigualdades conducen fácilmente a

$$\int_A (f + g) d\mu \geq \int_A f d\mu + \int_A g d\mu - \epsilon,$$

que junto con 3.12 nos proporciona la igualdad enunciada.  $\square$

Es inmediato ahora el siguiente corolario:

**Corolario 3.1** *Para  $\alpha$  y  $\beta$  constantes no negativas,  $\int_A (\alpha f + \beta g) d\mu = \alpha \int_A f d\mu + \beta \int_A g d\mu$ .*

Las siguientes propiedades muestran como afectan al valor de la integral algunas relaciones casi por todas partes, *cpp*, entre las funciones. Una relación *cpp* es aquella que se verifica sobre todo  $\Omega$  salvo en un conjunto de medida nula. Todas estas propiedades se basan en la primera de ellas: que los conjuntos de medida  $\mu$ -nula no aportan nada al valor de la integral

**Teorema 3.10** 1. Si  $A \in \mathcal{A}$  es tal que  $\mu(A) = 0$  entonces  $\int_A f d\mu = 0$ ,  $\forall f$ .

2. Si  $f \leq g$  *cpp* entonces  $\int_A f d\mu \leq \int_A g d\mu$ ,  $\forall A \in \mathcal{A}$ .

3. Si  $f = g$  *cpp* entonces  $\int_A f d\mu = \int_A g d\mu$ ,  $\forall A \in \mathcal{A}$ .

4. Si  $f = 0$  *cpp* entonces  $\int_A f d\mu = 0$ ,  $\forall A \in \mathcal{A}$ .

5. Si  $\int_A f d\mu < +\infty$  entonces  $f < +\infty$  *cpp*.

Todas ellas son de fácil demostración.

### 3.3.2. Funciones cualesquiera

**Definición 3.6** Sea  $f$  una función medible cualquiera, diremos que  $f$  es integrable en  $A$  si las funciones  $f^+ = \sup(f, 0)$ , parte positiva de  $f$ , y  $f^- = -\inf(f, 0)$ , parte negativa de  $f$ , verifican  $\int_A f^+ d\mu < +\infty$  y  $\int_A f^- d\mu < +\infty$ . La integral de  $f$  sobre  $A$  se define entonces mediante

$$\int_A f d\mu = \int_A f^+ d\mu - \int_A f^- d\mu.$$

**Observación 3.4** Para la definición de la integral de  $f$  hubiéramos podido elegir cualquier otra descomposición de la forma  $f = f_1 - f_2$ , con  $f_1$  y  $f_2$  no negativas. Cumpliéndose las condiciones anteriores la integral de  $f$  sería ahora

$$\int_A f d\mu = \int_A f_1 d\mu - \int_A f_2 d\mu,$$

y como la descomposición de  $f$  no es única la definición puede parecernos incoherente. No hay tal incoherencia, pues si  $f = f^+ - f^- = f_1 - f_2$  entonces  $f^+ + f_2 = f_1 + f^-$  de donde por lo visto anteriormente

$$\int_A f^+ + f_2 d\mu = \int_A f_1 + f^- d\mu,$$

es decir,

$$\int_A f^+ d\mu + \int_A f_2 d\mu = \int_A f_1 d\mu + \int_A f^- d\mu,$$

y finalmente

$$\int_A f^+ d\mu - \int_A f^- d\mu = \int_A f_1 d\mu - \int_A f_2 d\mu.$$

**Observación 3.5** Cuando la función es no negativa, a diferencia de lo que ocurre en el caso general, su integral está siempre definida aún cuando pueda valer  $+\infty$ , pero solamente diremos que es integrable cuando dicha integral sea finita. Como para  $f \geq 0$ ,  $f^- = 0$ , obvia y coherentemente,  $\int_A f d\mu = \int_A f^+ d\mu$ .

Como  $|f| = f^+ + f^-$ , una consecuencia inmediata de la definición es que la integrabilidad de una función y su módulo son equivalentes:

**Corolario 3.2**  $f$  es integrable si y solo si  $|f|$  lo es.

### Propiedades de la integral de una función medible cualquiera

Aquellas que se derivan de las demostradas en el caso de funciones no negativas nos limitaremos a enunciarlas.

**Teorema 3.11** 1. Si  $f$  y  $g$  son integrables en  $A$  y  $\alpha$  y  $\beta$  son dos reales finitos cualesquiera, entonces  $\alpha f + \beta g$  es integrable y  $\int_A (\alpha f + \beta g) d\mu = \alpha \int_A f d\mu + \beta \int_A g d\mu$ .

2. Si  $f$  y  $g$  son integrables y  $f \leq g$  c.p.p. en  $A$ , entonces  $\int_A f d\mu \leq \int_A g d\mu$ .

La desigualdad que demostramos a continuación se la conoce con el nombre de *desigualdad modular*.

**Teorema 3.12** Si  $f$  es integrable en  $A$ , se verifica

$$\left| \int_A f d\mu \right| \leq \int_A |f| d\mu.$$

**Demostración.-** Basta recordar la monotonía de la integral y la cadena de desigualdades  $-|f| \leq f \leq |f|$ .  $\square$

### 3.3.3. Teoremas de convergencia

En este apartado estudiaremos las condiciones bajo las cuales es posible permutar la integración y el paso al límite. Comenzaremos con sucesiones de funciones no negativas con un resultado que se conoce como el *teorema de la convergencia monótona de Lebesgue*.

**Teorema 3.13** Si  $\{f_n\}_{n \geq 1}$  una sucesión de funciones tales que  $0 \leq f_n \uparrow f$  cpp, entonces  $\forall A \in \mathcal{A}$

$$\lim_{n \rightarrow \infty} \int_A f_n d\mu = \int_A f d\mu.$$

**Demostración.-** Supongamos en primer lugar que la sucesión es monótona y converge en todo  $\Omega$ . La función límite,  $f$ , será medible, no negativa y se verificará  $\forall A \in \mathcal{A}$ ,  $\int_A f_n d\mu \leq \int_A f d\mu$  y al tomar límites

$$\lim_{n \rightarrow \infty} \int_A f_n d\mu \leq \int_A f d\mu. \quad (3.13)$$

Buscaremos la desigualdad contraria.

Si  $\int_A f d\mu = +\infty$  entonces  $\exists s = \sum_{i=1}^m v_i \mu(A_i) = +\infty$  con  $v_i = \inf_{\omega \in A_i} f(\omega)$ , siendo los  $\{A_i\}$  una partición de  $A$ , y podremos encontrar un  $v_{i_0}$  y un  $A_{i_0}$  tales que  $v_{i_0} \mu(A_{i_0}) = +\infty$ , bien porque  $v_{i_0} = +\infty$  o porque  $\mu(A_{i_0}) = +\infty$ . Supongamos que  $0 < x < v_{i_0} \leq +\infty$ ,  $0 < y < \mu(A_{i_0}) \leq +\infty$  y sea  $A_{i_0 n} = \{\omega \in A_{i_0}; f_n(\omega) > x\}$ , como  $f_n \uparrow f$  esto supone que  $A_{i_0 n} \uparrow A_{i_0}$  y  $\exists n_0$  tal que  $\mu(A_{i_0 n}) > y$ ,  $\forall n \geq n_0$ . Entonces

$$\int_A f_n d\mu \geq \int_{A_{i_0 n}} f_n d\mu \geq x \mu(A_{i_0 n}) \geq xy, \quad \forall n \geq n_0$$

y  $\lim_n \int_A f_n d\mu \geq xy$ . Si  $v_{i_0} = +\infty$  al hacer  $x \rightarrow +\infty$  tendremos que  $\lim_n \int_A f_n d\mu = +\infty$ , si  $\mu(A_{i_0}) = +\infty$  entonces haremos  $y \rightarrow +\infty$  y tendremos el mismo resultado.

Nos ocupamos ahora del caso en que  $\int_A f d\mu < +\infty$ . A partir de un  $\epsilon$ ,  $0 < \epsilon < 1$ , definimos la sucesión  $A_n = \{\omega \in A : f_n(\omega) \geq (1 - \epsilon)f(\omega)\}$ , cuyos elementos son medibles y verifican  $A_n \subset A_{n+1}$  y  $\lim_{n \rightarrow +\infty} A_n = A$ . Para cada  $A_n$  tenemos  $\int_A f_n d\mu \geq \int_{A_n} f_n d\mu \geq (1 - \epsilon) \int_{A_n} f d\mu$ , y siendo la sucesión expansiva y la integral una medida, al tomar límites se obtiene la desigualdad

$$\lim_{n \rightarrow \infty} \int_A f_n d\mu \geq (1 - \epsilon) \int_A f d\mu,$$

que siendo válida para todo  $\epsilon$ , junto con 3.13, demuestra el teorema. Para demostrar ahora el teorema con toda generalidad consideremos el conjunto  $B = \{\omega : \exists n_0, f_{n_0}(\omega) > f_{n_0+1}(\omega)\} \cup \{\omega : \lim_{n \rightarrow +\infty} f_n(\omega) \neq f(\omega)\}$  y definamos la sucesión auxiliar  $\{g_n\}$  cuyos

miembros coinciden con los de la original sobre  $B^c$  y valen 0 en  $B$ . Con igual criterio definimos la función  $g$  respecto de  $f$ . Estas nuevas funciones verifican las condiciones anteriores, y por tanto el teorema, sobre el conjunto  $A \cap B^c$ ; pero como el conjunto  $B$  tiene medida nula sus integrales coincidirán con las de las funciones originales sobre  $A$ , lo que demuestra el teorema.  $\square$

Los dos corolarios que siguen son consecuencias inmediatas del teorema.

**Corolario 3.3** Si  $\{f_n\}_{n \geq 1}$  una sucesión de funciones tales que  $0 \leq f_n \downarrow f$  cpp, entonces  $\forall A \in \mathcal{A}$  con  $\int_A f_1 d\mu < +\infty$  el teorema se verifica.

**Corolario 3.4** Sea  $\{f_n\}_{n \geq 1}$  una sucesión de funciones no negativas y sea  $f = \sum_{n \geq 1} f_n$ , entonces  $\int_A f d\mu = \sum_{n \geq 1} \int_A f_n d\mu$ ,  $\forall A \in \mathcal{A}$ .

**Ejemplo 3.3 (La integral en un espacio numerable)** Sea  $(\Omega, \mathcal{A}, \mu)$  en espacio de medida con  $\Omega = \{1, 2, \dots\}$ , numerable,  $\mathcal{A} = \wp(\Omega)$  y  $\mu$  la medida de conteo que asigna a cada conjunto el número de puntos que contiene. En este espacio una función medible no es más que una sucesión de reales,  $f = \{x_1, x_2, \dots\}$ , cuya integral queremos calcular. Para ello consideremos la sucesión de funciones

$$f_n = \{x_{nm}\}_m = \begin{cases} x_{nm} = x_m, & \text{si } m \leq n \\ 0, & \text{si } m > n, \end{cases}$$

que también puede escribirse como  $f_n = \sum_{m=1}^{n+1} x_{nm} \mathbf{1}_{A_m}$ , con  $A_m = \{m\}$ , si  $m \leq n$  y  $A_{n+1} = \{n+1, n+2, \dots\}$ . La integral de  $f_n$  sobre  $\Omega$  valdrá  $\int_{\Omega} f_n d\mu = \sum_{m=1}^n x_m \mu(A_m) = \sum_{m=1}^n x_m$ .

Si nuestra  $f$  inicial es no negativa, la familia  $f_n \uparrow f$  y aplicando el teorema de la convergencia monótona podemos escribir  $\int_{\Omega} f d\mu = \lim_n \int_{\Omega} f_n d\mu = \sum_{m \geq 1} x_m$ . Para el caso general, una sucesión de términos cualesquiera, la función  $f$  será integrable si la serie  $\sum_{m \geq 1} |x_m|$  es convergente.

La convergencia para sucesiones de funciones cualesquiera, que abordaremos a través de dos teoremas, requiere previamente un resultado conocido como el *lema de Fatou*.

**Lema 3.1** Sea  $\{f_n\}_{n \geq 1}$  una sucesión de funciones no negativas, entonces  $\forall A \in \mathcal{A}$  se verifica

$$\int_A \liminf_{n \rightarrow +\infty} f_n d\mu \leq \liminf_{n \rightarrow +\infty} \int_A f_n d\mu.$$

**Demostración.-** El  $\liminf_{n \rightarrow +\infty} f_n = \sup_k g_k$  con  $\{g_k = \inf_{n \geq k} f_n\}_{k \geq 1}$  una sucesión monótona creciente cuyo límite,  $\sup_k g_k = \liminf_{n \rightarrow +\infty} f_n$ . Como para  $k$  fijo,  $g_k \leq f_n$ ,  $\forall n \geq k$ , podemos escribir

$$\int_A g_k d\mu \leq \inf_{n \geq k} \int_A f_n d\mu \leq \liminf_{n \rightarrow +\infty} \int_A f_n d\mu.$$

Pasando al límite para  $k$  y aplicando el teorema de la convergencia monótona se obtiene la desigualdad enunciada.  $\square$

El primer resultado relativo a sucesiones de funciones cualesquiera se conoce con el nombre de *teorema de la convergencia dominada de Lebesgue*.

**Teorema 3.14** Sea  $\{f_n\}_{n \geq 1}$  una sucesión de funciones integrables en  $A$  tales que  $\lim_{n \rightarrow +\infty} f_n = f$  cpp. Si existe  $g$ , integrable en  $A$ , verificando  $|f_n| \leq g$ ,  $\forall n$ , entonces  $f$  es integrable en  $A$  y  $\lim_{n \rightarrow +\infty} \int_A f_n d\mu = \int_A f d\mu$ .

**Demostración.-** La integrabilidad de  $f$  se deduce de  $|f| \leq g$  cpp. Como  $|f_n - f| \leq 2g$  cpp, a la sucesión  $\{2g - |f_n - f|\}_{n \geq 1}$ , cuyo límite es  $2g$ , podemos aplicarle el lema de Fatou para obtener,

$$\int_A 2g d\mu \leq \int_A 2g d\mu - \limsup_{n \rightarrow +\infty} \int_A |f_n - f| d\mu,$$

y deducir de aquí que  $\lim_{n \rightarrow +\infty} \int_A |f_n - f| d\mu = 0$ , que combinado con la desigualdad modular demuestra el teorema.  $\square$

La función  $g$  que domina a la sucesión puede ser sustituida por una constante  $M$  a condición de que el espacio tenga medida finita. Obtenemos así, como corolario del teorema, el llamado *teorema de la convergencia acotada de Lebesgue*.

**Corolario 3.5** Sea  $\{f_n\}_{n \geq 1}$  una sucesión de funciones integrables en  $A$  definidas en un espacio de medida finita y tales que  $\lim_{n \rightarrow +\infty} f_n = f$  cpp. Si existe una constante  $M$ , de manera que  $|f_n| \leq M$ ,  $\forall n$ , entonces  $f$  es integrable en  $A$  y  $\lim_{n \rightarrow +\infty} \int_A f_n d\mu = \int_A f d\mu$ .

### 3.3.4. Cambio de variable. Densidades

#### Cambio de variable en una integral

Si  $f$ , entre  $(\Omega_1, \mathcal{A}_1, \mu_1)$  y  $(\Omega_2, \mathcal{A}_2)$ , es una aplicación medible, 3.1 nos permite inducir sobre  $\mathcal{A}_2$  una medida. Si además  $g$  es una función medible definida sobre  $\Omega_2$ , la composición  $g \circ f$  será una función medible sobre  $\Omega_1$ . La relación entre sus integrales respecto de  $\mu_1$  y  $\mu_f$ , respectivamente, se recoge en el siguiente *teorema del cambio de variable*.

**Teorema 3.15** Una función  $g$  es integrable en  $A_2$  respecto de  $\mu_f$  si y solo si  $g \circ f$  lo es en  $f^{-1}(A_2)$  respecto de  $\mu_1$ , y en ese caso se verifica

$$\int_{f^{-1}(A_2)} g(f(\omega_1)) d\mu_1 = \int_{A_2} g(\omega_2) d\mu_f. \quad (3.14)$$

Si  $g$  es no negativa 3.14 es siempre cierta.

**Demostración.-** Si  $g = \mathbf{1}_B$ , 3.14 se reduce a 3.1. Si  $g$  es ahora una función simple no negativa, 3.14 también se verifica por linealidad de las integrales. Para  $g$  no negativa basta aplicar el teorema de la convergencia monótona a la sucesión de funciones simples que crecen hacia ella. Así pues 3.14 es siempre cierta para funciones no negativas.

La afirmación acerca de la integrabilidad de una  $g$  cualquiera se deduce aplicando 3.14 a  $|g|$ . La aplicación a  $g^+$  y  $g^-$  de 3.14 permite su extensión a  $g$ .  $\square$

### Integración bajo densidades

Hemos visto en el capítulo anterior que, bajo ciertas ocasiones que estudiaremos en el capítulo 8, las medidas de probabilidad poseen densidades respecto de la medida de Lebesgue. En general, dadas dos medidas  $\nu$  y  $\mu$  definidas sobre el mismo espacio medible, decimos que  $f$  es la *densidad de  $\nu$  respecto de  $\mu$* , si  $\nu(A) = \int_A f d\mu$ ,  $\forall A \in \mathcal{A}$ , con  $f \geq 0$ .

Las integrales respecto de medidas ligadas a través de una densidad guardan una sencilla relación.

**Teorema 3.16** *Si  $\nu$  posee una densidad  $f$  respecto de  $\mu$ , una función  $g$  es integrable en  $A$  respecto de  $\nu$  si y solo si  $gf$  lo es respecto de  $\mu$ , y en ese caso se verifica*

$$\int_A g d\nu = \int_A gf d\mu. \quad (3.15)$$

Si  $g$  es no negativa 3.15 es siempre cierta.

La demostración la omitimos por ser análoga a la del teorema anterior. Este método de demostración, basado en el teorema de caracterización de las funciones medibles y en el teorema de la convergencia monótona, es una herramienta potente y de fácil manejo a la que recurriremos con frecuencia.

#### 3.3.5. Esperanza de una variable aleatoria

**Definición 3.7** *Si  $X$ , variable aleatoria definida sobre el espacio de probabilidad  $(\Omega, \mathcal{A}, \mathcal{P})$ , es integrable en  $\Omega$  respecto de  $\mathcal{P}$  diremos que existe su esperanza o valor esperado que se define como*

$$E(X) = \int_{\Omega} X d\mathcal{P}.$$

Sea  $g$  es una función medible definida sobre  $(\mathcal{R}, \mathcal{B})$ , entonces  $g(X)$  es una nueva variable aleatoria cuya esperanza suponemos que existe. Si aplicamos 3.14, la fórmula del cambio de variable en la integral, con  $f = X$  y  $\mu_f = \mathcal{P}_X$  tendremos

$$E[g(X)] = \int_{\Omega} g(X) d\mathcal{P} = \int_{\mathcal{R}} g d\mathcal{P}_X, \quad (3.16)$$

expresión que nos permite obtener la esperanza como una integral sobre  $\mathcal{R}$ . Aún así, 3.16 no es operativa y solamente observando las características de la distribución de probabilidad de  $X$  obtendremos expresiones manejables. En efecto,

$\mathcal{P}_X$  es **discreta** De acuerdo con 3.10, que nos da la integral respecto de una medida discreta, si  $D$  es el soporte de  $\mathcal{P}_X$

$$E[g(X)] = \int_{\mathcal{R}} g d\mathcal{P}_X = \sum_{x_i \in D} g(x_i) \mathcal{P}_X(\{x_i\}) = \sum_{x_i \in D} g(x_i) \mathcal{P}(X = x_i). \quad (3.17)$$

Si en 3.17 tomamos para  $g$  la identidad obtendremos

$$E(X) = \sum_{x_i \in D} x_i \mathcal{P}(X = x_i).$$

$\mathcal{P}_X$  **posee densidad** Si  $f$  es la densidad de  $\mathcal{P}_X$  respecto de la medida de Lebesgue en  $\mathcal{R}$ , 3.15 nos permitirá escribir

$$E[g(X)] = \int_{\mathcal{R}} g d\mathcal{P}_X = \int_{\mathcal{R}} gf d\lambda. \quad (3.18)$$

Cuando, como ocurre con frecuencia,  $gf$  es integrable Riemann, 3.18 adopta la forma de la correspondiente integral de Riemann  $E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$ , que para el caso particular  $g(x) = x$  nos proporciona la  $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$ .

**Observación 3.6** Es costumbre escribir también 3.16 de la forma

$$E[g(X)] = \int_{\Omega} g(X) d\mathcal{P} = \int_{\mathcal{R}} g dF_X,$$

que se justifica porque  $\mathcal{P}_X$  es la medida de Lebesgue-Stieltjes engendrada por  $F_X$ .

### Momentos de una variable aleatoria

Para  $g(X) = |X|^k$  la  $E(|X|^k)$  recibe el nombre de *momento absoluto de orden  $k$*  de la variable aleatoria  $X$ . Si  $g(X) = X^k$  hablamos de *momento de orden  $k$* . Diremos que  $X$  *posee* momento de orden  $k$  si  $|X|^k$  es integrable.

Si  $X$  posee momento de orden  $k$  como  $|X|^j \leq 1 + |X|^k$ ,  $j \leq k$ , poseerá también todos los momentos de orden inferior.

El interés de los momentos de una variable aleatoria estriba en que son características numéricas que resumen su comportamiento probabilístico. Bajo ciertas condiciones el conocimiento de todos los momentos nos permite conocer completamente la distribución de probabilidad de la variable. Especialmente relevante es el caso  $k = 1$ , cuyo



correspondiente momento coincide con la esperanza de la variable y recibe el nombre de *media*, que a su vez da lugar a una nueva familia de momentos, los *momentos centrales de orden  $k$* ,  $E[(X - m)^k]$ , donde  $m$  es la media de  $X$ . De entre todos ellos cabe destacar la *varianza*,  $Var(X) = \sigma_X^2 = E[(X - m)^2]$ , que verifica la relación  $Var(X) = E[(X - m)^2] = E[X^2] - m^2$ . El siguiente resultado nos ofrece una forma alternativa de obtener la  $E(X)$  cuando  $X \geq 0$ .

**Proposición 3.3** Si  $X \geq 0$ ,

$$E(X) = \int_0^{+\infty} \mathcal{P}(X > t) dt = \int_0^{+\infty} \mathcal{P}(X \geq t) dt. \quad (3.19)$$

**Demostración.-** Dada  $X \geq 0$  existe una sucesión de funciones simples verificando  $0 \leq X_n \uparrow X$  y por el teorema de la convergencia monótona,  $E(X_n) \uparrow E(X)$ . Además  $\{X_n > t\} \uparrow \{X > t\}$ ,  $\forall t$ , por lo que al tomar probabilidades  $f_n(t) = \mathcal{P}(X_n > t) \uparrow f(t) = \mathcal{P}(X > t)$ ,  $\forall t$  y el teorema de la convergencia monótona también les será de aplicación. Si 3.19 se cumple para las funciones simples, el teorema estará demostrado. Veámoslo.

Sea  $X = \sum_{i=1}^n x_i \mathbf{1}_{A_i}$ ,  $x_i > 0$ ,  $i = 1, \dots, n$  y supongamos que los  $x_i$  están ordenados de menor a mayor, entonces  $f(t) = \mathcal{P}(X > t) = \sum_{j=i+1}^n P(A_j)$ ,  $x_i \leq t < x_{i+1}$ . Al integrarla se obtiene,

$$\int_0^{+\infty} f(t) dt = \sum_{i=0}^n [(x_{i+1} - x_i) \sum_{j=i+1}^n P(A_j)] = \sum_{i=1}^n x_i P(A_i) = E(X),$$

donde  $x_0 = 0$ .

Señalemos por último que las funciones  $\mathcal{P}(X > t)$  y  $\mathcal{P}(X \geq t)$  difieren a lo sumo en una cantidad numerable de valores de  $t$  y son por lo tanto iguales *cpp* respecto de la medida de Lebesgue, lo que justifica la igualdad en 3.19.  $\square$

### Desigualdades

Para  $X \geq 0$ , de 3.19 tenemos

$$E(X) = \int_0^{+\infty} \mathcal{P}(X \geq t) dt = \int_0^{\alpha} \mathcal{P}(X \geq t) dt + \int_{\alpha}^{+\infty} \mathcal{P}(X \geq t) dt,$$

pero  $\mathcal{P}(X \geq t) \geq \mathcal{P}(X \geq \alpha)$  para  $t \in [0, \alpha]$ , por lo que  $E(X) \geq \alpha \mathcal{P}(X \geq \alpha) + \int_{\alpha}^{+\infty} \mathcal{P}(X \geq t) dt$  y como  $\int_{\alpha}^{+\infty} \mathcal{P}(X \geq t) dt \geq 0$  tendremos que  $E(X) \geq \alpha \mathcal{P}(X \geq \alpha)$  y de aquí la desigualdad

$$\mathcal{P}(X \geq \alpha) \leq \frac{E(X)}{\alpha}. \quad (3.20)$$

Si en 3.20 sustituimos  $X$  por  $|X|^k$  obtenemos una nueva desigualdad,

$$\mathcal{P}(|X| \geq \alpha) = \mathcal{P}(|X|^k \geq \alpha^k) \leq \frac{1}{\alpha^k} E(|X|^k), \quad (3.21)$$

conocida como la *desigualdad de Markov*. Un caso especial de 3.21 se conoce como la *desigualdad de Chebyshev* y se obtiene para  $k = 2$  y  $X = X - E(X)$ ,

$$\mathcal{P}(|X - E(X)| \geq \alpha) \leq \frac{1}{\alpha^2} \text{Var}(X). \quad (3.22)$$

## Capítulo 4

# Independencia y medida producto

### 4.1. Independencia de sucesos y variables aleatorias

#### 4.1.1. Dos sucesos independientes

Desde el punto de vista del observador de un experimento el espacio de probabilidad,  $(\Omega, \mathcal{A}, \mathcal{P})$ , representa su descripción del mismo. Sea  $A \in \mathcal{A}$  un suceso cualquiera y supongamos que *antes* de realizar la experiencia preguntamos al observador, ¿qué probabilidad crees que hay de que se produzca el suceso  $A$ ? Si es consecuente con su descripción debe responder  $\mathcal{P}(A)$ . Si una vez realizado el experimento le damos la siguiente información parcial sobre lo ocurrido, el resultado del experimento pertenece al conjunto medible  $B$  tal que  $\mathcal{P}(B) > 0$ , y volvemos a preguntar al observador, ¿qué probabilidad crees que hay de que se haya producido el suceso  $A$ ? ¿Cuál podría ser una respuesta lógica por su parte? Sabe ahora que  $\omega$  no pertenece a  $\Omega - B$ . Si  $\mathcal{P}(A|B)$  denota la probabilidad que el observador le asigna a  $A$  dado su nivel actual de información entonces para  $A \subset B^c$  es inmediato que  $\mathcal{P}(A|B)$  ha de ser nula. Si  $A \cap B \neq \emptyset$  entonces es lógico que  $\mathcal{P}(A|B) = \mathcal{P}(A \cap B|B)$ . Lo que estamos haciendo básicamente consiste en *quitar* los resultados que no están en  $B$ ; ¿por qué no definir  $\mathcal{P}(A|B) = \mathcal{P}(A \cap B)$ ? La respuesta es una pregunta, ¿cuál es la probabilidad de  $B$  dado  $B$ ,  $\mathcal{P}(B|B)$ ? Lógicamente ha de ser uno pues el observador *sabe* que  $B$  ha ocurrido. Sin embargo, utilizando la anterior definición se sigue  $\mathcal{P}(B|B) = \mathcal{P}(B)$ , que no tiene porqué valer uno.

**Definición 4.1 (Probabilidad condicionada de un suceso a otro)** Si  $A, B \in \mathcal{A}$  con  $\mathcal{P}(B) > 0$ , definimos la probabilidad condicionada del suceso  $A$  al (o dado el) suceso

$B$  como

$$\mathcal{P}(A | B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}. \quad (4.1)$$

Fácilmente se comprueba que verifica las exigencias anteriores. Y lo que es más importante, si tomamos dos subconjuntos medibles de  $B$ ,  $C_1$  y  $C_2$ , si  $\mathcal{P}(C_1) = \alpha \mathcal{P}(C_2)$  con  $\alpha$  un número real no negativo entonces  $\mathcal{P}(C_1|B) = \alpha \mathcal{P}(C_2|B)$ . Dicho de otro modo el observador no ha mejorado su conocimiento del experimento *dentro* de  $B$ . Si  $C_1$  era  $\alpha$  veces tan probable como  $C_2$  *antes* de saber que  $\omega \in B$  sigue siendo  $\alpha$  veces tan probable como  $C_2$  *después* de conocer que  $\omega \in B$ .

Notemos que la expresión 4.1 solamente tiene sentido si el suceso  $B$  tiene probabilidad estrictamente positiva. En un capítulo posterior generalizaremos las ideas aquí presentadas de modo que podamos definir la probabilidad condicionada a sucesos de probabilidad nula.

Se comprueba sin dificultad que  $(\Omega, \mathcal{A}, \mathcal{P}(\cdot|B))$  es un espacio de probabilidad.

Si  $\mathcal{P}(A|B) \neq \mathcal{P}(A)$  es porque la ocurrencia de  $A$  está ligada a la ocurrencia de  $B$ ,  $A$  depende de  $B$ . Si  $\mathcal{P}(A|B) = \mathcal{P}(A)$  significa que al observador no le aporta ninguna información sobre  $A$  saber que  $B$  ha tenido lugar.  $A$  es independiente de  $B$ , pero entonces

$$\mathcal{P}(A) = \mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)},$$

es decir,

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B). \quad (4.2)$$

Lo que justifica la siguiente definición.

**Definición 4.2** Dos sucesos  $A$  y  $B$  de  $\mathcal{A}$  son independientes cuando verifican la ecuación 4.2.

Si  $A$  y  $B$  son independientes y  $\mathcal{P}(B) > 0$  entonces  $\mathcal{P}(A|B) = \mathcal{P}(A)$ . Si  $A$  y  $B$  son independientes y  $\mathcal{P}(A) > 0$  entonces  $\mathcal{P}(B|A) = \mathcal{P}(B)$ .

Observemos que con la definición dada de independencia estamos englobando la situación donde o bien  $\mathcal{P}(A)$  o bien  $\mathcal{P}(B)$  es nula.

Una familia finita de conjuntos medibles  $\{A_1, \dots, A_n\}$  se dice que son **independientes** cuando

$$\mathcal{P}(A_{k_1} \cap \dots \cap A_{k_m}) = \prod_{i=1}^m \mathcal{P}(A_{k_i}) \quad (4.3)$$

siendo  $\{k_1, \dots, k_m\} \subset \{1, \dots, n\}$  y los  $k_i$  distintos. Estamos expresando para  $n$  sucesos la misma idea que para dos, si de los  $n$  sucesos tomamos dos subcolecciones finitas disjuntas ninguna de ellas nos dice nada sobre la otra. Observemos que la independencia de  $n$  sucesos supone que han de verificarse  $\binom{n}{n} + \binom{n}{n-1} + \dots + \binom{n}{2} = 2^n - n - 1$  ecuaciones

del tipo dado en 4.3. En principio parece posible que si se dieran las igualdades que implican dos elementos las demás se verifiquen por añadidura. Esto no es cierto.

**Ejemplo 4.1 (Un experimento sencillo)** *Tenemos un tetraedro con una cara de color rojo, otra de color negro, la tercera blanca y la cuarta tiene los tres colores. Nuestro tetraedro está bien construido de modo que cuando lo tiramos sobre una mesa tenemos la misma probabilidad de que se apoye sobre cada una de las cuatro caras ( $\frac{1}{4}$ ). El experimento consiste en lanzar el tetraedro y ver en que posición ha caído.  $A$ ,  $B$  y  $C$  son los sucesos consistentes en que sale color rojo, negro y blanco respectivamente en la cara en que se apoya el tetraedro. Es inmediato que son independientes dos a dos pero no son independientes.*

Una colección infinita de sucesos diremos que son independientes cuando cualquier subcolección finita lo sea.

#### 4.1.2. Independencia de clases de sucesos

Sean  $A$  y  $B$  dos conjuntos medibles independientes.  $A$  no nos da información sobre si se ha producido  $B$ . Parece lógico que tampoco nos diga mucho sobre si  $B$  no se ha producido. ¿Son  $A$  y  $B^c$  independientes?

$$\begin{aligned}\mathcal{P}(A \cap B^c) &= \\ \mathcal{P}(A) - \mathcal{P}(A \cap B) &= \mathcal{P}(A) - \mathcal{P}(A)\mathcal{P}(B) = \mathcal{P}(A)(1 - \mathcal{P}(B)) = \mathcal{P}(A)\mathcal{P}(B^c).\end{aligned}$$

Como era de esperar. Del mismo se comprueba que  $A^c$  es independiente tanto de  $B$  como de  $B^c$ .

Las clases de sucesos  $\mathcal{L}_1, \dots, \mathcal{L}_n \subset \mathcal{A}$  se dicen independientes si para cualesquiera  $A_i \in \mathcal{L}_i$  con  $i = 1, \dots, n$  los sucesos  $\{A_1, \dots, A_n\}$  son independientes. Lo que hemos probado anteriormente es que  $\{A, A^c\}$  y  $\{B, B^c\}$  son dos clases de sucesos independientes.

Notemos que no pedimos que los elementos de cada clase  $\mathcal{L}_i$  sean independientes entre sí. Por ejemplo,  $A$  y  $A^c$  no tienen porque serlo (para que lo fueran la probabilidad de  $A$  ha de ser cero o uno).

**Definición 4.3** *Una colección infinita de clases de sucesos diremos que es independiente o que las clases son independientes cuando cualquier subcolección finita lo sea.*

Volvamos al caso en que tenemos dos sucesos  $A$  y  $B$  independientes. Hemos comprobado que  $\{A, A^c\}$  y  $\{B, B^c\}$  son clases independientes. Pero el conjunto vacío y el espacio muestral son siempre independientes de cualquier otro suceso,  $\mathcal{P}(\Omega \cap A) = \mathcal{P}(A) = \mathcal{P}(A)\mathcal{P}(\Omega)$  y  $\mathcal{P}(\emptyset \cap A) = \mathcal{P}(\emptyset) = 0 = \mathcal{P}(\emptyset)\mathcal{P}(A)$ . Es inmediato que  $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$

y  $\sigma(B) = \{\emptyset, B, B^c, \Omega\}$  son independientes. Esto era previsible, la información que nos da  $A^c$  es redundante con  $A$ ; si  $A$  se produce no se produce  $A^c$  y viceversa. El conjunto vacío y el espacio muestral son sucesos que no *aportan* ninguna información, sabemos que  $\Omega$  siempre se produce mientras que el conjunto vacío nunca puede darse.

#### 4.1.3. Independencia de variables y vectores aleatorios

La definición de independencia entre variables nos da un ejemplo concreto que justifica la definición dada de independencia entre clases.

$X$  e  $Y$  denotan dos variables aleatorias definidas sobre el mismo espacio de probabilidad,  $(\Omega, \mathcal{A}, \mathcal{P})$ . Una realización del experimento nos da un  $\omega$  concreto y dos valores de las variables,  $X(\omega) = x$  e  $Y(\omega) = y$ . ¿Qué conocemos ahora de  $\omega$ ? Dos valores numéricos,  $x$  e  $y$ . El conocimiento del valor que toma  $X$  para un  $\omega$  dado, ¿cambia nuestro conocimiento del valor que toma  $Y$ ?

¿Qué información nos proporciona  $X$ ? Toda variable aleatoria  $X$  engendra una  $\sigma$  álgebra,  $\sigma_X$ , definida mediante<sup>1</sup>

$$\sigma_X = \{X^{-1}(H), H \in \mathcal{B}\}.$$

Así, dado el valor de la variable podemos conocer para cada boreliano  $H$  si  $X \in H$  o  $X \in H^c$ . Del mismo modo tenemos interés en conocer acerca de  $Y$  las probabilidades de los elementos de  $\sigma(Y)$ . El concepto de independencia entre variables formaliza la idea de que la información que  $X$  nos proporciona no cambia las probabilidades que nos interesan acerca de la variable  $Y$ .

**Definición 4.4 (Variables aleatorias independientes)** *Las variables aleatorias  $X$  e  $Y$  se dicen independientes si las  $\sigma$  álgebras  $\sigma(X)$  y  $\sigma(Y)$  son independientes.*

Es decir,  $X$  e  $Y$  independientes equivale a

$$\mathcal{P}(X \in H, Y \in H') = \mathcal{P}(X \in H)\mathcal{P}(Y \in H') \quad \forall H, H' \in \mathcal{B}. \quad (4.4)$$

La independencia de un número finito o infinito de variables aleatorias se define análogamente a partir de la independencia de las correspondientes  $\sigma$  álgebras engendradas.

Comprobar que dos variables son independientes mediante 4.4 es prácticamente imposible. Necesitamos poder determinar de un modo más sencillo la independencia de variables. El problema lo resolvemos con el siguiente resultado.

**Teorema 4.1** *Sea  $(\Omega, \mathcal{A}, \mathcal{P})$  un espacio de probabilidad. Entonces,*

---

<sup>1</sup>El concepto de  $\sigma$  álgebra engendrada es, como se imagina el lector, más general y puede definirse a partir de cualquier aplicación medible

1. Si  $\mathcal{T}_1, \dots, \mathcal{T}_n \subset \mathcal{A}$  son  $\pi$  sistemas independientes entonces  $\sigma(\mathcal{T}_1), \dots, \sigma(\mathcal{T}_n)$  son  $\sigma$  álgebras independientes.
2. Si  $\mathcal{T}_\theta \subset \mathcal{A}$  para  $\theta \in \Theta$  es una familia de  $\pi$  sistemas independientes entonces  $\sigma(\mathcal{T}_\theta)$  con  $\theta \in \Theta$  son  $\sigma$  álgebras independientes.

**Demostración.-** Sea  $\mathcal{S}_i = \mathcal{T}_i \cup \{\Omega\}$ . Los  $\mathcal{S}_i$  siguen siendo  $\pi$  sistemas independientes. Tomamos  $B_2 \in \mathcal{S}_2, \dots, B_n \in \mathcal{S}_n$  fijos y consideramos la familia

$$\mathcal{L} = \{B_1 \in \sigma(\mathcal{S}_1) : \mathcal{P}(B_1 \cap \dots \cap B_n) = \prod_{i=1}^n \mathcal{P}(B_i)\}.$$

Es inmediato comprobar que  $\mathcal{L}$  es un  $\lambda$  sistema que contiene al  $\pi$  sistema  $\mathcal{S}_1$ . Por el teorema  $\pi - \lambda$  de Dynkin se sigue que  $\mathcal{L} = \sigma(\mathcal{S}_1) = \sigma(\mathcal{T}_1)$ . Es decir, hemos probado que  $\sigma(\mathcal{T}_1), \mathcal{S}_2, \dots, \mathcal{S}_n$  son independientes. De un modo análogo probamos que  $\sigma(\mathcal{T}_1), \sigma(\mathcal{T}_2), \mathcal{S}_3, \dots, \mathcal{S}_n$  son independientes y así sucesivamente.

La segunda parte del teorema es un corolario de la primera si tenemos en cuenta que la independencia de cualquier colección infinita de clases de sucesos se comprueba a partir de subcolecciones finitas.  $\square$

Utilizando este resultado obtenemos una caracterización simple de la independencia entre variables aleatorias.

### Independencia y función de distribución

Sean  $\mathcal{T}_1 = \mathcal{T}_2 = \{(-\infty, x], x \in \mathcal{R}\}$  que como vimos son  $\pi$  sistemas tales que  $\sigma(\mathcal{T}_1) = \sigma(\mathcal{T}_2) = \mathcal{B}$ , es decir, generan la tribu de Borel. Por el anterior teorema  $X$  e  $Y$  son variables aleatorias independientes y se verifica

$$\mathcal{P}(X \leq x, Y \leq y) = \mathcal{P}(X \leq x)\mathcal{P}(Y \leq y) \quad \forall x, y \in \mathcal{R}, \quad (4.5)$$

pero si lo que se verifica es la igualdad anterior, observemos que  $\sigma(X) = \sigma(X^{-1}(\mathcal{T}_1))$  y  $\sigma(Y) = \sigma(Y^{-1}(\mathcal{T}_2))$ , y al tomar  $A \in X^{-1}(\mathcal{T}_1)$  y  $B \in Y^{-1}(\mathcal{T}_2)$ ,  $A = \{X \leq x\}$  y  $B = \{Y \leq y\}$ , por lo que de la aplicación (4.5) se deduce la independencia de  $\sigma(X)$  y  $\sigma(Y)$ . En definitiva, la relación (4.5) es una manera alternativa de definir la independencia de las variables aleatorias  $X$  e  $Y$ . Si tenemos en cuenta que dicha relación no es más que

$$F(x, y) = F_1(x)F_2(y) \quad \forall x, y \in \mathcal{R}, \quad (4.6)$$

podremos también afirmar que  $X$  e  $Y$  son independientes si y solo si se verifica (4.6).

Del mismo modo podemos afirmar que si  $X_1, \dots, X_n$  son  $n$  variables sobre el mismo espacio de probabilidad serán independientes sí y solo sí

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n)$$

para cualesquiera números reales  $x_1, \dots, x_n$ , siendo  $F$  la función de distribución conjunta de  $(X_1, \dots, X_n)$  y cada  $F_i$  la correspondiente marginal de  $X_i$ .

### Independencia de una cantidad finita o infinita de vectores

Se define análogamente en términos de la independencia de las  $\sigma$  álgebras engendradas por dichos vectores.

Supongamos que  $Z_1, \dots, Z_n$  son vectores independientes y sea  $X_i$  una de las componentes de  $Z_i$ : ¿son  $X_1, \dots, X_n$  variables aleatorias independientes? Fácilmente se comprueba que sí.

Notemos que en la definición anterior no hacemos ninguna exigencia acerca de la dependencia o independencia entre las distintas componentes de un mismo vector.

#### 4.1.4. Funciones de variables independientes

Si  $X$  e  $Y$  son dos variables aleatorias independientes y  $f, g$  denotan dos funciones medibles parece natural que las nuevas variables aleatorias  $f(X)$  y  $g(Y)$  sigan siendo independientes ya que las transformaciones  $f$  y  $g$  que les aplicamos son deterministas. Esto es cierto. De la independencia de  $X$  e  $Y$  se sigue que  $\sigma(X)$  y  $\sigma(Y)$  son independientes, pero  $\sigma(f(X)) \subset \sigma(X)$  y  $\sigma(g(Y)) \subset \sigma(Y)$ . Extender el resultado al caso en que tenemos cualquier cantidad de variables o de vectores es inmediato, funciones de variables o vectores independientes siguen siendo independientes.

## 4.2. Espacio producto

Sean  $X$  e  $Y$  dos variables aleatorias independientes definidas sobre el mismo espacio de probabilidad,  $(\Omega, \mathcal{A}, \mathcal{P})$ ;  $\mathcal{P}_1$  y  $\mathcal{P}_2$  denotan sus distribuciones de probabilidad respectivas. Si consideramos conjuntamente las dos variables tenemos el vector  $Z = (X, Y)$  de  $(\Omega, \mathcal{A}, \mathcal{P})$  en  $(\mathcal{R}^2, \mathcal{B}^2)$  y cuya distribución de probabilidad denotamos por  $\mathcal{P}_{12}$ . Vamos a fijarnos en este espacio de probabilidad,  $(\mathcal{R}^2, \mathcal{B}^2, \mathcal{P}_{12})$ .

**En primer lugar** consideremos los conjuntos medibles,  $\mathcal{B}^2$ , que hemos definimos como la menor  $\sigma$  álgebra que contiene a los rectángulos  $(a, b]$  con  $a, b \in \mathcal{R}^2$ . Consideremos la siguiente  $\sigma$  álgebra sobre el mismo espacio

$$\mathcal{B} \times \mathcal{B} = \sigma(A \times B : A, B \in \mathcal{B})$$

a la que vamos a llamar  **$\sigma$  álgebra producto** de  $\mathcal{B}$  por  $\mathcal{B}$ . Estamos hablando de la misma  $\sigma$  álgebra.

**Lema 4.1**  $\mathcal{B} \times \mathcal{B} = \mathcal{B}^2$ .



**Demostración.-** Es inmediato que  $\mathcal{B}^2 \subset \mathcal{B} \times \mathcal{B}$ . Veamos la inclusión contraria. Sean  $a, b \in \mathcal{R}$  y definamos  $\mathcal{L} = \{B \in \mathcal{B}, (a, b] \times B \in \mathcal{B}^2\}$ .  $\mathcal{L}$  es un  $\lambda$ -sistema que contiene al  $\pi$ -sistema  $\{(a, b] : a, b \in \mathcal{R}\}$ . Aplicando el teorema  $\pi - \lambda$  de Dynkin tenemos  $\mathcal{L} = \mathcal{B}$ .

Fijamos  $B \in \mathcal{B}$  y definimos  $\mathcal{L}' = \{A \in \mathcal{B} : A \times B \in \mathcal{B}^2\}$ . Del mismo modo que en el caso anterior llegamos a que  $\mathcal{L}' = \mathcal{B}$ .  $\square$

A los conjuntos de la forma  $A \times B$  con  $A$  y  $B$  conjuntos de Borel los llamaremos **rectángulos medibles**. Es decir,  $\mathcal{B}^2$  es la menor  $\sigma$  álgebra que contiene a los rectángulos medibles.

**Observación 4.1** Una demostración análoga sirve para probar que  $\mathcal{B}^n = \mathcal{B} \times \dots \times \mathcal{B}$  siendo  $\mathcal{B} \times \dots \times \mathcal{B}$  la  $\sigma$  álgebra engendrada por los rectángulos medibles  $A_1 \times \dots \times A_n$ .

**En segundo lugar** vamos a plantearnos qué relación liga  $\mathcal{P}_{12}$  con  $\mathcal{P}_1$  y  $\mathcal{P}_2$ . Si tomamos el rectángulo medible  $A \times B$  entonces,

$$\mathcal{P}_{12}(A \times B) = \mathcal{P}_1(A)\mathcal{P}_2(B), \quad (4.7)$$

que en términos de las variables  $X$  e  $Y$  no nos dice mas que  $\mathcal{P}(X \in A, Y \in B) = \mathcal{P}(X \in A)\mathcal{P}(Y \in B)$ . ¿Es  $\mathcal{P}_{12}$  una medida de probabilidad? Sí, puesto que  $\mathcal{P}_{12}(C) = \mathcal{P}(Z^{-1}(C))$  para  $C \in \mathcal{B}^2$  y  $Z$  es una aplicación medible. Olvidémonos por un momento que existen las variables  $X$  e  $Y$  y por lo tanto podemos definir  $Z$ . Supongamos que tenemos únicamente los dos espacios de probabilidad  $(\mathcal{R}, \mathcal{B}, \mathcal{P}_1)$  y  $(\mathcal{R}, \mathcal{B}, \mathcal{P}_2)$  y que pretendemos definir en el espacio medible  $(\mathcal{R} \times \mathcal{R}, \mathcal{B} \times \mathcal{B})$  una función de conjunto que denotamos por  $\mathcal{P}_{12}$  verificando 4.7. La pregunta que pretendemos responder es, ¿podemos extender  $\mathcal{P}_{12}$  sobre  $\mathcal{B} \times \mathcal{B}$  de modo que tengamos una medida de probabilidad? El problema planteado no es artificial puesto que pretendemos mostrar cómo cuando tenemos dos variables aleatorias independientes,  $X$  e  $Y$ , la distribución conjunta de las mismas se puede obtener a partir de las marginales de cada variable. Las marginales determinan a la conjunta bajo la hipótesis de independencia.

#### 4.2.1. Medida producto

Sean  $(\Omega_i, \mathcal{A}_i, \mu_i)$  con  $i = 1, 2$  dos espacios de medida  $\sigma$  finitos. Consideremos el espacio medible  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$  siendo  $\mathcal{A}_1 \times \mathcal{A}_2 = \sigma(A \times B, A \in \mathcal{A}_1 \text{ y } B \in \mathcal{A}_2)$  la  $\sigma$  álgebra producto de  $\mathcal{A}_1$  y  $\mathcal{A}_2$  y  $A \times B$  un rectángulo medible.

En este apartado respondemos las siguientes preguntas:

1. ¿Existe sobre el espacio medible  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$  alguna medida tal que a los rectángulos  $A \times B$  le asocie el valor  $\mu_1(A) \times \mu_2(B)$ ?
2. ¿Es única?

Todo lo que sigue está basado en Billingsley y allí se puede encontrar material complementario. Necesitamos un resultado previo.

**Lema 4.2** *Se verifica,*

1. Si  $E \in \mathcal{A}_1 \times \mathcal{A}_2$  entonces  $\{x_1 : (x_1, x_2) \in E\} \in \mathcal{A}_1$  para cualquier  $x_2 \in \Omega_2$  y  $\{x_2 : (x_1, x_2) \in E\} \in \mathcal{A}_2$  para cualquier  $x_1 \in \Omega_1$ .
2. Si  $f$  denota una función medible definida sobre  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$  entonces la función que obtenemos fijando un valor de  $x_2$  y haciendo variar  $x_1$  es medible sobre  $(\Omega_1, \mathcal{A}_1)$ . Análogamente si fijamos  $x_1$  y variamos  $x_2$  obtenemos una función medible sobre  $(\Omega_2, \mathcal{A}_2)$ .

**Demostración.-** Sea  $x_1 \in \Omega_1$  fijo. Denotemos  $E_{x_1} = \{x_2 : (x_1, x_2) \in E\}$  para  $E \in \mathcal{A}_1 \times \mathcal{A}_2$ . Sea  $\mathcal{L} = \{E \in \mathcal{A}_1 \times \mathcal{A}_2 : E_{x_1} \in \mathcal{A}_2\}$ . Es inmediata la comprobación de que  $\mathcal{L}$  es un  $\lambda$ -sistema. Si  $E = A \times B$  con  $A \in \mathcal{A}_1$  y  $B \in \mathcal{A}_2$  entonces  $E_{x_1} = B$  si  $x_1 \in A$  o  $\emptyset$  si  $x_1 \in A^c$ . En definitiva  $\mathcal{L}$  contiene a los rectángulos medibles. Por el teorema  $\pi - \lambda$  de Dynkin,  $\mathcal{L} = \mathcal{A}_1 \times \mathcal{A}_2$ .

Sea  $f_{x_1} : \Omega_2 \rightarrow \mathcal{R}$  definida como  $f_{x_1} = f(x_1, x_2)$ . Para  $H \in \mathcal{B}$ ,  $f_{x_1}^{-1}(H) = [f^{-1}(H)]_{x_1}$ .  $\square$

Antes de definir la medida producto justifiquemos con un ejemplo su definición. Supongamos que estamos en  $(\mathcal{R} \times \mathcal{R}, \mathcal{B} \times \mathcal{B})$  y suponemos definida la medida de Lebesgue,  $\lambda$ , en cada uno de los espacios factores ( $\mu_1 = \mu_2 = \lambda$ ). Consideremos el rectángulo  $[a, b] \times [c, d]$ . Entonces  $\{x_2 : (x_1, x_2) \in [a, b] \times [c, d]\} = [c, d]$  si  $x_1 \in [a, b]$  y  $\emptyset$  en otro caso. Por tanto,  $\mu_2(\{x_2 : (x_1, x_2) \in [a, b] \times [c, d]\}) = \lambda([c, d]) = d - c$  si  $x_1 \in [a, b]$  y cero en otro caso de modo que

$$\begin{aligned} \int_{\Omega_1} \mu_2(\{x_2 : (x_1, x_2) \in [a, b] \times [c, d]\}) d\mu_1(x_1) &= \\ &= \int_{\mathcal{R}} \lambda(\{x_2 : (x_1, x_2) \in [a, b] \times [c, d]\}) d\lambda(x_1) = \int_{[a, b]} \lambda([c, d]) d\lambda(x_1) = (b - a)(d - c). \end{aligned}$$

Análogamente,

$$\int_{\Omega_2} \mu_1(\{x_1 : (x_1, x_2) \in [a, b] \times [c, d]\}) d\mu_2(x_2) = \int_{[c, d]} \lambda([a, b]) d\lambda(x_2) = (b - a)(d - c).$$

Dicho de otro modo, la medida producto del rectángulo la obtenemos integrando la medida de las secciones y además son intercambiables los papeles de los dos espacios coordenados. En este hecho tan simple se basa el concepto de medida producto.

### Definición de medida producto

Sea  $E \in \mathcal{A}_1 \times \mathcal{A}_2$ . Supongamos primero que  $\mu_1$  y  $\mu_2$  son finitas. Por el lema anterior la función que a cada  $x_1$  de  $\Omega_1$  le asocia  $\mu_2\{x_2 : (x_1, x_2) \in E\}$  y que podemos denotar  $f_E$  está bien definida. Además es medible, porque si  $E$  es el rectángulo medible  $A \times B$  entonces  $f_E = 1_A(x_1)\mu_2(B)$  que es medible.  $\mathcal{L} = \{E \in \mathcal{A}_1 \times \mathcal{A}_2 : f_E \text{ es medible}\}$  es un  $\lambda$ -sistema que, por lo que acabamos de comprobar, contiene a los rectángulos. Por el teorema  $\pi - \lambda$  de Dynkin,  $\mathcal{L} = \mathcal{A}_1 \times \mathcal{A}_2$ . Tiene sentido definir la siguiente función de conjunto,

$$\pi^{(12)}(E) = \int_{\Omega_1} \mu_2\{x_2 : (x_1, x_2) \in E\} d\mu_1(x_1). \quad (4.8)$$

De hecho, por las propiedades de la integral y considerando que  $\mu_2$  es una medida, es inmediato que  $\pi^{(12)}$  es una medida finita (por serlo  $\mu_1$  y  $\mu_2$ ) definida sobre  $\mathcal{A}_1 \times \mathcal{A}_2$ . Análogamente,

$$\pi^{(21)}(E) = \int_{\Omega_2} \mu_1\{x_1 : (x_1, x_2) \in E\} d\mu_2(x_2) \quad (4.9)$$

es también una medida finita sobre  $\mathcal{A}_1 \times \mathcal{A}_2$ . Pero  $\pi^{(12)}(A \times B) = \pi^{(21)}(A \times B) = \mu_1(A) \times \mu_2(B)$  y como el  $\pi$ -sistema formado por los rectángulos medibles que engendra la  $\sigma$  álgebra producto, aplicando el teorema de unicidad,  $\pi^{(12)}$  y  $\pi^{(21)}$  coinciden también sobre  $\mathcal{A}_1 \times \mathcal{A}_2$ .

Supongamos  $\mu_1$  y  $\mu_2$  medidas  $\sigma$  finitas en sus espacios respectivos; existen particiones  $\{A_n\}_{n \geq 1}$  de  $\Omega_1$  y  $\{B_m\}_{m \geq 1}$  de  $\Omega_2$  tales que  $\mu_1(A_n) < +\infty$  y  $\mu_2(B_m) < +\infty$  para todo  $n$  y  $m$ . Por ser particiones  $\mu_2(B) = \sum_{m \geq 1} \mu_2(B \cap B_m)$ . En el integrando de (4.8), definición de  $\pi^{(12)}$ , tenemos una suma de funciones medibles y por lo tanto una función medible. Análogamente con  $\pi^{(21)}$ .

Si  $\pi_{nm}^{(12)}$  denota la medida del tipo (4.8) que definimos utilizando las medidas  $\mu_1(A \cap A_n)$  y  $\mu_2(B \cap B_m)$  entonces es inmediato que  $\pi^{(12)}(E) = \sum_{n,m} \pi_{nm}^{(12)}(E)$  y por lo tanto  $\pi^{(12)}$  es una medida. Análogamente tendríamos  $\pi^{(21)}(E) = \sum_{n,m} \pi_{nm}^{(21)}(E)$ . Además,  $\pi^{(12)}(E) = \sum_{n,m} \pi_{nm}^{(12)}(E) = \sum_{n,m} \pi_{nm}^{(21)}(E) = \pi^{(21)}(E)$  ya que  $\pi_{nm}^{(12)} = \pi_{nm}^{(21)}$ . Tenemos pues definida una única medida. Pero  $\pi^{(12)}(A \times B) = \sum_{n,m} \mu_1(A \cap A_n) \mu_2(B \cap B_m) = \mu_1(A) \mu_2(B)$ .

Denotaremos esta medida producto de  $\mu_1$  y  $\mu_2$  por  $\mu_1 \times \mu_2$ .  $\mu_1 \times \mu_2$  es  $\sigma$  finita ya que  $\{A_n \times B_m\}_{n,m}$  recubren  $\Omega_1 \times \Omega_2$  y  $\mu_1 \times \mu_2(A_n \times B_m) = \mu_1(A_n) \times \mu_2(B_m)$  que es finita.

$\mu_1 \times \mu_2$  es la única medida que a los rectángulos le asocia el producto de la medida de sus lados. Esta afirmación es evidente pues cualquier otra medida sería  $\sigma$  finita sobre el  $\pi$ -sistema de los rectángulos medibles.

**Ejemplo 4.2 (Distribución conjunta de variables independientes)** Si  $X$  e  $Y$  son variables independientes y  $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_{12}$  son sus distribuciones marginales y conjunta respectivamente entonces,  $X$  e  $Y$  son independientes si y solo si  $\mathcal{P}_{12} = \mathcal{P}_1 \times \mathcal{P}_2$ . ¿Hemos ganado algo? En Probabilidad no interesan tanto las variables en si mismas como las distribuciones que inducen. Variables que tienen diferentes definiciones pero con la misma distribución son, probabilísticamente hablando, una misma cosa. Mediante la construcción propuesta tenemos la distribución conjunta,  $\mathcal{P}_{12}$ , de cualquier par de variables aleatorias independientes cuyas distribuciones marginales sean,  $\mathcal{P}_1$  y  $\mathcal{P}_2$ .

Un ejemplo de lo anterior. Si  $((0, 1], \mathcal{B}((0, 1]), \lambda)$  representa el experimento consistente en obtener un punto aleatorio uniforme en el intervalo unitario y  $k$  es un entero positivo mayor que la unidad entonces  $((0, 1]^k, \mathcal{B}((0, 1]) \times \dots \times \mathcal{B}((0, 1]), \lambda \times \dots \times \lambda)$  representa el experimento consistente en obtener un punto aleatorio uniforme en el cubo (o hipercubo)  $(0, 1]^k$ .

**Ejemplo 4.3 (Otra construcción de la medida de Lebesgue en  $\mathcal{R}^k$ )** La medida de Lebesgue de dimensión  $k$  (con  $k \geq 2$ ),  $\lambda_k$ , fue definida como aquella tal que

$$\lambda_k((a, b]) = \prod_{i=1}^k (b_i - a_i) \text{ siendo } a, b \in \mathcal{R}^k.$$

Lo que la anterior igualdad afirma es que  $\lambda_k((a, b]) = \lambda \times \dots \times \lambda((a, b])$  y como  $\{(a, b] : a, b \in \mathcal{R}^k\}$  es un  $\pi$ -sistema que engendra  $\mathcal{B}^k$  tenemos que  $\lambda_k = \lambda \times \dots \times \lambda$ .

#### 4.2.2. Teorema de Fubini

En este apartado pretendemos justificar el sentido de una frase célebre en Matemáticas que afirma que *la integral doble es igual a las integrales iteradas*.

Como en el apartado anterior,  $(\Omega_i, \mathcal{A}_i, \mu_i)$  con  $i = 1, 2$  son dos espacios de medida  $\sigma$  finitos y  $f$  una función medible definida en el espacio producto  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ . Formalmente la frase anterior se expresa con las dos igualdades siguientes

$$\int_{\Omega_1 \times \Omega_2} f(x_1, x_2) d\mu_1(x_1) \times d\mu_2(x_2) = \int_{\Omega_1} \left[ \int_{\Omega_2} f(x_1, x_2) d\mu_2(x_2) \right] d\mu_1(x_1), \quad (4.10)$$

$$\int_{\Omega_1 \times \Omega_2} f(x_1, x_2) d\mu_1(x_1) \times d\mu_2(x_2) = \int_{\Omega_2} \left[ \int_{\Omega_1} f(x_1, x_2) d\mu_1(x_1) \right] d\mu_2(x_2). \quad (4.11)$$

La integral doble se refiere al miembro de la izquierda en las dos igualdades, la integral de la función  $f$  respecto de la medida producto, mientras que las integrales iteradas hacen referencia a los dos miembros de la derecha en 4.10 y 4.11. Vamos a ocuparnos de 4.10. La prueba de 4.11 es similar.

Aparece la integral respecto  $\mu_1$  de la siguiente función de  $x_1$ ,

$$\int_{\Omega_2} f(x_1, x_2) d\mu_2(x_2). \quad (4.12)$$

Como vimos en el lema previo la función que estamos integrando (con  $x_2$  variable) es medible por lo que la integral tiene sentido. Una vez hemos integrado respecto  $\mu_2$  tenemos una función de  $x_1$  que puede tomar valores finitos, valores infinitos o ni tan siquiera estar definida. Nuestro primer problema es imponer condiciones a la función  $f$  para que 4.12 como función de  $x_1$  esté definida (aunque tome un valor infinito), sea medible (respecto  $\mathcal{A}_1$ ) y, por último, verifique la igualdad 4.10.

- Si  $f = 1_E$  con  $E \in \mathcal{A}_1 \times \mathcal{A}_2$  entonces 4.12 es  $\mu_2\{x_2 : (x_1, x_2) \in E\}$  que ya comprobamos que era medible y 4.10 no es más que una de las definiciones de la medida producto.
- Si  $f$  es una función simple no negativa entonces 4.12 es una combinación lineal de funciones medibles y por lo tanto medible. Por lo visto para funciones indicatrices y la linealidad de la integral se tiene 4.10.
- Si  $f$  es una función medible no negativa entonces sabemos que existe una sucesión no decreciente de funciones simples que converge a  $f$ . Tanto la medibilidad de 4.12 como la igualdad 4.10 son consecuencias inmediatas del teorema de la convergencia monótona. En ninguno de los casos considerados hasta ahora hemos afirmado que 4.12 tome valores finitos (pueden tomar el valor  $+\infty$ ) ni que 4.10 sea por lo tanto una igualdad entre valores finitos (podemos tener la igualdad  $+\infty = +\infty$ ).
- Si  $f$  es una función integrable entonces  $|f|$  también lo es. Se verifica 4.10 para  $|f|$ , es decir,

$$\int_{\Omega_1} \left[ \int_{\Omega_2} |f(x_1, x_2)| d\mu_2(x_2) \right] d\mu_1(x_1) < +\infty, \quad (4.13)$$

de donde  $\int_{\Omega_2} |f(x_1, x_2)| d\mu_2(x_2)$  es finita sobre un subconjunto  $\Lambda_1$  de  $\Omega_1$  tal que  $\mu_1(\Omega_1 - \Lambda_1) = 0$ . Si  $x_1 \in \Lambda_1$  entonces

$$\int_{\Omega_2} f(x_1, x_2) d\mu_2(x_2) = \int_{\Omega_2} f^+(x_1, x_2) d\mu_2(x_2) - \int_{\Omega_2} f^-(x_1, x_2) d\mu_2(x_2). \quad (4.14)$$

Notemos que por lo visto para funciones no negativas las integrales de la derecha en 4.14 son funciones sobre  $\Lambda_1$  (considerando la restricción de la  $\sigma$  álgebra  $\mathcal{A}_1$  a  $\Lambda_1$ ) y además a valores finitos. Por tanto 4.12 es medible sobre  $\Lambda_1$ . Integrando los

dos miembros de 4.14 sobre  $\Lambda_1$  por lo demostrado para funciones no negativas se sigue 4.10 sustituyendo  $\Omega_1$  por  $\Lambda_1$ .

Ya que  $\Lambda_1 \times \Omega_2$  tiene medida producto nula se tiene

$$\int_{\Omega_1 \times \Omega_2} f(x_1, x_2) d\mu_1(x_1) \times d\mu_2(x_2) = \int_{\Lambda_1 \times \Omega_2} f(x_1, x_2) d\mu_1(x_1) \times d\mu_2(x_2).$$

Si a 4.12 le damos un valor constante arbitrario sobre  $\Omega_1 - \Lambda_1$  entonces es medible sobre  $\mathcal{A}_1$  y

$$\int_{\Omega_1} \left[ \int_{\Omega_2} f(x_1, x_2) d\mu_2(x_2) \right] d\mu_1(x_1) = \int_{\Lambda_1} \left[ \int_{\Omega_2} f(x_1, x_2) d\mu_2(x_2) \right] d\mu_1(x_1),$$

lo que prueba que 4.10 es cierta.

**Teorema 4.2 (de Fubini)**  $(\Omega_i, \mathcal{A}_i, \mu_i)$  con  $i = 1, 2$  son dos espacios de medida  $\sigma$  finitos entonces,

1. Si  $f$  es una función medible no negativa sobre  $\Omega_1 \times \Omega_2$  las funciones

$$\int_{\Omega_2} f(x_1, x_2) d\mu_2(x_2) \text{ y } \int_{\Omega_1} f(x_1, x_2) d\mu_1(x_1) \quad (4.15)$$

son medibles sobre  $(\Omega_1, \mathcal{A}_1)$  y  $(\Omega_2, \mathcal{A}_2)$  respectivamente. Además se verifican las igualdades 4.10 y 4.11.

2. Si suponemos  $f$  una función integrable respecto  $\mu_1 \times \mu_2$  entonces las funciones 4.15 son medibles y finitas sobre conjuntos  $\Lambda_1$  y  $\Lambda_2$  respectivamente tales que  $\mu_1(\Lambda_1^c) = \mu_2(\Lambda_2^c) = 0$ . Son ciertas las igualdades 4.10 y 4.11.

### 4.3. Independencia entre variables y función de densidad

$X$  e  $Y$  son dos variables aleatorias discretas sobre un mismo espacio de probabilidad.  $f, f_1$  y  $f_2$  son las funciones de probabilidad conjunta y las marginales de  $X$  e  $Y$  respectivamente. Si  $X$  e  $Y$  son independientes entonces  $\mathcal{P}(X = x, Y = y) = \mathcal{P}(X = x)\mathcal{P}(Y = y)$ , es decir,  $f(x, y) = f_1(x)f_2(y)$ .

¿Es cierto el recíproco? ¿Si podemos factorizar la función de probabilidad conjunta como producto de marginales entonces son independientes las variables? Para  $A, B \in \mathcal{B}$ ,

$$\begin{aligned} \mathcal{P}(X \in A, Y \in B) &= \sum_{x \in A} \sum_{y \in B} f(x, y) = \sum_{x \in A} \sum_{y \in B} f_1(x) f_2(y) \\ &= \sum_{x \in A} f_1(x) \sum_{y \in B} f_2(y) = \mathcal{P}(X \in A) \mathcal{P}(Y \in B). \end{aligned}$$

Supongamos ahora que  $X$  e  $Y$  son tales que su distribución conjunta,  $\mathcal{P}_{12}$ , tienen densidad  $f$  respecto de la medida de Lebesgue de dimensión dos,  $\lambda_2$ . Las funciones  $f_1(x) = \int_{\mathcal{R}} f(x, y) dy$ ,  $f_2(y) = \int_{\mathcal{R}} f(x, y) dx$  son finitas casi por todas partes. En los valores que no lo sean les damos un valor constante positivo. Bajo las condiciones indicadas se tiene el siguiente resultado.

**Proposición 4.1**  *$X$  e  $Y$  son independientes si y solo si  $f(x, y) = f_1(x)f_2(y)$  cpp.*

**Demostración.-**  $X$  e  $Y$  son independientes es equivalente con  $\mathcal{P}(X \in A, Y \in B) = \mathcal{P}(X \in A)\mathcal{P}(Y \in B)$  para todo  $A, B \in \mathcal{B}$ . Expresándolo en términos de las funciones de densidad tenemos

$$\int_{A \times B} f(x, y) dx dy = \int_A f_1(x) dx \int_B f_2(y) dy, \quad \forall A, B \in \mathcal{B}.$$

Por el teorema de Fubini,

$$\int_A f_1(x) dx \int_B f_2(y) dy = \int_{A \times B} f_1(x) f_2(y) dx dy,$$

de modo que la independencia de  $X$  e  $Y$  es equivalente a

$$\int_{A \times B} f(x, y) dx dy = \int_{A \times B} f_1(x) f_2(y) dx dy, \quad \forall A, B \in \mathcal{B}. \quad (4.16)$$

Tenemos pues que las medidas de probabilidad  $\mathcal{P}$  y  $\mathcal{P}'$ , con densidades  $f$  y  $f_1 \times f_2$  respectivamente, coinciden sobre el  $\pi$ -sistema de los rectángulos  $\{A \times B : A \in \mathcal{B}, B \in \mathcal{B}\}$  y, al ser ambas finitas, coincidirán sobre la  $\sigma$ -álgebra engendrada por dicho  $\pi$ -sistema que es la  $\sigma$ -álgebra producto  $\mathcal{B}^2$ . Tendremos pues

$$\int_E f(x, y) dx dy = \int_E f_1(x) f_2(y) dx dy, \quad \forall E \in \mathcal{B}^2,$$

y aplicando un conocido resultado que afirma que para  $f$  y  $g$  no negativas y  $\mu$   $\sigma$ -finita, si  $\int_A f d\mu = \int_A g d\mu$ ,  $\forall A \in \mathcal{A}$ , entonces  $f = g$  cpp, se sigue el resultado.  $\square$

Resumiendo, *las variables aleatorias  $X_1, \dots, X_n$  con función de densidad conjunta (respectivamente variables discretas con función de probabilidad conjunta)  $f(x_1, \dots, x_n)$  y marginales  $f_1(x_1), \dots, f_n(x_n)$  son independientes si y sólo si*

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$$

*cpp (respectivamente para todos los  $x_1, \dots, x_n$ ).*

**Ejemplo 4.4 (Una sucesión de variables aleatorias independientes)** *Consideremos el espacio de probabilidad correspondiente a la elección de un punto al azar en el intervalo unitario,  $((0, 1], \mathcal{B}((0, 1]), \lambda)$ . Si  $\omega \in (0, 1]$ ,  $X_n(\omega)$  denota el  $n$ -ésimo elemento del desarrollo binario de  $\omega$ , Por tanto  $\omega = \sum_{n \geq 1} \frac{X_n(\omega)}{2^n}$  con  $X_n = 0$  ó  $1$ . No siempre es único el desarrollo binario de un número, por ejemplo si  $\omega = 1/2$  podemos tomar o  $X_1(\omega) = 1$  y  $X_n(\omega) = 0$  para  $n > 1$  o  $X_1(\omega) = 0$  y  $X_n(\omega) = 1$  para  $n > 1$ . Cuando se plantea esta situación tomamos como desarrollo el segundo. ¿Son las funciones  $X_n$  variables aleatorias? y si lo son ¿son independientes?*

$$X_1 = 1_{(1/2, 1]}, \quad X_2 = 1_{(1/4, 1/2] \cup (3/4, 1]}, \dots$$

y por lo tanto las  $X_n$  son efectivamente variables aleatorias. También son independientes pues

$$\{X_1 = x_1, \dots, X_n = x_n\} = \left( \sum_{i=1}^n \frac{x_i}{2^i}, \sum_{i=1}^n \frac{x_i}{2^i} + \frac{1}{2^n} \right]$$

siendo  $x_i = 0$  ó  $1$ . De esta igualdad, y porque  $\omega$  es uniforme en  $(0, 1]$ , deducimos  $\lambda(\{X_1 = x_1, \dots, X_n = x_n\}) = \lambda(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{2^n}$ . Por la expresión vista para las  $X_i$  se tiene que  $\mathcal{P}(X_i = x_i) = \frac{1}{2}$ . Se verifica pues la igualdad  $\mathcal{P}(X_1 = x_1, \dots, X_n = x_n) = \mathcal{P}(X_1 = x_1) \dots \mathcal{P}(X_n = x_n)$  de donde deducimos que las  $X_i$  con  $i = 1, \dots, n$  son variables independientes y esta afirmación es cierta para cualquier  $n$  finito que tomemos. Estamos diciendo que la sucesión  $\{X_n\}_{n \geq 1}$  está formada por variables independientes.

Para ver la utilidad de esta sucesión supongamos que mediante algún procedimiento generamos un punto aleatorio uniforme en  $(0, 1]$ ,  $\omega$ . Dado ese  $\omega$  tenemos determinados todos los valores  $X_n(\omega)$ . También tenemos una moneda bien construida, esto es, que tiene la misma probabilidad de cara y cruz. Lanzamientos sucesivos e independientes de la moneda pueden de un modo natural identificarse con la sucesión de los  $X_n(\omega)$  si interpretamos  $X_n(\omega) = 1$  como que el  $n$ -ésimo lanzamiento ha producido el resultado de cara y  $X_n(\omega) = 0$  como cruz en dicho lanzamiento. Para que esta identificación tenga sentido es preciso que las  $X_n$  sean independientes (lanzamientos independientes) y  $\mathcal{P}(X_n(\omega) = 0) = \mathcal{P}(X_n(\omega) = 1) = \frac{1}{2}$  (moneda bien construida). Utilizando esta identificación podremos responder preguntas sobre el juego como, ¿si lanzamos un gran número de veces la moneda, tendremos aproximadamente el mismo número de caras que de cruces?



## 4.4. Algunas aplicaciones

### 4.4.1. Estimación del volumen de objetos tridimensionales

Un animal ha muerto como consecuencia de un cancer de hígado. Se pretende evaluar el grado de desarrollo que había alcanzado la enfermedad estimando el volumen total ocupado por la zona afectada dentro del hígado,  $C$ . Uno de los métodos más utilizados es el siguiente: elegimos una dirección cualquiera en  $\mathcal{R}^3$ , por ejemplo, la del eje de abscisas. Sea  $S$  el plano ortogonal a la dirección dada y que pasa por el origen, en nuestro caso el plano determinado por los dos ejes restantes. Nuestro problema es estimar el volumen de  $C$ ,  $\lambda_3(C)$ . Para ello tomamos una serie de planos paralelos al plano  $S$  equidistantes entre sí una distancia  $h_0$ . Si  $S_h = S + (0, 0, h)$  entonces  $S_h \cap C$  denota la sección de  $C$  mediante el plano  $S_h$ . Vamos a seccionar el hígado mediante los planos  $S_{h_n}$  con  $h_n = nh_0$  y  $n \in \mathbb{Z}_+$ . Sobre la sección  $S_h$  medimos el área que ocupa la zona afectada,  $\lambda_2(C \cap S_h)$ . Con los valores  $\{\lambda_2(C \cap S_{h_n})\}$  y utilizando algún algoritmo de integración numérica podemos estimar

$$\int_{\mathcal{R}} \lambda_2(C \cap S_h) dh,$$

pero

$$\int_{\mathcal{R}} \lambda_2(C \cap S_h) dh = \int_{\mathcal{R}} \int_{\mathcal{R}} \int_{\mathcal{R}} 1_C(x, y, h) dx dy dh = \lambda_3(C).$$

### 4.4.2. Covarianza de variables independientes

Sean  $X$  e  $Y$  dos variables independientes con distribuciones marginales respectivas  $\mathcal{P}_1$  y  $\mathcal{P}_2$  y conjunta  $\mathcal{P}_{12}$  tales que existen los valores esperados de  $X$ ,  $Y$  y  $XY$ . Aplicando el teorema de Fubini se tiene,

$$\begin{aligned} E(XY) &= \int_{\mathcal{R}^2} xy d\mathcal{P}_{12}(x, y) \\ &= \int_{\mathcal{R}} \int_{\mathcal{R}} xy d\mathcal{P}_1(x) d\mathcal{P}_2(y) \\ &= EX \int_{\mathcal{R}} y d\mathcal{P}_2(y) = EXEY, \end{aligned}$$

por lo que  $cov(X, Y) = E(XY) - E(X)E(Y) = 0$ . En particular el coeficiente de correlación de Pearson

$$\rho^2 = \frac{cov(X, Y)}{var(X)^{1/2} var(Y)^{1/2}}$$

también se anula para  $X$  e  $Y$  independientes. El recíproco no es cierto en general.

Dos variables aleatorias cuyo coeficiente de correlación es 1 o -1 verifican que una es función lineal de la otra.

#### 4.4.3. Una igualdad muy práctica

Sean  $X$  e  $Y$  dos vectores aleatorios independientes de dimensiones respectivas  $n$  y  $m$  y cuyas distribuciones denotamos por  $\mathcal{P}_1$  y  $\mathcal{P}_2$ .  $(X, Y)$  es pues un vector a valores en  $\mathcal{R}^{n+m}$ .

**Proposición 4.2** Si  $B \in \mathcal{B}^{n+m}$  entonces

$$\begin{aligned} \mathcal{P}((X, Y) \in B) &= \int_{\mathcal{R}^m} P(\{\omega : (X(\omega), y) \in B\}) \mathcal{P}_2(dy) \\ &= \int_{\mathcal{R}^n} P(\{\omega : (x, Y(\omega)) \in B\}) \mathcal{P}_1(dx). \end{aligned}$$

La prueba es inmediata si consideramos que la distribución del vector  $(X, Y)$  es  $\mathcal{P}_1 \times \mathcal{P}_2$ . Veamos una aplicación importante de esta proposición.

#### 4.4.4. Problema de la aguja de Buffon

Lo planteó el naturalista Buffon en 1777 dentro de su libro *Essai d'arithmétique morale* es el siguiente. Consideremos en  $\mathcal{R}^2$  una serie de líneas paralelas ortogonales al eje  $x$  y separadas entre si por una distancia fija  $d$ . Tomemos una aguja de coser no muy larga ( $l < d$ ) y la lanzamos *aleatoriamente* sobre el plano. ¿Cuál es la probabilidad de que la aguja toque alguna de las líneas?

Cuando Buffon usa la palabra *aleatoriamente* está queriendo decir que puede caer de cualquier forma y en cualquier sitio. Hemos de definir con precisión qué entendemos por caer aleatoriamente. La situación de la aguja queda caracterizada si conocemos la posición de su punto medio  $(x, y)$  y su orientación o ángulo que forma con el eje de abscisas  $\theta$ . Asumiremos en primer lugar que ninguna banda tiene una especial predisposición a que la aguja caiga en ella. Podemos suponer por tanto que  $(x, y)$  está en la delimitada por el  $x = 0$  y  $x = d$ . Observemos que el valor de  $y$  no influye a la hora de considerar si la aguja toca alguna de estas líneas. Sea  $X$  la variable aleatoria que nos da la abscisa del punto medio.

También vamos a asumir que son independientes la posición en la que cae con la orientación. Si  $\Theta$  denota la orientación aleatoria de la aguja estamos asumiendo en realidad la independencia de  $\Theta$  y  $X$ . Por último, dentro de los valores que puede tomar  $X$  y  $\Theta$  no debe de haber subconjuntos que con la misma longitud tengan más probabilidad uno que otro. En definitiva,  $X$  la suponemos uniforme en  $[0, d]$  y  $\Theta$  uniforme en  $(-\frac{\pi}{2}, +\frac{\pi}{2}]$ .

¿Qué valores  $(x, \theta)$  son tales que la aguja toca a una de las dos líneas? Son los valores que constituyen el siguiente conjunto,

$$B = \{(x, \theta) : 0 \leq x \leq \frac{l}{2} \cos \theta \text{ o } d - \frac{l}{2} \cos \theta \leq x \leq d\}.$$

Aplicando la anterior proposición la probabilidad que buscamos es

$$\begin{aligned}\mathcal{P}((X, \Theta) \in B) &= \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \frac{1}{\pi} \mathcal{P}(X \in [0, \frac{l}{2} \cos \theta] \cup [d - \frac{l}{2} \cos \theta, d]) d\theta \\ &= \frac{1}{\pi d} \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} l \cos \theta d\theta = \frac{2l}{\pi d}.\end{aligned}$$



## Capítulo 5

# Sucesiones de Variables Aleatorias. Leyes de los Grandes Números

Los capítulos anteriores no han permitido familiarizarnos con el concepto de variable y vector aleatorio, dotándonos de las herramientas que nos permiten conocer su comportamiento probabilístico. En el caso de un vector aleatorio somos capaces de estudiar el comportamiento conjunto de un número finito de variables aleatorias. Pero imaginemos por un momento los modelos probabilísticos asociados a los siguientes fenómenos:

1. lanzamientos sucesivos de una moneda,
2. tiempos transcurridos entre dos llamadas consecutivas a una misma centralita y
3. sucesión de estimadores de un parámetro cuando se incrementa el tamaño de la muestra.

En todos los casos las variables aleatorias involucradas lo son en cantidad numerable y habremos de ser capaces de estudiar su comportamiento conjunto y, tal y como siempre sucede en ocasiones similares, de conocer cuanto haga referencia a su comportamiento límite. El objetivo de este y posteriores capítulos es dar respuesta estas necesidades.

### 5.1. Convergencia en probabilidad

Sobre un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathcal{P})$  consideremos la sucesión de variables aleatorias  $\{X_n\}$  y la variable aleatoria  $X$ ,

**Definición 5.1** Decimos que  $\{X_n\}$  converge a  $X$  en probabilidad,  $X_n \xrightarrow{\mathcal{P}} X$ , si para cada  $\delta > 0$ ,

$$\lim_n \mathcal{P}\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\} = 0.$$

Al igual que ocurre con otros tipos de convergencia también ahora se verifica la igualdad de los límites, si bien casi seguramente.

**Teorema 5.1** Si  $X_n \xrightarrow{\mathcal{P}} X$  y  $X_n \xrightarrow{\mathcal{P}} Y$  entonces  $X = Y$  a.s..

**Demostración.-** Sean  $A_k = \{\omega : |Y(\omega) - X(\omega)| \geq \frac{1}{k}\}$ ,  $A_{kn} = \{\omega : |Y(\omega) - X_n(\omega)| \geq \frac{1}{2k}\}$  y  $B_{kn} = \{\omega : |X(\omega) - X_n(\omega)| \geq \frac{1}{2k}\}$ . Por la desigualdad triangular,  $|Y(\omega) - X(\omega)| \leq |Y(\omega) - X_n(\omega)| + |X(\omega) - X_n(\omega)|$ , se verifica  $A_k \subset A_{kn} \cup B_{kn}$ ,  $\forall n$ , y en consecuencia  $\mathcal{P}(A_k) \leq \mathcal{P}(A_{kn}) + \mathcal{P}(B_{kn})$ ,  $\forall n$ . Por otra parte

$$\{\omega : Y(\omega) \neq X(\omega)\} = \bigcup_{k \geq 1} \{\omega : |Y(\omega) - X(\omega)| \geq \frac{1}{k}\},$$

y teniendo en cuenta la convergencia en probabilidad de  $X_n$  a  $X$  y  $Y$ ,  $\mathcal{P}(\{\omega : Y(\omega) \neq X(\omega)\}) = 0$ .  $\square$

Los dos teoremas que siguen relacionan la convergencia *en probabilidad* y la convergencia *casi segura*.

**Teorema 5.2**  $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{\mathcal{P}} X$ .

**Demostración.-** Para  $\delta > 0$  sean  $A = \{\omega : \lim X_n(\omega) \neq X(\omega)\}$  y  $A_n = \{\omega : |X_n(\omega) - X(\omega)| \geq \delta\}$ , entonces  $\limsup A_n \subset A$  y  $\mathcal{P}(\limsup A_n) = 0$ , y de aquí  $\lim \mathcal{P}(A_n) = 0$ .  $\square$

**Teorema 5.3**  $X_n \xrightarrow{\mathcal{P}} X$  si y solo si *para cada subsucesión  $\{X_{n_k}\}$ , existe una subsubsucesión  $\{X_{n_{k(m)}}\}$  verificando  $X_{n_{k(m)}} \xrightarrow{a.s.} X$*

**Demostración.-** Si  $X_n \xrightarrow{\mathcal{P}} X$ , consideremos una subsucesión cualquiera  $\{X_{n_k}\}$ . Dado  $m$ , existe  $k(m)$  tal que para  $A_{n_k} = \{\omega : |X_{n_k}(\omega) - X(\omega)| \geq \frac{1}{m}\}$ ,  $\mathcal{P}(A_{n_k}) \leq 2^{-m}$ ,  $\forall k \geq k(m)$ . Se comprueba fácilmente que,

$$\mathcal{P}(\limsup A_{n_{k(m)}}) = \mathcal{P}\{\omega : \bigcap_m \bigcup_{i \geq m} |X_{n_{k(i)}}(\omega) - X(\omega)| \geq \frac{1}{i}\} = 0,$$

con lo que  $X_{n_{k(m)}} \xrightarrow{a.s.} X$  con  $m$ .

Para demostrar la suficiencia supongamos que  $X_n$  no converge en probabilidad a  $X$ . Ello supone que existirán un  $\delta > 0$  y un  $\epsilon > 0$  tal que  $\mathcal{P}\{\omega : |X_{n_k}(\omega) - X(\omega)| \geq \delta\} > \epsilon$  para alguna subsucesión  $\{X_{n_k}\}$ , y por lo tanto ninguna de sus subsubsucesiones podrá converger *a.s.* a  $X$  de acuerdo con el teorema anterior.  $\square$

**Corolario 5.1** Si  $g$  es continua y  $X_n \xrightarrow{\mathcal{P}} X$  entonces  $g(X_n) \xrightarrow{\mathcal{P}} g(X)$ .

Conviene señalar la distinta naturaleza de las convergencias en *probabilidad* y *casi segura*. Mientras esta última es una convergencia de tipo puntual, aquella es de tipo conjuntista. El ejemplo que sigue ilustra bien esta diferencia y pone de manifiesto que la implicación contraria al teorema 5.2 no se verifica.

**Ejemplo 5.1 La convergencia en probabilidad no implica la convergencia casi segura.** Para nuestro espacio de probabilidad consideraremos el intervalo unidad dotado de la  $\sigma$  álgebra de Borel y de la medida de Lebesgue. Definimos la sucesión  $X_n = \mathbf{1}_{I_n}$ ,  $\forall n$  con  $I_n = ]\frac{p}{2^q}, \frac{p+1}{2^q}]$ , siendo  $p$  y  $q$  los únicos enteros positivos que verifican,  $p + 2^q = n$  y  $0 \leq p < 2^q$ . Por ejemplo.

$$\begin{array}{lll} n = 1 & q = 0, p = 0 & X_1 = \mathbf{1}_{]0,1]} \\ n = 2 & q = 1, p = 0 & X_2 = \mathbf{1}_{]0, \frac{1}{2}]} \\ n = 3 & q = 1, p = 1 & X_3 = \mathbf{1}_{] \frac{1}{2}, 1]} \\ n = 4 & q = 2, p = 0 & X_4 = \mathbf{1}_{]0, \frac{1}{4}]} \\ n = 5 & q = 2, p = 1 & X_5 = \mathbf{1}_{] \frac{1}{4}, \frac{1}{2}]} \\ n = 6 & q = 2, p = 2 & X_6 = \mathbf{1}_{] \frac{1}{2}, \frac{3}{4}]} \\ n = 7 & q = 2, p = 3 & X_7 = \mathbf{1}_{] \frac{3}{4}, 1]} \\ \dots\dots & \dots\dots & \dots\dots \end{array}$$

donde claramente se ve que  $q = q(n)$  y  $\lim_n q(n) = +\infty$ .

Por su construcción las  $X_n$  no convergen puntualmente en  $[0, 1]$ , pero si  $X = 0$ ,  $\forall \delta > 0$  se tiene  $\lambda\{\omega : |X_n(\omega)| > \delta\} = \lambda\{\omega : |X_n(\omega)| = 1\} = \lambda(I_n) = 2^{-q}$ ,  $2^q \leq n < 2^{q+1}$  y  $X_n \xrightarrow{\lambda} 0$ .

## 5.2. Lemas de Borel-Cantelli

Con frecuencia aparecen en Probabilidad unos sucesos que tienen la particularidad de ser independientes de si mismos. Si el conjunto medible  $A$  es independiente de si mismo entonces,  $\mathcal{P}(A) = \mathcal{P}(A \cap A) = \mathcal{P}(A)\mathcal{P}(A) = \mathcal{P}(A)^2$  por lo que  $\mathcal{P}(A)$  ha de ser cero o uno. Los lemas de Borel-Cantelli tratan sobre un suceso importante de este tipo.

**Teorema 5.4 (Primer lema de Borel-Cantelli)** Sea  $\{A_n\}_{n \geq 1}$  una sucesión de conjuntos medibles tal que  $\sum_n \mathcal{P}(A_n)$  es una serie convergente, entonces

$$\mathcal{P}(\limsup_n A_n) = 0$$

**Demostración.-** De  $\limsup_n A_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n$  se deduce  $\limsup_n A_n \subset \bigcup_{n \geq k} A_n$  para cualquier  $k$ , luego

$$\mathcal{P}(\limsup_n A_n) \leq \mathcal{P}(\bigcup_{n \geq k} A_n) \leq \sum_{n \geq k} \mathcal{P}(A_n),$$

pero  $\sum_{n \geq k} \mathcal{P}(A_n)$  tiende a cero con  $k$  por ser convergente la serie correspondiente.  $\square$

**Teorema 5.5 (Segundo lema de Borel-Cantelli)** *Si  $\{A_n\}_{n \geq 1}$  es una sucesión de sucesos independientes tal que  $\sum_n \mathcal{P}(A_n)$  diverge, entonces*

$$\mathcal{P}(\limsup_n A_n) = 1$$

.

**Demostración.-** Veamos que  $\mathcal{P}((\limsup_n A_n)^c) = 0$ . De  $(\limsup_n A_n)^c = \bigcup_{k \geq 1} \bigcap_{n \geq k} A_n^c$  y por la continuidad desde abajo de la medida de probabilidad,

$$\mathcal{P}((\limsup_n A_n)^c) = \lim_k \mathcal{P}(\bigcap_{n \geq k} A_n^c).$$

De la independencia de los  $\{A_n\}_{n \geq 1}$  se deduce la de los  $\{A_n^c\}_{n \geq 1}$ . Además  $1 - x \leq e^{-x}$  si  $0 \leq x \leq 1$ . Tenemos pues

$$\mathcal{P}(\bigcap_{n=k}^{k+m} A_n^c) = \prod_{n=k}^{k+m} \mathcal{P}(A_n^c) =$$

$$\prod_{n=k}^{k+m} (1 - \mathcal{P}(A_n)) \leq \prod_{n=k}^{k+m} e^{-\mathcal{P}(A_n)} = e^{-\sum_{n=k}^{k+m} \mathcal{P}(A_n)}.$$

De la divergencia de  $\sum_n \mathcal{P}(A_n)$  se deduce que el último término tiende a cero con  $m$  para cada  $k$  de modo que  $\mathcal{P}(\bigcap_{n \geq k} A_n^c) = 0$  y se sigue el resultado.  $\square$

De los dos lemas de Borel-Cantelli se desprende que si los  $\{A_n\}$  son independientes entonces  $\limsup A_n$  es independiente de si mismo. De otro modo,  $\limsup A_n$  tiene probabilidad *cero* o *uno*. Además tenemos un criterio para saber cuando es *cero* o cuando es *uno*, según que  $\sum_n \mathcal{P}(A_n)$  sea convergente o divergente. En el apartado siguiente veremos otro resultado de este tipo, la ley cero-uno de Kolmogorov, para una clase mayor de sucesos.



### 5.3. Ley Cero-Uno de Kolmogorov

Recordemos que para una variable aleatoria  $X$ ,  $\sigma(X) = \{X^{-1}(B), B \in \mathcal{B}\}$ , con  $\mathcal{B}$  la  $\sigma$  álgebra de Borel, representa la  $\sigma$  álgebra engendrada por  $X$ . Si se trata de una familia finita de variables aleatorias  $X_1, X_2, \dots, X_n$ , la familia  $\{X_i^{-1}(B), B \in \mathcal{B}, i = 1, \dots, n\}$  no es una  $\sigma$  álgebra pero podemos engendrar una a partir de ella mediante,

$$\sigma(X_1, \dots, X_n) = \sigma\{X_i^{-1}(B), B \in \mathcal{B}, i = 1, \dots, n\},$$

a la que llamamos  $\sigma$  álgebra engendrada por  $X_1, \dots, X_n$  y que, como fácilmente se comprueba, admite esta otra expresion,

$$\sigma(X_1, \dots, X_n) = \sigma\left(\bigcup_{i=1}^n \sigma(X_i)\right).$$

Si nuestra familia es ahora numerable,  $\{X_n\}_{n \geq 1}$ , un razonamiento análogo nos permite definir la  $\sigma$  álgebra engendrada por  $X_n, X_{n+1}, \dots$ , como

$$\sigma(X_n, X_{n+1}, \dots) = \sigma\left(\bigcup_{i \geq n} \sigma(X_i)\right).$$

Obtendremos así la familia decreciente de  $\sigma$  álgebras,  $\sigma(X_n, X_{n+1}, \dots)$ ,  $n \geq 1$ , cuya intersección será otra  $\sigma$  álgebra

$$\tau = \bigcap_{n \geq 1} \sigma(X_n, X_{n+1}, \dots),$$

que denominaremos  $\sigma$  álgebra cola de la sucesión  $\{X_n\}$ . Cuando las  $X_n$  son independientes esta  $\sigma$  álgebra tiene interesantes propiedades.

**Teorema 5.6 (Ley Cero-Uno de Kolmogorov)** *Si las variables  $\{X_n\}_{n \geq 1}$  son independientes,  $\forall E \in \tau$ ,  $\mathcal{P}(E) = 0$  o  $\mathcal{P}(E) = 1$ .*

**Demostración.-** La familia  $\mathcal{A}_0 = \bigcup_{n \geq 1} \sigma(X_1, \dots, X_n)$  verifica que  $\sigma(X_n) \subset \mathcal{A}_0$ ,  $\forall n$ , entonces  $\sigma(\mathcal{A}_0) = \sigma(X_1, \dots)$ . Pero  $\tau \subset \sigma(X_n, X_{n+1}, \dots)$  y será independiente de  $\sigma(X_1, \dots, X_n)$ ,  $\forall n$ , lo que supone que lo será de  $\sigma(\mathcal{A}_0)$  y como  $\tau \subset \sigma(X_1, \dots)$ , sus elementos serán independientes de si mismos y se ha de verificar  $\forall E \in \tau$  que  $\mathcal{P}(E \cap E) = (\mathcal{P}(E))^2 = \mathcal{P}(E)$ , relación que exige necesariamente  $\mathcal{P}(E) = 0$  o  $\mathcal{P}(E) = 1$ .  $\square$

Se derivan de este teorema importantes resultados relacionados con la convergencia de sucesiones de variables aleatorias.

**Corolario 5.2** *Si  $Y$  es una variable aleatoria tal que  $\sigma_Y \subset \tau$ , entonces  $Y$  es constante con probabilidad uno.*

**Demostración.-** Según la hipótesis  $\{Y \leq y\} \subset \tau$ ,  $\forall y \in \mathcal{R}$  y por lo tanto  $\mathcal{P}(Y \leq y) = 1$  o  $\mathcal{P}(Y \leq y) = 0$ , por lo que necesariamente debe existir  $y_0$  tal que  $\mathcal{P}(Y = y_0) = 1$ .  $\square$

**Corolario 5.3** Si  $\{X_n\}$  es una sucesión de variables aleatorias independientes,

1.  $\{X_n\}$  converge a.s. a una constante o diverge a.s.,
2.  $\sum_{n \geq 1} X_n$  converge a.s. a un límite finito o diverge a.s. y
3. para toda sucesión de constantes,  $\{a_n\}$ , con  $\lim a_n = 0$ ,  $\{a_n \sum_{i=1}^n X_i\}$  converge a.s. a una constante o diverge a.s.

**Demostración.-** Observemos previamente que para dos variables aleatorias cualesquiera,  $X_i$  y  $X_j$ , y para una constante cualquiera,  $k$ , podemos escribir

$$A = \{\omega : X_i(\omega) + X_j(\omega) < k\} = \bigcup_{r \in \mathcal{Q}} \{\omega : X_i(\omega) < r\} \cap \{\omega : X_j(\omega) < k - r\},$$

lo que supone que  $A \in \sigma(X_i, X_j)$ . De forma análoga, para  $\epsilon > 0$  y para  $X_{i_1}, \dots, X_{i_m}$ , podemos afirmar que  $\{\omega : |\sum_{j=1}^m (\pm X_{i_j}(\omega))| < \epsilon\} \in \sigma(X_{i_1}, \dots, X_{i_m})$ . Pasemos ahora a la demostración.

1. Sea  $E = \{\omega : \exists \lim X_n(\omega) \text{ y es finito}\}$ , para cada  $n$  se verifica que  $E = \{\omega : \exists \lim_{k \geq 0} X_{n+k}(\omega) \text{ y es finito}\}$  y por el criterio de Cauchy podremos expresar  $E$  de la siguiente forma,

$$E = \bigcap_{m \geq 1} \bigcup_{k \geq n} \bigcap_{i \geq k} \bigcap_{j \geq i} \{\omega, |X_i(\omega) - X_j(\omega)| < \frac{1}{m}\},$$

lo que de acuerdo con el comentario inicial supone que  $E \in \sigma(X_n, X_{n+1}, \dots)$ ,  $\forall n$  y por tanto  $E \in \tau$ . La ley cero-uno de Kolmogorov asegura que  $\mathcal{P}(E) = 0$  o  $\mathcal{P}(E) = 1$ . En este último caso, si  $X_n \xrightarrow{a.s.} X$ ,

$$\{\omega : X(\omega) < c\} = \bigcup_{m \geq 1} \bigcap_{n \geq m} \{\omega : X_n(\omega) < c\}, \quad \forall c \in \mathcal{R},$$

y  $\sigma_X \subset \tau$ . El corolario 5.2 nos dice que  $X$  es a.s. constante.

2. Si tenemos en cuenta que  $\sum_{n \geq 1} X_n$  converge si y solo si lo hace  $\sum_{i \geq n} X_i$ , razonando como antes, el correspondiente conjunto  $E$  en el que existe límite finito podrá escribirse,

$$E = \bigcap_{m \geq 1} \bigcup_{k \geq n} \bigcap_{i \geq k} \bigcap_{j \geq i} \{\omega, |\sum_{l=i}^j X_l(\omega)| < \frac{1}{m}\},$$

deduciendo, también como antes, que  $\mathcal{P}(E) = 0$  o  $\mathcal{P}(E) = 1$ .

3.  $a_n \sum_{i=1}^n X_i = a_n \sum_{i=1}^N X_i + a_n \sum_{i=N+1}^n X_i$ ,  $n > N$ , y como las  $a_n \rightarrow 0$ ,  $\{a_n \sum_{i=1}^n X_i\}$  converge si y solo si  $\{a_n \sum_{i=N+1}^n X_i\}$  lo hace, con lo que podremos repetir el argumento de los apartados anteriores, en particular la última parte del razonamiento del 1) para establecer que el límite, si es finito, es también ahora a.s. constante.  $\square$

**Observación 5.1** Señalemos que para el segundo de los apartados del corolario, a diferencia de lo afirmado para los otros dos, el límite, aún siendo finito, no tiene porque ser constante. En efecto, como más tarde veremos, para variables independientes, centradas y con varianzas tales que  $\sum_{n \geq 1} \text{var}(X_n) < +\infty$ , la serie  $\sum X_n$  converge a.s. a una variable aleatoria  $X$  con media 0 y  $\text{var}(X) = \sum \text{var}(X_n)$ .

## 5.4. Leyes de los Grandes Números

El nombre de *leyes de los grandes números* hace referencia al estudio de un tipo especial de límites derivados de la sucesión de variables aleatorias  $\{X_n\}$ . Concretamente los de la forma  $\lim_n \frac{S_n - a_n}{b_n}$ , con  $S_n = \sum_{i=1}^n X_i$  y siendo  $\{a_n\}$  y  $\{b_n\}$  sucesiones de constantes tales que  $\lim b_n = +\infty$ . La ley cero-uno de Kolmogorov a través del corolario 5.3 nos da una respuesta parcial, recordemos que si las  $X_n$  son independientes el límite anterior converge a.s. a una constante o diverge a.s. En esta sección fijaremos las condiciones para saber cuando existe convergencia a.s., y como nos ocuparemos también de la convergencia en probabilidad, las leyes se denominarán *fuerte* y *débil* respectivamente.

**Teorema 5.7 (Ley débil)** Sea  $\{X_k\}$  una sucesión de variables aleatorias independientes tales que  $E(X_k^2) < +\infty$  para cualquier  $k$  y  $\lim_n \frac{1}{n^2} \sum_{k=1}^n \text{var}(X_k) = 0$ , entonces

$$\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow{P} 0.$$

**Demostración.-** Para  $S_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k))$ , se verifica que  $E(S_n) = 0$  y  $\text{var}(S_n) = \frac{1}{n^2} \sum_{k=1}^n \text{var}(X_k)$ . Por la desigualdad de Chebyshev,  $\forall \epsilon > 0$

$$\mathcal{P}(|S_n| \geq \epsilon) \leq \frac{\text{var}(S_n)}{\epsilon^2} = \frac{1}{n^2 \epsilon^2} \sum_{k=1}^n \text{var}(X_k),$$

que al pasar al límite nos asegura la convergencia en probabilidad de  $S_n$  a 0.  $\square$

**Corolario 5.4** Si las  $X_n$  son i.i.d. con varianza finita y esperanza común  $EX_1$ , entonces  $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} EX_1$ .

**Demostración.-** Si  $\text{var}(X_k) = \sigma^2$ ,  $\forall k$ , tendremos  $\frac{1}{n^2} \sum_{k=1}^n \text{var}(X_k) = \frac{\sigma^2}{n}$  que tiende a cero con  $n$ . Es por tanto de aplicación la ley débil que conduce al resultado enunciado.  $\square$

Este resultado fué demostrado por primera vez por J. Bernoulli para variables con distribución binomial, lo que se conoce como la *ley de los grandes números de Bernoulli*. En efecto, si  $Y_n \sim B(n, p)$  podemos considerar  $Y_n = \sum_{k=1}^n X_k$ , con  $X_k \sim B(1, p)$  e independientes, por lo que la aplicación del corolario conduce a  $\frac{Y_n}{n} \xrightarrow{P} p$ .

El siguiente paso será fijar las condiciones para que el resultado sea válido bajo convergencia a.s.

**Teorema 5.8 (Ley fuerte)** Si  $\{X_k\}$  es una sucesión de variables aleatorias i.i.d. con media finita, entonces

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{a.s.} E(X_1).$$

**Demostración.-** Sin pérdida de generalidad podemos suponer  $X_k \geq 0$ ,  $\forall k$ . Para cada  $k$  definamos la variable truncada  $Y_k = X_k \mathbf{1}_{\{X_k \leq k\}}$  y consideremos la suma parcial  $S_n^* = \sum_{k=1}^n Y_k$ . Por la independencia de la  $X_k$ ,

$$\begin{aligned} \text{var}(S_n^*) &= \sum_{k=1}^n \text{var}(Y_k) \leq \sum_{k=1}^n E(Y_k^2) = \\ &= \sum_{k=1}^n E(X_k^2 \mathbf{1}_{\{X_k \leq k\}}) \leq n E(X_1^2 \mathbf{1}_{\{X_1 \leq n\}}). \end{aligned}$$

Sea ahora  $\alpha > 1$ , fijo y  $u_n = [\alpha^n]$ , por la desigualdad de Chebyshev

$$\begin{aligned} \mathcal{P}\left(\left|\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n}\right| > \epsilon\right) &\leq \frac{\text{var}(S_{u_n}^*)}{\epsilon^2 u_n} \leq \\ &\leq \frac{u_n E(X_1^2 \mathbf{1}_{\{X_1 \leq n\}})}{\epsilon^2 u_n} = \frac{1}{\epsilon^2} E\left(\frac{X_1^2}{u_n} \mathbf{1}_{\{X_1 \leq n\}}\right), \end{aligned}$$

y sumando para  $n$

$$\begin{aligned} \sum_{n \geq 1} \mathcal{P}\left(\left|\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n}\right| > \epsilon\right) &\leq \\ &\leq \frac{1}{\epsilon^2} \sum_{n \geq 1} E\left(\frac{X_1^2}{u_n} \mathbf{1}_{\{X_1 \leq n\}}\right) = \frac{1}{\epsilon^2} E\left(X_1^2 \sum_{n \geq 1} \frac{1}{u_n} \mathbf{1}_{\{X_1 \leq n\}}\right). \end{aligned} \quad (5.1)$$

Si  $n_0$  es el menor  $n$  tal que  $u_n \geq x$ ,  $\alpha^{n_0} \geq x$  y como para  $y \geq 1$  se verifica  $y \leq 2[y]$ , tendremos que

$$\sum_{u_n \geq x} u_n^{-1} = \sum_{[\alpha^n] \geq x} [\alpha^n]^{-1} = \sum_{n \geq n_0} [\alpha^n]^{-1} \leq$$

$$\leq 2 \sum_{n \geq n_0} \alpha^{-n} = 2 \frac{\alpha}{\alpha - 1} \alpha^{-n_0} \leq M x^{-1}, \quad M = 2 \frac{\alpha}{\alpha - 1},$$

y de aquí,  $\sum_{n \geq 1} u_n^{-1} 1_{X_1 \leq u_n} \leq M x^{-1}$ , que aplicado a 5.1 nos da

$$\sum_{n \geq 1} \mathcal{P}(|\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n}| > \epsilon) \leq \frac{M}{\epsilon^2} E(X_1) < +\infty. \quad (5.2)$$

Aplicando el primer lema de Borel-Cantelli, para  $\epsilon$  racional, concluimos que  $\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n} \xrightarrow{a.s.} 0$ , y por el teorema de las medias de Cesàro  $u_n^{-1} E(S_{u_n}^*) = u_n^{-1} \sum_{k=1}^{u_n} E(Y_k)$  tiene el mismo límite que  $E(Y_k)$ , es decir,  $E(X_1)$ . En consecuencia  $\frac{S_{u_n}^*}{u_n} \xrightarrow{a.s.} E(X_1)$ .

Por otra parte,

$$\sum_{n \geq 1} \mathcal{P}(X_n \neq Y_n) = \sum_{n \geq 1} \mathcal{P}(X_1 > n) \leq \int_0^{+\infty} \mathcal{P}(X_1 > t) dt = E(X_1) < +\infty$$

y aplicando nuevamente Borel-Cantelli,  $\frac{S_n - S_n^*}{n} \xrightarrow{a.s.} 0$  y por tanto  $\frac{S_{u_n}}{u_n} \xrightarrow{a.s.} E(X_1)$ .

Si  $u_n \leq k \leq u_{n+1}$ ,  $\frac{u_n}{u_{n+1}} \frac{S_{u_n}}{u_n} \leq \frac{S_k}{k} \leq \frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}}{u_{n+1}}$  y pasando al límite, como  $\frac{u_n}{u_{n+1}} \rightarrow \alpha > 1$ , obtendremos

$$\frac{1}{\alpha} E(X_1) \leq \liminf_k \frac{S_k}{k} \leq \limsup_k \frac{S_k}{k} \leq \alpha E(X_1) \quad a.s.$$

y utilizando todos los  $\alpha > 1$  racionales, la intersección sobre los correspondientes conjuntos nos conduce a

$$\frac{S_k}{k} \xrightarrow{a.s.} E(X_1). \square$$

**Corolario 5.5** Si  $\{X_k\}$  es una sucesión de variables aleatorias i.i.d. con  $E(X_1^-) < +\infty$  y  $E(X_1^+) = +\infty$ , entonces  $\frac{S_n}{n} \xrightarrow{a.s.} \infty$ .

**Demostración.-** Por el teorema sabemos que  $\frac{1}{n} \sum_{k=1}^n X_k^- \xrightarrow{a.s.} E(X_1^-)$ , por lo que bastará comprobar el corolario para  $X_k = X_k^+ \geq 0$ . Definamos

$$X_n^u = \begin{cases} X_k, & \text{si } 0 \leq X_k \leq u, \\ 0, & \text{si } X_k > u, \end{cases}$$

entonces

$$\frac{1}{n} \sum_{k=1}^n X_k \geq \frac{1}{n} \sum_{k=1}^n X_k^u \xrightarrow{a.s.} E(X_1^u),$$

y al tender  $u$  a  $+\infty$  obtenemos el resultado buscado.  $\square$

La demostración de la *Ley fuerte* que hemos presentado se debe a Etemadi y supone una gran simplificación respecto de los argumentos clásicos ofrecidos por Kolmogorov que presentaba la *Ley fuerte* como el resultado final de una cadena de propiedades previas, que si bien alargan su obtención, tienen a su favor, en casi todos los casos, el ser resultados interesantes y útiles por si mismos. Es en beneficio de la concisión por lo que hemos elegido la versión de Etemadi, pero aconsejamos vivamente al estudiante que encuentre ocasión de hojear el desarrollo de Kolmogorov en cualquiera de los textos habituales (Burrill, Billingsley,...)

## 5.5. Aplicaciones

### 5.5.1. El teorema de Glivenko-Cantelli

Para las variables aleatorias  $X_1, X_2, \dots, X_n$  se define la función de distribución empírica mediante

$$F_n(x, \omega) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{]-\infty, x]}(X_k(\omega)).$$

Cuando todas las variables tienen la misma distribución,  $F_n(x, \omega)$  es el estimador natural de la función de distribución común,  $F(x)$ . El acierto en la elección de este estimador se pone de manifiesto en el siguiente resultado.

**Teorema 5.9** *Sea  $\{X_k\}$  una sucesión de variables aleatorias i.i.d. con función de distribución común  $F(x)$ , entonces  $F_n(x, \omega) \xrightarrow{a.s.} F(x)$ .*

**Demostración.-** Para cada  $x$ ,  $F_n(x, \omega)$  es una variable aleatoria resultante de sumar las  $n$  variables aleatorias independientes,  $\mathbf{1}_{]-\infty, x]}(X_k(\omega))$ ,  $k = 1, \dots, n$ , cada una de ellas con la misma esperanza,  $E(\mathbf{1}_{]-\infty, x]}(X_k(\omega))) = \mathcal{P}(X_k \leq x) = F(x)$ . Aplicando la ley fuerte de los grandes números,

$$F_n(x, \omega) \xrightarrow{a.s.} F(x). \square$$

Este resultado es previo al teorema que da nombre al apartado y que nos permite contrastar la hipótesis de suponer que  $F$  es la distribución común a toda la sucesión.

**Teorema 5.10 (Glivenko-Cantelli)** *Sea  $\{X_k\}$  una sucesión de variables aleatorias i.i.d. con función de distribución común  $F(x)$ . Hagamos  $D_n(x) = \sup_x |F_n(x, \omega) - F(x)|$ , entonces  $D_n \xrightarrow{a.s.} 0$ .*

La demostración, muy técnica, la omitimos y dejamos al interés del lector consultarla en el texto de Billingsley.

### 5.5.2. Cálculo aproximado de integrales por el método de Monte-Carlo

Sea  $f(x) \in \mathcal{C}([0, 1])$  con valores en  $[0, 1]$ . Una aproximación al valor de  $\int_0^1 f(x)dx$  puede obtenerse a partir de la sucesión de pares de variables aleatorias distribuidas uniformemente en  $[0, 1]$ ,  $(X_1, Y_1), (X_2, Y_2), \dots$ . Para ello hagamos,

$$Z_i = \begin{cases} 1, & \text{si } f(X_i) \geq Y_i \\ 0, & \text{si } f(X_i) < Y_i. \end{cases}$$

Así definidas las  $Z_i$  son variables Bernoulli con parámetro  $p = E(Z_i) = \mathcal{P}(f(X_i) \geq Y_i) = \int_0^1 f(x)dx$ , y aplicándoles la ley fuerte de los grandes números tendremos que

$$\frac{1}{n} \sum_{i=1}^n Z_i \xrightarrow{a.s.} \int_0^1 f(x)dx,$$

lo que en términos prácticos supone simular los pares  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , con  $X_i$  e  $Y_i \sim U(0, 1)$ , y calcular la proporción de ellos que caen por debajo de la gráfica  $y = f(x)$ .

### 5.5.3. Números normales de Borel

En el intervalo  $[0, 1]$  consideremos el desarrollo binario con infinitos 1's de sus elementos. Es decir,  $\omega \in [0, 1]$ ,  $\omega = 0.d_1(\omega)d_2(\omega), \dots, d_i(\omega), \dots$ , con  $d_i(\omega) \in \{0, 1\}$ . La elección de un número al azar en  $[0, 1]$  implica probabilizar el espacio  $([0, 1], \mathcal{B}_{[0, 1]})$  con la medida de Lebesgue, lo que supone, como ya hemos visto, que  $\mathcal{P}(d_i = 1) = \mathcal{P}(d_i = 0) = \frac{1}{2}$ ,  $\forall i$ .

Define Borel un *número normal* como aquel  $\omega$  que verifica

$$\lim_n \frac{1}{n} \sum_{i=1}^n d_i(\omega) = \frac{1}{2},$$

lo que supone que en el límite la frecuencia relativa de 1's y 0's es la misma. Como cada  $d_i \sim B(1, \frac{1}{2})$ , aplicando la ley fuerte de los grandes números podemos afirmar que

$$\frac{1}{n} \sum_{i=1}^n d_i \xrightarrow{a.s.} \frac{1}{2}.$$

Este resultado se conoce como *Teorema de los Números Normales de Borel* y supone que al elegir al azar un número entre 0 y 1, con probabilidad uno, será *normal*.





## Capítulo 6

# Convergencia débil

### 6.1. Convergencia débil

Cualquier generador de números aleatorios uniformes en el intervalo  $(0, 1]$  se basa en la idea de tomar un  $n$  entero no negativo muy grande y elegimos un valor de  $D_n = \{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$ , de modo que cada elemento de  $D_n$  tenga la misma probabilidad de ser seleccionado. El valor es el número aleatorio buscado. Los diferentes generadores se diferencian en el modo en que seleccionamos el elemento de  $D_n$ . En temas anteriores hemos formalizado el concepto de número aleatorio uniforme como el espacio de probabilidad  $((0, 1], \mathcal{B}((0, 1]), \mathcal{P})$  siendo  $\mathcal{P}$  la medida de Lebesgue en dicho intervalo. Sin embargo, ahora estamos identificándolo con  $((0, 1], \mathcal{B}((0, 1]), \mathcal{P}_n)$  donde  $\mathcal{P}_n$  es en realidad una medida de probabilidad discreta uniforme en  $D_n$ . Obviamente  $\mathcal{P}_n$  y  $\mathcal{P}$  son medidas diferentes para cualquier  $n$  finito que tomemos, aunque cabe esperar que cuando  $n$  es muy grande ambas medidas sean similares. ¿Son similares en el sentido de que

$$\lim_n \mathcal{P}_n(A) = \mathcal{P}(A) \quad (6.1)$$

para cualquier boreliano  $A$  que tomemos? No, porque  $\mathcal{P}_n$  es una medida discreta concentrada sobre  $D_n$  que es un subconjunto de los racionales,  $\mathcal{Q}$ , por tanto  $\mathcal{P}_n(\mathcal{Q}) = 1$  para cualquier  $n$ , de donde  $\lim_n \mathcal{P}_n(\mathcal{Q}) = 1$  pero  $\mathcal{P}(\mathcal{Q}) = 0$ .

Aunque 6.1 no es cierta para un conjunto de Borel en  $(0, 1]$  cualquiera, sí se verifica para una familia de conjuntos medibles muy importantes. En efecto, si  $x \in (0, 1]$  entonces  $\mathcal{P}_n(0, x] = \frac{[nx]}{n}$ , donde  $[nx]$  denota la parte entera por defecto de  $nx$ . Tendremos que  $\lim_n \mathcal{P}_n(0, x] = \lim_n \frac{[nx]}{n} = x = \mathcal{P}(0, x]$ , lo que significa que para valores grandes de  $n$  la función de distribución de  $\mathcal{P}_n$  es muy parecida a la función de distribución de  $\mathcal{P}$ . Es a este tipo de similitud al que nos estamos refiriendo.

**Definición 6.1 (Convergencia débil de medidas de probabilidad)** Sean  $\{\mathcal{P}_n\}_{n \geq 1}$  y  $\mathcal{P}$  medidas de probabilidad sobre la recta real, diremos que las  $\mathcal{P}_n$  convergen débilmente a  $\mathcal{P}$  si  $\lim_n \mathcal{P}_n(-\infty, x] = \mathcal{P}(-\infty, x]$  para cualquier  $x \in \mathcal{R}$  tal que  $\mathcal{P}\{x\} = 0$ . Lo denotamos por  $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$ .

Si  $F_n$  y  $F$  denotan las funciones de distribución de  $\mathcal{P}_n$  y  $\mathcal{P}$  entonces diremos que  $\{F_n\}_{n \geq 1}$  converge débilmente a  $F$  si  $\lim_n F_n(x) = F(x)$  para cualquier  $x$  punto de continuidad de  $F$ . Lo denotamos por  $F_n \xrightarrow{w} F$ .

Puesto que  $\mathcal{P}\{x\} = F(x) - \lim_{y \rightarrow x-} F(y)$  ambas definiciones son equivalentes. En el ejemplo anterior estamos aproximando una distribución continua con una distribución discreta. También podemos aproximar una distribución discreta mediante otra distribución discreta.

**Ejemplo 6.1** Una distribución binomial puede aproximarse por una distribución de Poisson cuando tenemos un gran número de pruebas con una probabilidad de éxito muy pequeña. Si  $\mathcal{P}_n$  denota una distribución binomial con parámetros  $n$  y  $p_n$  y  $\lim_n np_n = \lambda$  con  $0 < \lambda < +\infty$ , y mediante  $\mathcal{P}$  denotamos una distribución de Poisson con media  $\lambda$ , haciendo  $\lambda_n = np_n$  tendremos, para  $k = 0, \dots, n$ ,

$$\mathcal{P}_n\{k\} = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1) \dots (n-k+1)}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{-k} \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n$$

y

$$\lim_n \frac{n(n-1) \dots (n-k+1)}{n^k} = 1.$$

De  $\lim_n \lambda_n = \lambda$ , finito, se sigue que  $\lim_n (1 - \frac{\lambda_n}{n})^{-k} = 1$ ;  $\lim_n \lambda_n^k = \lambda^k$  y  $\lim_n (1 - \frac{\lambda_n}{n})^n = e^{-\lambda}$ . Por tanto:

$$\lim_n \mathcal{P}_n\{k\} = \mathcal{P}\{k\} \quad (6.2)$$

Puesto que  $\mathcal{P}_n(-\infty, x] = \sum_{k=0}^{\lfloor x \rfloor} \mathcal{P}_n\{k\}$ , utilizando 6.2 se tiene,

$$\lim_n \mathcal{P}_n(-\infty, x] = \mathcal{P}(-\infty, x], \quad \forall x$$

y por tanto  $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$ .

En la definición propuesta de convergencia débil hemos exigido que tanto los elementos de la sucesión  $\mathcal{P}_n$  como  $\mathcal{P}$  fueran medidas de probabilidad. No es una hipótesis gratuita.

**Ejemplo 6.2** Sea  $\mathcal{P}_n = \delta_{\{n\}}$  la delta de Dirac en  $\{n\}$ . Tenemos  $F_n = 1_{[n, +\infty)}$  que converge a la función idénticamente nula que no es una función de distribución de probabilidad. Si  $\mathcal{P}(A) = 0$  para cualquier boreliano  $A$  entonces  $\lim_n \mathcal{P}_n(-\infty, x] = \mathcal{P}(-\infty, x]$  para cualquier  $x$  real pero  $\mathcal{P}$  no es una medida de probabilidad.

Estamos definiendo la convergencia de una sucesión de medidas de probabilidad a otra medida de probabilidad bien en términos de las propias medidas, bien en términos de sus funciones de distribución; pero hasta ahora, los tipos de convergencia estudiados hasta el momento (casi segura o en probabilidad) se referían a convergencia de una sucesión de variables aleatorias a otra variable aleatoria. Fácilmente podemos extender el concepto a sucesiones de variables aleatorias.

**Definición 6.2 (Convergencia en ley de variables aleatorias)** Sean  $\{X_n\}_{n \geq 1}$  y  $X$  variables aleatorias definidas sobre un mismo espacio de probabilidad. Denotamos por  $\{\mathcal{P}_n\}_{n \geq 1}$  y  $\mathcal{P}$  sus distribuciones de probabilidad respectivas. Se dice que la sucesión  $\{X_n\}_{n \geq 1}$  converge en ley a  $X$ ,  $X_n \xrightarrow{L} X$ , si  $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$  o, equivalentemente, si  $\lim_n \mathcal{P}(X_n \leq x) = \mathcal{P}(X \leq x)$  para todo  $x$  tal que  $\mathcal{P}(X = x) = 0$ .

Recordemos los tres tipos de convergencia para sucesiones de variables que hemos definido:

1.  $X_n \xrightarrow{a.s.} X$  si  $\mathcal{P}(\lim_n X_n = X) = 1$ .
2.  $X_n \xrightarrow{p} X$  si  $\lim_n \mathcal{P}(|X_n - X| > \epsilon) = 0$  para cualquier  $\epsilon > 0$ .
3.  $X_n \xrightarrow{L} X$  si  $\lim_n \mathcal{P}(X_n \leq x) = \mathcal{P}(X \leq x)$  para cualquier  $x$  verificando que  $\mathcal{P}(X = x) = 0$ .

Notemos que en la definición de las convergencias *casi segura* y *en probabilidad* estamos considerando probabilidades de sucesos en cuya definición intervienen las  $X_n$  y la  $X$  simultáneamente. Obviamente esto obliga a que tanto las  $X_n$  como la  $X$  estén definidas sobre un mismo espacio de probabilidad. Por el contrario, en la definición de convergencia en ley tenemos por una parte sucesos definidos únicamente a partir de cada  $X_n$  y por otra parte probabilidades de sucesos donde sólo interviene  $X$ . En ningún momento comparamos  $X_n$  con  $X$ . Una consecuencia inmediata es que las  $X_n$  y  $X$  pueden estar definidas sobre diferentes espacios de probabilidad y la definición sigue siendo válida. Sin embargo, esta generalización no nos aporta ninguna ventaja en lo que sigue.

¿Qué relación existe entre estos tres tipos de convergencia? En el siguiente teorema vemos que el nombre de convergencia débil no es gratuito porque es el tipo de convergencia más débil de las tres.

**Teorema 6.1 (Relaciones entre convergencias)** Sean  $X_n$  y  $X$  variables aleatorias definidas sobre un mismo espacio de probabilidad entonces,

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X \implies X_n \xrightarrow{L} X$$

**Demostración.-** La primera implicación ya la hemos probado en el capítulo anterior. Supongamos que  $X_n \xrightarrow{p} X$  y sea  $x$  tal que  $\mathcal{P}(X = x) = 0$ . Se verifica que

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

y

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

Expresando conjuntamente las dos desigualdades tenemos,

$$\mathcal{P}(X \leq x - \epsilon) - \mathcal{P}(|X_n - X| > \epsilon) \leq \mathcal{P}(X_n \leq x) \leq \mathcal{P}(X \leq x + \epsilon) + \mathcal{P}(|X_n - X| > \epsilon),$$

pero  $\lim_n \mathcal{P}(|X_n - X| > \epsilon) = 0$  para cualquier  $\epsilon$  positivo por lo que

$$\mathcal{P}(X \leq x - \epsilon) \leq \liminf_n \mathcal{P}(X_n \leq x) \leq \limsup_n \mathcal{P}(X_n \leq x) \leq \mathcal{P}(X \leq x + \epsilon)$$

y, puesto que  $\mathcal{P}(X = x) = 0$  es equivalente a que  $F(x) = \mathcal{P}(X \leq x)$  sea continua en  $x$ , se sigue el resultado.  $\square$

Ninguna de las implicaciones contrarias en el teorema anterior es cierta. Ya vimos un ejemplo donde mostrábamos que la convergencia en probabilidad no implica la convergencia casi segura. Tampoco la convergencia débil implica en general la convergencia en probabilidad.

**Ejemplo 6.3** Si  $\{X_n\}_{n \geq 1}$  y  $X$  son variables Bernoulli con  $p = 1/2$ , entonces para  $0 < \epsilon < 1$  tendremos  $\mathcal{P}(|X_n - X| > \epsilon) = \mathcal{P}(|X_n - X| = 1) = \frac{1}{2}$  por lo que  $X_n$  no converge en probabilidad a  $X$  y sin embargo si que hay convergencia débil ya que  $X_n$  y  $X$  tienen la misma distribución de probabilidad.

Existe una situación de gran interés práctico donde la convergencia en probabilidad y en ley son equivalentes. En efecto, si  $a \in \mathcal{R}$  entonces  $X_n \xrightarrow{p} a$  sí y solo sí  $X_n \xrightarrow{L} a$ . Probar este resultado es muy simple y lo dejamos al cuidado del lector.

La importancia de la convergencia débil en Probabilidad se deriva del hecho de que el Teorema Central del Límite asegura una convergencia de este tipo. Este resultado (más bien conjunto de resultados pues existen diferentes versiones del mismo) en la versión de Lindeberg y Levy afirma que si tenemos una sucesión de variables aleatorias  $\{X_n\}$  independientes y con la misma distribución, con media  $m$  y varianza  $\sigma^2$  finitas, entonces  $n^{1/2} \frac{\bar{X}_n - m}{\sigma} \xrightarrow{L} Z$ , siendo  $Z$  una variable con distribución normal estándar. Este resultado es una de las razones que justifican el uso intensivo que se hace de la distribución normal en el análisis de datos. El estudio que estamos haciendo de la convergencia débil y el que haremos de la función característica está justificado en orden a probar el teorema central.

## 6.2. Relación entre convergencia débil de medidas y convergencia de integrales

Hemos definido un tipo de convergencia de medidas de probabilidad,  $\{\mathcal{P}_n\}$ , a otra medida de probabilidad,  $\mathcal{P}$ , que supone que a medida que  $n$  crece  $\mathcal{P}_n$  se *aproxima*, en algún sentido, a  $\mathcal{P}$ . Si  $f$  es una función integrable entonces es de esperar que  $\int f d\mathcal{P}_n$  se aproxime a  $\int f d\mathcal{P}$ . Esto no es cierto en general, pero si  $f$  es *continua y acotada* entonces  $\lim_n \int f d\mathcal{P}_n = \int f d\mathcal{P}$ .

En efecto, tomemos  $\epsilon > 0$  fijo, si  $F$  denota la función de distribución de  $\mathcal{P}$ , entonces  $D = \{x : x \text{ es punto de continuidad de } F\} = \{x : \mathcal{P}\{x\} = 0\}$  es un subconjunto denso de  $\mathcal{R}$ . Existirán por tanto  $a, b \in D$  con  $a < b$  tales que  $\mathcal{P}(a, b]^c < \epsilon$ . Tendremos,

$$\begin{aligned} d_n &= \left| \int f d\mathcal{P}_n - \int f d\mathcal{P} \right| \leq \\ &\leq \left| \int_{(a,b]} f d\mathcal{P}_n - \int_{(a,b]} f d\mathcal{P} \right| + \left| \int_{(a,b]^c} f d\mathcal{P}_n \right| + \left| \int_{(a,b]^c} f d\mathcal{P} \right| \\ &\leq \sup_{x \in \mathcal{R}} f(x)(\epsilon + \mathcal{P}_n(a, b]^c) + \left| \int_{(a,b]} f d\mathcal{P}_n - \int_{(a,b]} f d\mathcal{P} \right|, \end{aligned}$$

y  $f$  continua sea uniformemente continua en  $[a, b]$  por lo que existe un  $\delta$  positivo verificando que si  $|x - y| < \delta$  entonces  $|f(x) - f(y)| < \epsilon$ . Considerando otra vez que  $D$  es denso existirán  $a_0 = a < a_1 < \dots < a_n < a_{n+1} = b$  con  $a_{i+1} - a_i < \delta$  para  $i = 1, \dots, n$ .

$$\begin{aligned} \left| \int_{(a,b]} f d\mathcal{P}_n - \int_{(a,b]} f d\mathcal{P} \right| &\leq \sum_{i=0}^n \left| \int_{(a_i, a_{i+1}]} f d\mathcal{P}_n - \int_{(a_i, a_{i+1}]} f d\mathcal{P} \right| \\ &= \sum_{i=0}^n \left| \int_{(a_i, a_{i+1}]} (f - f(a_i)) d\mathcal{P}_n - \int_{(a_i, a_{i+1}]} (f - f(a_i)) d\mathcal{P} + f(a_i)(\mathcal{P}_n(a_i, a_{i+1}] - \mathcal{P}(a_i, a_{i+1}]) \right| \\ &\leq 2\epsilon(\mathcal{P}_n(a, b] + \mathcal{P}(a, b]) + [\sup_{x \in \mathcal{R}} f(x)](\mathcal{P}_n(a_i, a_{i+1}] - \mathcal{P}(a_i, a_{i+1}])). \end{aligned}$$

Pero  $a_i, a_{i+1} \in D$  de donde  $\lim_n \mathcal{P}_n(a_i, a_{i+1}] = \mathcal{P}(a_i, a_{i+1}]$  para  $i = 0, \dots, n$ . Pasando al límite

$$\lim_n d_n \leq 2\epsilon[\sup_{x \in \mathcal{R}} f(x)] + 4\epsilon\mathcal{P}(a, b] \leq 2\epsilon[\sup_{x \in \mathcal{R}} f(x) + 2], \quad \forall \epsilon > 0$$

y por lo tanto  $\lim_n \int f d\mathcal{P}_n = \int f d\mathcal{P}$ . El recíproco también es cierto.

**Teorema 6.2** *Son equivalentes las dos afirmaciones siguientes:*

1.  $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$ .

2.  $\lim_n \int f d\mathcal{P}_n = \int f d\mathcal{P}$  para cualquier función  $f$  continua y acotada.

**Demostración.-** Supongamos que la segunda afirmación es cierta. Tomamos  $x$  un punto de continuidad de  $F$  y  $\epsilon > 0$ . Consideremos

$$f_1(y) = \begin{cases} 1, & \text{si } y \leq x - \epsilon, \\ \frac{x-y}{\epsilon}, & \text{si } x - \epsilon < y \leq x, \\ 0, & \text{en el resto,} \end{cases}$$

y

$$f_2(y) = \begin{cases} 1, & \text{si } y \leq x, \\ 1 - \frac{x-y}{\epsilon}, & \text{si } x < y \leq x + \epsilon, \\ 0, & \text{en el resto.} \end{cases}$$

Tenemos las siguientes desigualdades,

$$\begin{aligned} F(x - \epsilon) = \mathcal{P}(-\infty, x - \epsilon] &\leq \int f_1 d\mathcal{P} = \lim_n \int f_1 d\mathcal{P}_n \leq \liminf_n \mathcal{P}_n(-\infty, x] \leq \\ \limsup_n \mathcal{P}_n(-\infty, x] &\leq \lim_n \int f_2 d\mathcal{P}_n = \int f_2 d\mathcal{P} \leq \mathcal{P}(-\infty, x + \epsilon] = F(x + \epsilon), \end{aligned}$$

Puesto que  $x$  es un punto de continuidad de  $F$  haciendo tender  $\epsilon$  a cero

$$F(x) = \liminf_n \mathcal{P}_n(-\infty, x] = \limsup_n \mathcal{P}_n(-\infty, x] = \lim_n F_n(x)$$

, lo que prueba la implicación contraria.  $\square$

Visto el teorema anterior podíamos haber definido la convergencia débil de medidas de probabilidad a partir de la convergencia de las integrales de las funciones continuas y acotadas. Hay además razones de conveniencia matemática para hacerlo (ver Shiriyayev, 1984, páginas 306 a 313), sin embargo preferimos 6.1 por ser mucho más intuitiva y porque en la recta real, donde nos proponemos trabajar, son equivalentes.

### 6.3. Un criterio de convergencia para sucesiones ajustadas

Dada una sucesión de medidas de probabilidad,  $\{\mathcal{P}_n\}_{n \geq 1}$ , antes de plantearnos a qué medida converge habremos de comprobar si converge o no. En este apartado proponemos un criterio de convergencia ya conocido en una situación simple, la de las sucesiones numéricas. Supongamos  $\{x_n\}_{n \geq 1}$  una sucesión acotada de números reales. Un conocido resultado afirma que  $\{x_n\}_{n \geq 1}$  es convergente si y solo si todas sus subsucesiones convergentes convergen a un mismo valor. O si se prefiere,  $\lim_n x_n = x$  sí y solo sí para cualquier subsucesión convergente  $\{x_{n_k}\}_{k \geq 1}$  de  $\{x_n\}_{n \geq 1}$ ,  $\lim_k x_{n_k} = x$ . Sea

ahora  $\delta_z$  la delta de Dirac en el punto  $z$ . Para cualquier sucesión numérica, en particular  $\{x_n\}_{n \geq 1}$ , se verifica que  $\lim_n x_n = x$  equivale a (w)  $\lim_n \delta_{x_n} = \delta_x$ . Por tanto el resultado enunciado para sucesiones numéricas podemos expresarlo como,  $\delta_{x_n} \xrightarrow{a.s.} \delta_x$  sí y solo sí cualquier subsucesión débilmente convergente,  $\{\delta_{x_{n_k}}\}_{k \geq 1}$  de  $\{\delta_{x_n}\}_{n \geq 1}$  verifica  $\delta_{x_{n_k}} \xrightarrow{a.s.} \delta_x$ .

Este resultado enunciado para las delta de Dirac es cierto en un contexto más amplio. Solamente hemos impuesto una condición previa, la acotación de  $\{x_n\}_{n \geq 1}$ . En términos de las medidas  $\{\delta_{x_n}\}_{n \geq 1}$  esto lo podemos expresar diciendo que para cualquier  $\epsilon$  positivo, y en particular para valores entre cero y uno, encontramos valores  $a, b \in \mathcal{R}$  tales que  $\delta_{x_n}(a, b]^c < \epsilon$  para todos los  $n$ . Consideremos las dos definiciones siguientes.

- Definición 6.3** 1.  $\{\mathcal{P}_n\}_{n \geq 1}$  diremos que es relativamente compacta si cada sucesión de elementos de  $\{\mathcal{P}_n\}_{n \geq 1}$  contiene una subsucesión débilmente convergente a una medida de probabilidad.
2.  $\{\mathcal{P}_n\}_{n \geq 1}$  es ajustada si para cada  $\epsilon$  positivo encontramos dos valores reales  $a, b$  tales que  $\mathcal{P}_n(a, b]^c < \epsilon$  para cualquier  $n$ .

Es importante notar que en la definición de sucesión relativamente compacta no exigimos que la medida límite pertenezca a la sucesión. Por esta razón hablamos de relativamente compacta y no de compacta. Prohorov probó que ambas definiciones son equivalentes. Nuestro criterio de convergencia será una consecuencia inmediata de este resultado.

**Teorema 6.3 (Prohorov)** Si  $\{\mathcal{P}_n\}_{n \geq 1}$  son medidas de probabilidad sobre la recta real entonces la sucesión es relativamente compacta sí y solo sí es ajustada.

**Demostración.-** Para probar la necesidad, supongamos que  $\{\mathcal{P}_n\}_{n \geq 1}$  es relativamente compacta pero no ajustada. Entonces, dado  $\epsilon$  positivo y  $m$  entero no negativo encontramos  $n_m$  tal que  $\mathcal{P}_{n_m}(-m, m]^c > \epsilon$ . La sucesión  $\{\mathcal{P}_{n_m}\}_{m \geq 1}$  es relativamente compacta por lo que tendrá una subsucesión  $\{\mathcal{P}_{n_{m(k)}}\}_{k \geq 1}$  y una medida de probabilidad  $\mathcal{P}$  tal que  $\mathcal{P}_{n_{m(k)}} \xrightarrow{w} \mathcal{P}$ . Dada  $\mathcal{P}$  encontramos  $a < b$  tales que  $\mathcal{P}\{a\} = \mathcal{P}\{b\} = 0$  y  $\mathcal{P}(a, b]^c < \epsilon$ . A partir de un cierto  $k$ ,  $(a, b] \subset (-m(k), m(k)]$  y  $\mathcal{P}_{n_{m(k)}}(a, b]^c \geq \mathcal{P}_{n_{m(k)}}(-m(k), m(k)]^c > \epsilon$ ; pero:  $\mathcal{P}_{n_{m(k)}}(a, b]^c = \mathcal{P}_{n_{m(k)}}(-\infty, a] + 1 - \mathcal{P}_{n_{m(k)}}(b, +\infty)$  tiende con  $k$  a  $\mathcal{P}(-\infty, a] + 1 - \mathcal{P}(b, +\infty)$ . En consecuencia  $\mathcal{P}(a, b]^c > \epsilon$  en contra de la elección de  $a$  y  $b$ .  $\square$

Para probar la suficiencia necesitamos el siguiente teorema previo.

**Teorema 6.4 (Teorema de Helly-Bray)** Para una sucesión cualquiera de funciones de distribución  $\{F_n\}_{n \geq 1}$  siempre existe una subsucesión,  $\{F_{n_k}\}_{k \geq 1}$ , y una función continua por la derecha y no decreciente  $F$ , tal que  $\lim_k F_{n_k}(x) = F(x)$  si  $x$  es un punto de continuidad de  $F$ .

La prueba del resultado se apoya en el método diagonal de Cantor.

**Lema 6.1 (Método diagonal de Cantor)** Sean

$$\begin{array}{l} x_{11}, x_{12}, x_{13}, \dots \\ x_{21}, x_{22}, x_{23}, \dots \\ \dots \dots \dots \end{array}$$

tales que la sucesión numérica formada por cada fila está acotada. Existe  $\{n_k\}_{k \geq 1}$  tal que  $\{x_{rn_k}\}_{k \geq 1}$  es convergente para cualquier  $r$ .

**Demostración.-**  $\{x_{1j}\}_{j \geq 1}$  está acotada por lo que tendrá una subsucesión convergente  $\{x_{1n_{1j}}\}_{j \geq 1}$ ,  $\{x_{2n_{1j}}\}_{j \geq 1}$  está también acotada y podremos determinar una subsucesión  $\{x_{2n_{2j}}\}_{j \geq 1}$  que converge y así sucesivamente. Tomemos  $n_k = n_{kk}$ , para  $r$  fijo tendremos que  $\{x_{rn_k}\}_{k \geq r}$  es una subsucesión de  $\{x_{rn_{rj}}\}_{j \geq 1}$  elegida de forma que converja. En definitiva  $\{x_{rn_k}\}_{k \geq 1}$  es convergente para cualquier  $r$ .  $\square$

**Demostración del teorema de Helly-Bray.-** Tenemos una cantidad numerable de sucesiones  $\{F_n(r)\}_{n \geq 1}$  cuando  $r$  varía en el conjunto de los racionales. Son acotadas por lo que podemos aplicar el método diagonal de Cantor. Encontramos una sucesión  $\{n_k\}_{k \geq 1}$  de enteros no negativos tal que existe  $G(r) = \lim_k F_{n_k}(r)$  para todos los racionales. Definimos:  $F(x) = \inf\{G(r) : x < r, r \in \mathcal{Q}\}$  que es trivialmente no decreciente. También es continua por la derecha: para  $x$  y  $\epsilon > 0$  por la definición de  $F$  encontramos un  $r \in \mathcal{Q}$  tal que  $G(r) < F(x) + \epsilon$ ; todo  $y$  que verifique  $x \leq y < r$  verificará  $F(x) \leq F(y) \leq G(r) < F(x) + \epsilon$ .

Veamos que  $\lim_k F_{n_k}(x) = F(x)$  si  $F$  es continua en  $x$ . Dado  $\epsilon$  positivo encontramos un  $y$  tal que  $y < x$  y  $F(x) - \epsilon < F(y)$  así como dos racionales  $r$  y  $s$  que verifican  $y < r < x < s$  con  $G(s) < F(x) + \epsilon$ . Tenemos las siguientes desigualdades:  $F(x) - \epsilon < F(y) < G(r) \leq G(s) < F(x) + \epsilon$  y  $F_{n_k}(r) \leq F_{n_k}(x) \leq F_{n_k}(s)$ . Pero  $\lim_k F_{n_k}(r) = F(r)$  y  $\lim_k F_{n_k}(s) = F(s)$  por ser  $r$  y  $s$  racionales. Para  $k$  suficientemente grande  $F(x) - \epsilon < F_{n_k}(r) \leq F_{n_k}(x) \leq F_{n_k}(s) < F(x) + \epsilon$  de donde se sigue el resultado.  $\square$

En el teorema de Helly-Bray no tiene porqué verificarse que la función límite  $F$  sea de distribución,  $F_n = 1_{[n, +\infty)}$  sería un ejemplo.

**Demostración de la suficiencia del teorema de Prohorv.-** Suponemos  $\{\mathcal{P}_n\}_{n \geq 1}$  ajustada.  $F_n$  es la función de distribución de  $\mathcal{P}_n$ . Dada  $\{F_n\}_{n \geq 1}$  por el teorema de Helly y Bray existe una subsucesión  $\{F_{n_k}\}_{k \geq 1}$  y  $F$  monótona creciente y continua por la derecha con  $\lim_k F_{n_k}(x) = F(x)$  en los puntos de continuidad de  $F$ . Hemos de deducir del ajustamiento de  $\{F_n\}_{n \geq 1}$  que  $F$  verifica  $F(+\infty) = 1$  y  $F(-\infty) = 0$  y el resultado estará probado pues  $\mathcal{P}_{n_k}$  convergerá débilmente a la medida de probabilidad con función de distribución  $F$ . Tomamos  $\epsilon > 0$  arbitrario y encontramos  $a < b$  con  $\mathcal{P}_{n_k}(a, b]^c < \epsilon$  para cualquier  $k$ . Elegimos  $a'$  y  $b'$  puntos de continuidad de  $F$  con  $a' < a$  y  $b < b'$ . Entonces

$$\max\{F_{n_k}(a'), 1 - F_{n_k}(b')\} \leq \mathcal{P}_{n_k}(a', b']^c \leq \mathcal{P}_{n_k}(a, b]^c < \epsilon, \quad (6.3)$$



pero  $\lim_k F_{n_k}(a') = F(a')$  y  $\lim_k F_{n_k}(b') = F(b')$  y de 6.3 obtenemos que  $F(a') < \epsilon$  y  $F(b') > 1 - \epsilon$ .  $F$  es pues una función de distribución.  $\square$

**Teorema 6.5 (Criterio de convergencia)** *Si  $\{\mathcal{P}_n\}_{n \geq 1}$  es una sucesión ajustada de medidas de probabilidad tal que cualquier subsucesión débilmente convergente converge a una misma medida de probabilidad  $\mathcal{P}$  entonces  $(w) \lim_n \mathcal{P}_n = \mathcal{P}$ .*

**Demostración.-** Si no es cierto que  $(w) \lim_n \mathcal{P}_n = \mathcal{P}$  encontramos un  $x$  con  $\mathcal{P}\{x\} = 0$  verificando  $\lim_n \mathcal{P}_n(-\infty, x] \neq \mathcal{P}(-\infty, x]$ . Dado  $\epsilon > 0$  podemos determinar  $\{n_k\}_{k \geq 1}$  verificando

$$|\mathcal{P}_{n_k}(-\infty, x] - \mathcal{P}(-\infty, x]| > \epsilon. \quad (6.4)$$

Pero  $\{\mathcal{P}_{n_k}\}_{k \geq 1}$  es ajustada y, por el teorema de Prohorov, relativamente compacta de donde deducimos que ha de tener una subsucesión  $\{\mathcal{P}_{n_{k(j)}}\}_{j \geq 1}$  que converge débilmente y, por la hipótesis del teorema, ha de converger a  $\mathcal{P}$  lo que por 6.4 no es posible.  $\square$



## Capítulo 7

# Teorema Central del Límite

La convergencia de sucesiones de variables aleatorias fue estudiada en el Capítulo 6 a través de las Leyes de los Grandes números, aunque solamente resolvíamos el caso de convergencia a constantes. Una situación sin duda más rica, ya esbozada en uno de los corolarios de la Ley Cero-Uno de Kolmogorov, es aquella que conduce a la convergencia a otra variable aleatoria. Para poder resolver con éxito este problema ha sido necesario esperar a disponer de la utilísima herramienta que *la función característica* supone (Lyapounov dixit) y sobretodo de los teoremas de continuidad y unicidad de Lévy, a los que hay que añadir la interesante propiedad que permite obtener la función característica de una suma de variables aleatorias independientes mediante el producto de las correspondientes funciones características.

El precedente de los resultados que a continuación presentaremos se debe a De Moivre y Laplace quienes demostraron que una  $B(n, p)$ , adecuadamente normalizada, convergía en ley a una  $N(0, 1)$ . Este resultado lo recogeremos como un caso particular del Teorema Central del Límite en su versión clásica conocida, como Teorema de Lindeberg-Lévy.

Para cerrar esta pequeña introducción digamos que fue Polya quien acuñó el término *Teorema Central del Límite* (TCL) para referirse al conjunto de resultados que tratan el problema de la convergencia en ley de sucesiones de variables aleatorias, pretendiendo con dicho nombre reflejar el hecho de que este problema había sido el centro de las investigaciones en Teoría de la Probabilidad desde que De Moivre y Laplace presentaron sus resultados.

### 7.1. Función característica

La *función característica* es una herramienta de gran utilidad en Teoría de la Probabilidad, una de cuyas mayores virtudes reside en facilitar la obtención de la distribución

de probabilidad de la suma de variables aleatorias y la del límite de sucesiones de variables aleatorias, situaciones ambas que aparecen con frecuencia en Inferencia Estadística.

### 7.1.1. Definición y propiedades

**Definición 7.1** Sea  $X$  una variable aleatoria y sea  $t \in \mathbb{R}$ . La función característica de  $X$ ,  $\phi_X(t)$ , se define como  $E(e^{itX}) = E(\cos tX) + iE(\sin tX)$ .

Como  $|e^{itX}| \leq 1$ ,  $\forall t$ ,  $\phi_X(t)$  existe siempre y está definida  $\forall t \in \mathbb{R}$ . Para su obtención recordemos que,

**Caso discreto.-** Si  $X$  es una v. a. discreta con soporte  $D$  y función de probabilidad  $f(x)$ ,

$$\phi_X(t) = \sum_{x \in D} e^{itx} f(x). \quad (7.1)$$

**Caso continuo.-** Si  $X$  es una v. a. continua con función de densidad de probabilidad  $f(x)$ ,

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx. \quad (7.2)$$

De la definición se derivan, entre otras, las siguientes propiedades:

**P1)**  $\phi_X(0) = 1$

**P2)**  $|\phi_X(t)| \leq 1$

**P3)**  $\phi_X(t)$  es uniformemente continua

**P4)** Si definimos  $Y = aX + b$ ,

$$\phi_Y(t) = E(e^{itY}) = E(e^{it(aX+b)}) = e^{itb} \phi_X(at)$$

**P5)** Si  $E(X^n)$  existe, la función característica es  $n$  veces diferenciable y  $\forall k \leq n$  se verifica  $\phi_X^{(k)}(0) = i^k E(X^k)$

La propiedad 5 establece un interesante relación entre las derivadas de  $\phi_X(t)$  y los momentos de  $X$  cuando estos existen, relación que permite desarrollar  $\phi_X(t)$  en serie de potencias. En efecto, si  $E(X^n)$  existe  $\forall n$ , entonces,

$$\phi_X(t) = \sum_{k \geq 0} \frac{i^k E(X^k)}{k!} t^k. \quad (7.3)$$

### 7.1.2. Función característica e independencia

Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes y definimos  $Y = X_1 + X_2 + \dots + X_n$ , tendremos

$$\phi_Y(t) = E(e^{it(X_1+X_2+\dots+X_n)}) = E\left(\prod_{k=1}^n e^{itX_k}\right) = \prod_{k=1}^n E(e^{itX_k}) = \prod_{k=1}^n \phi_{X_k}(t), \quad (7.4)$$

expresión que permite obtener con facilidad la función característica de la suma de variables independientes y cuya utilidad pondremos de manifiesto de inmediato en los ejemplos que siguen.

### 7.1.3. Ejemplos de funciones características

**Bernoulli.-** Si  $X \sim B(1, p)$ , su función de probabilidad vale  $f(x) = p^x(1-p)^{1-x}$ , si  $x \in \{0, 1\}$ , y  $f(x) = 0$ , si  $x \notin \{0, 1\}$ . Aplicando (7.1),

$$\phi_X(t) = e^{it,0}q + e^{it,1}p = q + pe^{it}.$$

**Binomial.-** Si  $X \sim B(n, p)$ , recordemos que  $X = X_1 + X_2 + \dots + X_n$ , con las  $X_k \sim B(1, p)$ ,  $\forall k$  e independientes. Si aplicamos (7.4),

$$\phi_X(t) = \prod_{k=1}^n (q + pe^{it}) = (q + pe^{it})^n.$$

**Poisson.-** Si  $X \sim P(\lambda)$ , su función de probabilidad es  $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ , si  $x \in \{0, 1, \dots\}$ , y  $f(x) = 0$ , en caso contrario. Aplicando (7.1),

$$\phi_X(t) = \sum_{x \geq 0} e^{itx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x \geq 0} \frac{(\lambda e^{it})^x}{x!} = e^{\lambda(e^{it}-1)}.$$

**Normal tipificada.-** Si  $Z \sim N(0, 1)$ , sabemos que existen los momentos de cualquier orden y en particular,  $E(Z^{2n+1}) = 0$ ,  $\forall n$  y  $E(Z^{2n}) = \frac{(2n)!}{2^n n!}$ ,  $\forall n$ . Aplicando (7.3),

$$\phi_Z(t) = \sum_{n \geq 0} \frac{i^{2n}(2n)!}{2^n(2n)!n!} t^{2n} = \sum_{n \geq 0} \frac{\left(\frac{it}{2}\right)^{2n}}{n!} = \sum_{n \geq 0} \frac{\left(-\frac{t^2}{2}\right)^n}{n!} = e^{-\frac{t^2}{2}}.$$

Para obtener  $\phi_X(t)$  si  $X \sim N(\mu, \sigma^2)$ , podemos utilizar el resultado anterior y P4. En efecto, recordemos que  $X$  puede expresarse en función de  $Z$  mediante  $X = \mu + \sigma Z$  y aplicando P4,

$$\phi_X(t) = e^{i\mu t} \phi_Z(\sigma t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

**Observación 7.1** Obsérvese que  $\text{Im}(\phi_Z(t)) = 0$ . El lector puede comprobar que no se trata de un resultado exclusivo de la Normal tipificada, si no de una propiedad que poseen todas las v. a. con distribución de probabilidad simétrica, es decir, aquellas que verifican  $(P(X \geq x) = P(X \leq -x))$ .

**Gamma.-** Si  $X \sim G(\alpha, \lambda)$ , su función de densidad de probabilidad viene dada por

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0, \end{cases}$$

por lo que aplicando (7.2),

$$\phi_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{itx} x^{\alpha-1} e^{-\lambda x} dx,$$

que con el cambio  $y = x(1 - it/\lambda)$  conduce a

$$\phi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-\alpha}.$$

Hay dos casos particulares que merecen ser mencionados:

**Exponencial.-** Cuando  $\alpha = 1$ ,  $X \sim \text{Exp}(\lambda)$  y

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

**Chi-cuadrado.-** Cuando  $\alpha = n/2$  y  $\lambda = 1/2$ ,  $X \sim \chi_n^2$  y

$$\phi_X(t) = (1 - 2it)^{-\frac{n}{2}}.$$

#### 7.1.4. Teorema de inversión. Unicidad

Hemos obtenido la función característica de una v. a.  $X$  a partir de su distribución de probabilidad, pero es posible proceder de manera inversa por cuanto el conocimiento de  $\phi_X(t)$  permite obtener  $F(x)$ .

**Teorema 7.1 (Inversión)** Sean  $\phi_X(t)$  y  $F(x)$  las funciones característica y de distribución de la v. a.  $X$  y sean  $x_1 \leq x_2$  sendos puntos de continuidad de  $F(x)$ , entonces

$$F(x_2) - F(x_1) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \phi_X(t) dt.$$

Este resultado permite obtener  $F(x)$  si  $x$  es un punto de continuidad. Basta para ello que  $x_1 \rightarrow \infty$  a través de puntos de continuidad, y como  $F(x)$  es continua por la derecha, la tendremos completamente determinada.

Pero este teorema tiene una trascendencia mayor por cuanto implica la *unicidad* de la función característica, que no por casualidad recibe este nombre, porque *caracteriza*, al igual que lo hacen otras funciones asociadas a  $X$  (la de distribución, la de probabilidad o densidad de probabilidad, ...), su distribución de probabilidad. Podemos afirmar que si dos variables  $X$  e  $Y$  comparten la misma función característica tienen idéntica distribución de probabilidad. La combinación de este resultado con las propiedades antes enunciadas da lugar a una potente herramienta que facilita el estudio y obtención de las distribuciones de probabilidad asociadas a la suma de variables independientes. Veámoslo con algunos ejemplos.

1. Si  $X_k \sim B(n_k, p)$ ,  $k = 1, 2, \dots, m$  y son independientes, al definir  $X = \sum_{k=1}^m X_k$ , sabemos por (7.4) que

$$\phi_X(t) = \prod_{k=1}^m \phi_{X_k}(t) = \prod_{k=1}^m (q + pe^{it})^{n_k} = (q + pe^{it})^n,$$

con  $n = n_1 + n_2 + \dots + n_m$ . Pero siendo esta la función característica de una variable  $B(n, p)$ , podemos afirmar que  $X \sim B(n, p)$ .

2. Si nuestra suma es ahora de variables Poisson independientes,  $X_k \sim P(\lambda_k)$ , entonces

$$\phi_X(t) = \prod_{k=1}^m \phi_{X_k}(t) = \prod_{k=1}^m e^{\lambda_k(e^{it}-1)} = e^{\lambda(e^{it}-1)},$$

con  $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_m$ . Así pues,  $X \sim P(\lambda)$ .

3. En el caso de que la suma esté formada por variables  $N(\mu_k, \sigma_k^2)$  independientes,

$$\phi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}, \quad (7.5)$$

con  $\mu = \mu_1 + \mu_2 + \dots + \mu_m$  y  $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_m^2$ . Así pues  $X \sim N(\mu, \sigma^2)$ .

4. En el caso de que la suma esté formada por  $n$  variables independientes todas ellas  $Exp(\lambda)$ ,

$$\phi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^n = \left(1 - \frac{it}{\lambda}\right)^{-n},$$

y su distribución será la de una  $G(n, \lambda)$ .

5. Sea ahora  $Y = X^2$  con  $X \sim N(0, 1)$ , su función característica viene dada por

$$\phi_Y(t) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{x^2}{2}(1-2it)} = \frac{1}{(1-2it)^{\frac{1}{2}}},$$

lo que nos asegura que  $Y \sim \chi_1^2$ .

### 7.1.5. Teorema de continuidad

Se trata del último de los resultados que presentaremos y permite conocer la convergencia de una sucesión de v. a. a través de la convergencia puntual de la sucesión de sus funciones características.

**Teorema 7.2 (Directo)** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de v. a. y  $\{F_n(x)\}_{n \geq 1}$  y  $\{\phi_n(t)\}_{n \geq 1}$  las respectivas sucesiones de sus funciones de distribución y características. Sea  $X$  una v. a. y  $F(x)$  y  $\phi(t)$  sus funciones de distribución y característica, respectivamente. Si  $F_n \xrightarrow{w} F$  (es decir,  $X_n \xrightarrow{L} X$ ), entonces

$$\phi_n(t) \longrightarrow \phi(t), \quad \forall t \in R.$$

Resultado que se completa con el teorema inverso.

**Teorema 7.3 (Inverso)** Sea  $\{\phi_n(t)\}_{n \geq 1}$  una sucesión de funciones características y  $\{F_n(x)\}_{n \geq 1}$  la sucesión de funciones de distribución asociadas. Sea  $\phi(t)$  una función continua, si  $\forall t \in R, \phi_n(t) \rightarrow \phi(t)$ , entonces

$$F_n \xrightarrow{w} F,$$

donde  $F(x)$  es una función de distribución cuya función característica es  $\phi(t)$ .

## 7.2. El Teorema de Lindeberg-Lévy

**Teorema 7.4** Sea  $\{X_n\}$  una sucesión de variables aleatorias i. i. d. con  $E(X_n) = m$  y  $\text{var}(X_n) = \sigma^2, \forall n$ . Para  $S_n = \sum_{k=1}^n X_k$  se verifica que

$$\frac{\frac{S_n}{n} - m}{\sigma/\sqrt{n}} = \frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{L} N(0, 1).$$

**Demostración.-** A partir de  $X_n$  definimos  $Z_n = \frac{X_n - m}{\sigma}$ , que constituirán una nueva sucesión de variables independientes con media 0 y varianza 1. La función característica de  $Z_n$  podrá expresarse  $\varphi_{Z_n} = 1 - (t^2/2) + h(t)$ , con

$$h(t) = -t^2 \int_0^1 (1-s) \int_{-\infty}^{+\infty} z^2 (e^{istz_n} - 1) dF_{Z_n} ds.$$



Por otra parte  $\frac{S_n - nm}{\sigma\sqrt{n}} = \sum_{k=1}^n \frac{Z_k}{\sqrt{n}}$  y su función característica,  $\varphi_n(t)$ , vendrá dada por

$$\varphi_n(t) = \prod_{k=1}^n \varphi_{Z_k} \left( \frac{t}{\sqrt{n}} \right) = \left( 1 - \frac{t^2}{2n} + h \left( \frac{t}{\sqrt{n}} \right) \right)^n = \left( 1 + \frac{g(n)}{n} \right)^n.$$

El teorema de la convergencia dominada nos permite comprobar con facilidad que  $\lim_n nh(\frac{t}{\sqrt{n}}) = 0$ , con lo que la función  $g(n) = -\frac{t^2}{2} + nh(\frac{t}{\sqrt{n}})$  tiende a  $-\frac{t^2}{2}$  cuando  $n \rightarrow +\infty$ . Aplicando un resultado clásico de límites tendremos que

$$\lim_n \varphi_n(t) = \lim_n \left( 1 + \frac{g(n)}{n} \right)^n = e^{-\frac{t^2}{2}},$$

y aplicando el teorema de continuidad obtenemos el resultado enunciado.  $\square$

Si las variables del teorema anterior fueran Bernoulli de parámetro  $p$ , su esperanza común sería  $p$  y su varianza  $p(1-p)$  siendo además  $S_n$  una  $B(n, p)$ . La aplicación del teorema nos permite afirmar que  $\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{L} N(0, 1)$ , lo que se conoce como *teorema de De Moivre-Laplace*.

### 7.3. Las condiciones de Lindeberg y Lyapounov

Una generalización del teorema básico que acabamos de exponer supone relajar la condición de equidistribución exigida a las variables de la sucesión. Los trabajos de Lindeberg y Lyapounov van en este sentido y previamente a su exposición hemos de introducir un nuevo concepto.

**Definición 7.2 (Arreglo triangular)** *Una colección de variables aleatorias*

$$X_{n1}, \dots, X_{nr_n},$$

*independientes para cada  $n$ , recibe el nombre de arreglo triangular. El cambio de subíndice  $n$  no excluye el cambio de espacio de probabilidad.*

Supondremos, no implica pérdida de generalidad, que  $E(X_{nk}) = 0$ ,  $\forall n$  y  $\forall k$  y para cuanto sigue definimos  $S_n = X_{n1} + \dots + X_{nr_n}$ ,  $\sigma_{nk}^2 = E(X_{nk}^2)$  y  $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2$ .

#### 7.3.1. La condición de Lyapounov

La primera generalización se debe a Lyapounov expresada a través del siguiente teorema.

**Teorema 7.5** Si el arreglo triangular  $X_{n1}, \dots, X_{nr_n}$  verifica la existencia de un  $\delta > 0$ , tal que hace integrables a las  $X_{nk}^{2+\delta}$  y

$$\lim_n \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E(|X_{nk}^{2+\delta}|) = 0, \quad (7.6)$$

entonces  $\frac{S_n}{s_n} \xrightarrow{L} N(0, 1)$ .

A (7.6) se la conoce como *condición de Lyapounov*. La demostración del teorema la omitimos porque será una corolario, como comprobaremos más tarde, de la generalización debida a Lindeberg.

### 7.3.2. La condición de Lindeberg

La condición (7.6) tiene el inconveniente de exigir la existencia de momentos estrictamente superiores a 2. Lindeberg sustituye dicha condición por la que lleva su nombre y que se expresa:

$$\lim_n \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}| \geq \epsilon s_n} X_{nk}^2 dP = 0, \quad \forall \epsilon > 0. \quad (7.7)$$

Con esta condición el correspondiente teorema se enuncia de la siguiente forma,

**Teorema 7.6** Si el arreglo triangular  $X_{n1}, \dots, X_{nr_n}$  verifica la condición de Lindeberg, (7.7), entonces  $\frac{S_n}{s_n} \xrightarrow{L} N(0, 1)$ .

**Demostración.-** Podemos suponer que  $s_n^2 = 1$ , para ello basta reemplazar  $X_{nk}$  por  $X_{nk}/s_n$ . Por otra parte, de la relación entre función característica y momentos de la variable aleatoria, podemos aproximar  $\varphi_{X_{nk}} = \varphi_{nk}$  mediante  $1 - \frac{1}{2}t^2\sigma_{nk}^2$ , con un error menor o igual que  $E[\min(|tX_{nk}|^2, |tX_{nk}|^3)]$ . Esta cota,  $\forall \epsilon > 0$ , vale a lo sumo

$$\int_{|X_{nk}| < \epsilon} |tX_{nk}|^2 dP + \int_{|X_{nk}| \geq \epsilon} |tX_{nk}|^2 dP \leq \epsilon |t|^2 \sigma_{nk}^2 + t^2 \int_{|X_{nk}| \geq \epsilon} X_{nk}^2 dP.$$

Al sumar y pasar al límite, como  $\sigma_{nk}^2$  suma 1 y  $\epsilon$  es arbitrario, la condición de Lindeberg nos lleva a

$$\lim_n \sum_{k=1}^{r_n} |\varphi_{nk}(t) - (1 - \frac{1}{2}t^2\sigma_{nk}^2)| = 0,$$

para  $t$  fijo.

Por otra parte de la desigualdad  $\sigma_{nk}^2 \leq \epsilon^2 + \int_{|X_{nk}| \geq \epsilon} X_{nk}^2 dP$ , cierta para cualquier  $\epsilon$ , siempre a través de la condición de Lindeberg se deduce que

$$\max_{1 \leq k \leq r_n} \sigma_{nk}^2 \xrightarrow{n} 0. \quad (7.8)$$

Ello supone que para  $n$  suficientemente grande  $0 \leq (1 - \frac{1}{2}t^2\sigma_{nk}^2) \leq 1$ , y por una conocida propiedad de los complejos de módulo inferior a 1

$$\left| \prod_{k=1}^{r_n} \varphi_{nk}(t) - \prod_{k=1}^{r_n} (1 - \frac{1}{2}t^2\sigma_{nk}^2) \right| \leq \sum_{k=1}^{r_n} |\varphi_{nk}(t) - (1 - \frac{1}{2}t^2\sigma_{nk}^2)|.$$

Esta misma propiedad permite alcanzar la siguiente desigualdad,

$$\left| \prod_{k=1}^{r_n} e^{-t^2\sigma_{nk}^2/2} - \prod_{k=1}^{r_n} (1 - \frac{1}{2}t^2\sigma_{nk}^2) \right| \leq \sum_{k=1}^{r_n} |e^{-t^2\sigma_{nk}^2/2} - 1 + \frac{1}{2}t^2\sigma_{nk}^2| \leq t^4 \sum_{k=1}^{r_n} \sigma_{nk}^4,$$

tendiendo a 0 el último miembro de la cadena de desigualdades debido a (7.8) y a que la suma de los  $\sigma_{nk}^2$  es 1. Hemos por tanto comprobado que la función característica de  $S_n$ ,  $\varphi_{S_n}(t) = \prod_{k=1}^{r_n} \varphi_{nk}(t)$  converge a la de una  $N(0, 1)$ , con lo que el teorema de continuidad hace el resto.  $\square$

Como afirmábamos antes, el teorema debido a Lyapounov es un colorario de este resultado, para ello basta comprobar que (7.6) implica (7.7). En efecto,

$$\begin{aligned} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}| \geq \epsilon s_n} X_{nk}^2 dP &\leq \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}| \geq \epsilon s_n} \frac{|X_{nk}|^{2+\delta}}{\epsilon^\delta s_n^\delta} dP \\ &\leq \frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} \int_{\Omega} |X_{nk}|^{2+\delta} dP \\ &= \frac{1}{\epsilon^\delta} \sum_{k=1}^{r_n} \frac{1}{s_n^{2+\delta}} E(|X_{nk}|^{2+\delta}). \end{aligned}$$

Señalemos también que el teorema clásico de Lindeberg-Lévy en un caso particular del teorema anterior si hacemos  $X_{nk} = X_k$  y  $r_n = n$ , entonces la sucesión entera de las  $X_k$  es independiente y  $E(X_k) = 0$ ,  $\sigma^2 = E(X_k^2)$ ,  $\forall k$  y  $s_n^2 = n\sigma^2$ . La condición de Lindeberg se verificará ahora puesto que

$$\sum_{k=1}^n \frac{1}{n\sigma^2} \int_{|X_1| \geq \epsilon\sigma\sqrt{n}} X_1^2 dP = \frac{1}{n\sigma^2} \int_{|X_1| \geq \epsilon\sigma\sqrt{n}} X_1^2 dP,$$

y tiende a 0 cuando  $n$  tiende a  $+\infty$  porque  $\{|X_1| \geq \epsilon\sigma\sqrt{n}\} \xrightarrow{n} \emptyset$ .

## 7.4. El Teorema de Lindeberg-Feller

Hemos demostrado en la sección anterior la suficiencia de las condiciones de Lindeberg y Lyapounov, la necesidad no es sin embargo siempre cierta. Una sucesión de

normales independientes, adecuadamente elegidas, nos puede servir de contraejemplo. En efecto, tomemos  $X_k \sim N(0, \sigma_k^2)$ , independientes y con sus varianzas verificando  $\sigma_1^2 = 1$  y  $\sigma_n^2 = ns_{n-1}^2$ , lo que supone que  $s_n^2 \approx \sigma_n^2$  pues  $\lim_n \frac{\sigma_n^2}{s_n^2} = 1$ . Es fácil comprobar que la condición de Lindeberg no se cumple en estas circunstancias y ello es debido a que uno de los sumandos de  $S_n$  aporta tanta varianza como todos los demás, impidiendo que se verifique (7.8) que era una consecuencia de la mencionada condición de Lindeberg. Si controlamos el aporte de cada sumando mediante (7.8) podemos conseguir la necesidad de la condición como muestra el siguiente teorema.

**Teorema 7.7 (Teorema de Lindeberg-Feller)** Sea  $X_{n1}, \dots, X_{nr_n}$ ,  $n \geq 1$ , un arreglo triangular que verifica,

$$\max_{1 \leq k \leq r_n} \frac{\sigma_{nk}^2}{s_n^2} \xrightarrow{n} 0,$$

entonces la condición de Lindeberg es necesaria y suficiente para que  $\frac{S_n}{s_n} \xrightarrow{L} N(0, 1)$ .

La desmotivación puramente técnica, como las anteriores, la omitiremos pero el lector interesado puede consultarla en los libros de Shiryaev o Billingley.

## 7.5. Una curiosa aplicación del TCL

De Moivre y Laplace dieron en primer lugar una *versión local* del TCL al demostrar que si  $X \sim B(n, p)$ ,

$$P(X = m) \sqrt{np(1-p)} \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad (7.9)$$

para  $n$  suficientemente grande y  $x = \frac{m-np}{\sqrt{np(1-p)}}$ . Esta aproximación nos va a servir para estudiar la credibilidad de algunas aproximaciones al número  $\pi$  obtenidas a partir del problema de la *aguja de Buffon*.

Recordemos que en el problema planteado por Buffon se pretende calcular la probabilidad de que una aguja de longitud  $2l$ , lanzada al azar sobre una trama de paralelas separadas entre si una distancia  $2a$ , con  $a > l$ , corte a alguna de las paralelas. Puestos de acuerdo sobre el significado de *lanzada al azar*, la respuesta es

$$P(\text{corte}) = \frac{2l}{a\pi},$$

resultado que permite obtener una aproximación de  $\pi$  si, conocidos  $a$  y  $l$ , sustituimos en  $\pi = \frac{2l}{aP(\text{corte})}$  la probabilidad de corte por su estimador natural, *la frecuencia relativa de corte*,  $p$ , a lo largo de  $n$  lanzamientos. Podremos escribir, si en lugar de trabajar con  $\pi$  lo hacemos con su inverso,

$$\frac{1}{\pi} = \frac{am}{2ln},$$

donde  $m$  es el número de cortes en los  $n$  lanzamientos.

El año 1901 Lazzarini realizó 3408 lanzamientos obteniendo para  $\pi$  el valor 3,1415929 con 6!! cifras decimales exactas. La aproximación es tan buena que merece como mínimo alguna pequeña reflexión. Para empezar supongamos que el número de cortes aumenta en una unidad, las aproximaciones de los inversos de  $\pi$  correspondientes a los  $m$  y  $m+1$  cortes diferirían en

$$\frac{a(m+1)}{2ln} - \frac{am}{2ln} = \frac{a}{2ln} \geq \frac{1}{2n},$$

que si  $n \approx 5000$ , da lugar a  $\frac{1}{2n} \approx 10^{-4}$ . Es decir, un corte más produce una diferencia mayor que la precisión de  $10^{-6}$  alcanzada. *No queda más alternativa que reconocer que Lazzarini tuvo la suerte de obtener exactamente el número de cortes,  $m$ , que conducía a tan excelente aproximación.* La pregunta inmediata es, *¿cuál es la probabilidad de que ello ocurriera?*, y para responderla podemos recurrir a (7.9) de la siguiente forma,

$$P(X = m) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(m-np)^2}{2np(1-p)}} \leq \frac{1}{\sqrt{2\pi np(1-p)}},$$

que suponiendo  $a = 2l$  y  $p = 1/\pi$  nos da para  $P(X = m)$  la siguiente cota

$$P(X = m) \leq \sqrt{\frac{\pi}{2n(\pi - 1)}}.$$

Para el caso de Lazzarini  $n=3408$  y  $P(X = m) \leq 0,0146$ ,  $\forall m$ . *Parece ser que Lazzarini era un hombre de suerte, quizás demasiada.*



## Capítulo 8

# Teorema de Radon-Nikodym

### 8.1. Funciones aditiva y completamente aditiva

Los precedentes del concepto de medida, de los que ésta no sería más que un caso particular, son los de *funciones aditiva y completamente aditiva*.

**Definición 8.1 (Función aditiva)** Sea  $(\Omega, \mathcal{A})$  un espacio medible. Una función  $\nu$ , definida sobre  $\mathcal{A}$ , decimos que es aditiva si

1.  $\nu(\emptyset) = 0$ , y
2. dados  $A$  y  $B$  de  $\mathcal{A}$ , disjuntos,  $\nu(A \cup B) = \nu(A) + \nu(B)$ .

Si extendemos la aditividad a familias numerables disjuntas tenemos

**Definición 8.2 (Función completamente aditiva)** En el contexto de la anterior definición decimos que  $\nu$  es completamente aditiva si

1. es aditiva, y
2. dada  $\{A_n\}_{n \geq 1} \subset \mathcal{A}$ , cuyos elementos son dos a dos disjuntos,  $\nu(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} \nu(A_n)$ .

Con estas definiciones una medida  $\mu$  no es más que *una función completamente aditiva y no negativa*.

Para el resto del capítulo serán de interés las **propiedades** que enumeramos a continuación y cuya demostración omitimos por sencilla. En todas ellas  $\nu$  es una función aditiva.

1. Si  $B \subset A$  y  $|\nu(B)| < +\infty$ , entonces  $\nu(A - B) = \nu(A) - \nu(B)$ .

2. Si  $B \subset A$  y  $|\nu(B)| = +\infty$ , entonces  $\nu(A) = \nu(B)$ .
3. Si  $B \subset A$  y  $|\nu(A)| < +\infty$ , entonces  $|\nu(B)| < +\infty$ .
4. Dados  $A$  y  $B$ , si  $\nu(A) = +\infty$ , entonces  $\nu(B) \neq -\infty$ , y si  $\nu(A) = -\infty$ , entonces  $\nu(B) \neq +\infty$ .

## 8.2. Descomposiciones de Jordan y Hahn

**Definición 8.3** Sea  $(\Omega, \mathcal{A})$  un espacio medible y  $\nu$  una función completamente aditiva definida de  $\mathcal{A}$  en  $\mathcal{R}^*$ . Llamaremos variación superior de  $\nu$  en  $A \in \mathcal{A}$  a

$$\bar{\nu}_\nu(A) = \sup\{\nu(B), B \in \mathcal{A}, B \subset A\}.$$

De esta definición se deriva la siguiente propiedad,

**Teorema 8.1**  $\bar{\nu}_\nu$  es una medida sobre  $\mathcal{A}$ .

**Demostración.-** Comprobemos que verifica las tres condiciones que definen a una medida.

1.  $\bar{\nu}_\nu(\emptyset) = 0$ , porque  $\nu(\emptyset) = 0$ .
2.  $\bar{\nu}_\nu(A) \geq 0$ ,  $\forall A \in \mathcal{A}$ , puesto que  $\emptyset \subset A$  y  $\nu(\emptyset) = 0$ .
3. Para  $\{A_1, A_2, \dots\}$  disjuntos dos a dos,  $\bar{\nu}_\nu(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} \bar{\nu}_\nu(A_n)$ .

En efecto, dado  $A \in \mathcal{A}$  y  $A \subset \bigcup_{n \geq 1} A_n$ , la aditividad numerable de  $\nu$  y la monotonía de  $\bar{\nu}_\nu$ , de fácil comprobación, nos permiten escribir

$$\nu(A) \leq \sum_{n \geq 1} \bar{\nu}_\nu(A \cap A_n) \leq \sum_{n \geq 1} \bar{\nu}_\nu(A_n),$$

desigualdad que, al ser cierta para cualquier subconjunto de  $\bigcup_{n \geq 1} A_n$ , conduce, al tomar supremos, a

$$\bar{\nu}_\nu\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \bar{\nu}_\nu(A_n).$$

Para obtener la desigualdad contraria supondremos que  $\bar{\nu}_\nu(A_n) < +\infty$ ,  $\forall n$  (en caso contrario la igualdad buscada es trivialmente cierta). Dado ahora  $\epsilon > 0$ , para cada  $A_n$ ,  $\exists E_n$  verificando  $\nu(E_n) \geq \bar{\nu}_\nu(A_n) - \epsilon/2^n$ . Al sumar para todo  $n$ , como  $\bigcup_{n \geq 1} E_n \subset \bigcup_{n \geq 1} A_n$ , obtendremos  $\bar{\nu}_\nu(\bigcup_{n \geq 1} A_n) \geq \nu(\bigcup_{n \geq 1} E_n) \geq \sum_{n \geq 1} \bar{\nu}_\nu(A_n) - \epsilon$ .  $\square$



**Definición 8.4** Llamaremos variación inferior de  $\nu$  en  $A \in \mathcal{A}$  a

$$\underline{\nu}(A) = \inf\{\nu(B), B \in \mathcal{A}, B \subset A\}.$$

Observemos que  $-\underline{\nu}(A) = \bar{\nu}_{-\nu}(A)$  y siendo  $-\nu$  completamente aditiva,  $-\underline{\nu}$  será también una medida sobre  $\mathcal{A}$ .

**Proposición 8.1** Si  $\bar{\nu}(A) = +\infty$ , entonces  $\nu(A) = +\infty$ .

**Demostración.-** Si  $\bar{\nu}(A) = +\infty$ , dado  $n$ ,  $\exists A_n \subset A$  verificando  $\nu(A_n) \geq n$ . Para  $n = 1$  consideremos  $A_1$  y  $A - A_1$  y siendo  $\bar{\nu}$  una medida, necesariamente tendremos  $\bar{\nu}(A_1) = +\infty$  o  $\bar{\nu}(A - A_1) = +\infty$ . Designemos por  $E_1$  aquel que lo verifique y repitamos el razonamiento con  $E_1$  en lugar del  $A$  inicial. Una iteración de este proceso dará lugar a las sucesiones  $\{E_n, n \geq 1\}$  y  $\{A_n, n \geq 1\}$ . Si  $\exists n_0$ , tal que  $\nu(E_{n_0}) = +\infty$ , la proposición está demostrada. Supondremos que  $|\nu(E_n)| < +\infty, \forall n$  y consideraremos las dos alternativas siguientes:

1.  $\exists n_0$  tal que  $E_n = A_n, \forall n \geq n_0$ .

Como los  $E_n$  son decrecientes,

$$\nu\left(\bigcap_{n \geq n_0} E_n\right) = \lim_{n \xrightarrow{n \geq n_0} \infty} \nu(E_n) = \lim_{n \xrightarrow{n \geq n_0} \infty} \nu(A_n) = +\infty,$$

y como  $\bigcap_{n \geq n_0} E_n \subset A$ ,  $\nu(A) = +\infty$ .

2. Solo eventualmente  $E_n = A_n$ .

Eliminamos los  $A_n$  que coinciden con algún  $E_n$  obteniendo así la subsucesión de conjuntos disjuntos,  $\{A_{n_i}, i \geq 1\}$ , para la que se verifica,

$$\nu\left(\bigcup_{i \geq 1} A_{n_i}\right) = \sum_{i \geq 1} \nu(A_{n_i}) \geq \sum_{i \geq 1} n_i = +\infty,$$

y como también  $\bigcup_{i \geq 1} A_{n_i} \subset A$ ,  $\nu(A) = +\infty$ .  $\square$

La variación inferior goza de una propiedad semejante.

**Corolario 8.1** Si  $\underline{\nu}(A) = -\infty$ , entonces  $\nu(A) = -\infty$ .

Estamos ahora en condiciones de demostrar las descomposiciones que dan título a la sección.

**Teorema 8.2 (Descomposición de Jordan)** Sea  $(\Omega, \mathcal{A})$  un espacio medible y  $\nu$  una función completamente aditiva definida de  $\mathcal{A}$  en  $\mathcal{R}^*$ . Existen entonces dos medidas  $\mu_1$  y  $\mu_2$  definidas sobre  $\mathcal{A}$  tales que,

$$\nu(A) = \mu_1(A) - \mu_2(A), \quad \forall A \in \mathcal{A}.$$

**Demostración.-** Consideremos las dos situaciones:

1.  $|\nu(A)| = +\infty$ .

Si  $\nu(A) = +\infty$ , la variación superior también lo será y por el Corolario anterior,  $\underline{\nu}(A) \neq -\infty$ . De análoga manera razonaríamos para el caso  $\nu(A) = -\infty$ , por lo que en ambas situaciones podemos escribir

$$\nu(A) = \bar{\nu}(A) + \underline{\nu}(A).$$

2.  $|\nu(A)| < +\infty$ .

Para cualquier  $E \subset A$ ,  $\nu(E) = \nu(A) - \nu(A - E)$ . Como  $\nu(A - E) \leq \bar{\nu}(A)$  y  $\nu(A - E) \geq \underline{\nu}(A)$ , al sustituir obtenemos las desigualdades  $\nu(A) - \bar{\nu}(A) \leq \nu(E) \leq \nu(A) - \underline{\nu}(A)$ , que son ciertas  $\forall E \subset A$ . Tomando ínfimos y supremos podemos de nuevo escribir

$$\nu(A) = \bar{\nu}(A) + \underline{\nu}(A).$$

La descomposición es  $\mu_1 = \bar{\nu}$  y  $\mu_2 = -\underline{\nu}$ .  $\square$

**Teorema 8.3 (Descomposición de Hahn)** Sea  $(\Omega, \mathcal{A})$  un espacio medible y  $\nu$  una función completamente aditiva y finita definida sobre  $\mathcal{A}$ . Existe  $E \in \mathcal{A}$  tal que

$$\nu(A) \geq 0, \quad \forall A \in \mathcal{A}, \quad A \subset E,$$

y

$$\nu(B) \leq 0, \quad \forall B \in \mathcal{A}, \quad B \subset E^c.$$

**Demostración.-** La finitud de  $\nu$  implica la de  $\bar{\nu}$  y para un  $\epsilon = 1/2^n$ ,  $\exists E_n \in \mathcal{A}$  tal que  $\nu(E_n) > \bar{\nu}(\Omega) - 1/2^n$ . De aquí,

$$\bar{\nu}(E_n^c) = \bar{\nu}(\Omega) - \bar{\nu}(E_n) \leq \bar{\nu}(\Omega) - \nu(E_n) < 1/2^n$$

y aplicando que  $\nu = \bar{\nu} + \underline{\nu}$

$$-\underline{v}_\nu(E_n) = \bar{v}_\nu(E_n) - \nu(E_n) \leq \bar{v}_\nu(\Omega) - \nu(E_n) < 1/2^n.$$

El conjunto  $E = \bigcap_{k \geq 1} \bigcup_{n \geq k} E_n$  será el buscado si comprobamos que  $\bar{v}_\nu(E^c) = 0$  y  $\underline{v}_\nu(E) = 0$ . Veámoslo,

$$\bar{v}_\nu(E^c) = \lim_{k \rightarrow \infty} \bar{v}_\nu\left(\bigcap_{n \geq k} E_n^c\right) \leq \limsup_{k \geq 1} \bar{v}_\nu(E_k^c) = 0,$$

y

$$-\underline{v}_\nu(E) \leq -\sum_{n \geq k} \underline{v}_\nu(E_n) < \frac{1}{2^{k-1}}, \quad \forall k,$$

de donde  $\underline{v}_\nu(E) = 0$ .  $\square$

**Observación 8.1** *Las demostraciones de los dos teoremas precedentes ponen de manifiesto que ninguna de las dos descomposiciones es única.*

### 8.3. Continuidad absoluta y singularidad

**Definición 8.5 (Continuidad absoluta)** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $\nu$  una función de  $\mathcal{A}$  en  $\mathcal{R}^*$ . Se dice que  $\nu$  es absolutamente continua respecto de  $\mu$  si  $\nu(A) = 0$ ,  $\forall A \in \mathcal{A}$  tal que  $\mu(A) = 0$ .

El concepto de continuidad absoluta nos es mucho más familiar de lo que este primer encuentro pueda hacernos pensar. En efecto, si recurrimos a la noción de integral y a una de sus propiedades, a saber, *que la integral sobre conjuntos de medida nula es nula*, construiremos con facilidad funciones absolutamente continuas respecto a cualquier medida. Para ello consideremos una función medible,  $f$ , y definamos sobre  $\mathcal{A}$  una función  $\nu$  mediante,

$$\nu(A) = \int_A f d\mu.$$

Es inmediato comprobar, apoyándonos en la propiedad citada, la continuidad absoluta de  $\nu$  respecto de  $\mu$ .

En el caso de funciones puntuales hemos hablado también, capítulo 2, de continuidad absoluta. La equivalencia en la recta de real de ambos conceptos, de la que nos ocuparemos en la última sección del capítulo, explica esta coincidencia de nombre.

La continuidad absoluta puede también caracterizarse, bajo ciertos supuestos, mediante una condición de las llamadas  $\epsilon - \delta$ . El siguiente teorema recoge dicha caracterización.

**Teorema 8.4** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $\nu$  una función sobre  $\mathcal{A}$  completamente aditiva y finita. Entonces,  $\nu$  es absolutamente continua respecto de  $\mu$  si y sólo si

$$\text{dado } \epsilon > 0, \exists \delta > 0, \text{ tal que } |\nu(A)| < \epsilon \text{ si } \mu(A) < \delta. \quad (8.1)$$

**Demostración.-** Veamos las dos implicaciones.

**La condición (8.1)  $\implies$  la continuidad absoluta.-** Si se verifica (8.1) y  $A$  es tal que  $\mu(A) = 0$ , entonces  $\mu(A) < \delta$  cualquiera que sea el  $\delta$  elegido y por tanto  $|\nu(A)| < \epsilon$ ,  $\forall \epsilon > 0$ . En consecuencia  $\nu(A) = 0$  y será absolutamente continua respecto de  $\mu$ .

**La continuidad absoluta  $\implies$  la condición (8.1).-** Supongamos ahora la continuidad absoluta de  $\nu$  respecto de  $\mu$ . Sea  $v_\nu = \bar{v}_\nu - \underline{v}_\nu$  la *variación total* de  $\nu$ , que también es absolutamente continua respecto de  $\mu$ . Si no lo fuera podríamos encontrar un  $A$  tal que  $\mu(A) = 0$  y  $v_\nu(A) > 0$  y dada su definición  $\bar{v}_\nu(A)$ , o bien  $\underline{v}_\nu(A)$ , serían positivas. Supongamos por ejemplo que  $\bar{v}_\nu(A) > 0$ , entonces  $\exists B \subset A$  con  $\nu(B) > 0$ , pero  $\mu(B) = 0$  y por continuidad absoluta,  $\nu(B) = 0$ , lo que contradice lo anterior.

Por otra parte como  $|\nu(A)| \leq v_\nu(A)$ ,  $\forall A \in \mathcal{A}$ , bastará con demostrar que  $v_\nu$  verifica (8.1). De no hacerlo existiría  $\epsilon > 0$  tal que, para  $\delta = 1/2^n$ , encontraríamos un  $E_n$  con  $\mu(E_n) < 1/2^n$  y  $v_\nu(E_n) \geq \epsilon$ . Definamos  $E = \bigcap_{k \geq 1} \bigcup_{n \geq k} E_n$ , entonces

$$\mu(E) \leq \mu\left(\bigcup_{n \geq k} E_n\right) < 1/2^{k-1}, \quad \forall k.$$

Es decir, que  $\mu(E) = 0$ . Pero

$$v_\nu\left(\bigcap_{k \geq 1} \bigcup_{n \geq k} E_n\right) = \lim_{k \rightarrow \infty} v_\nu\left(\bigcup_{n \geq k} E_n\right) \geq \liminf_{k \rightarrow \infty} v_\nu(E_n) \geq \epsilon > 0,$$

lo que es absurdo y por tanto  $v_\nu$  es absolutamente continua respecto de  $\mu$ .  $\square$

Otra relación importante entre una medida y una función completamente aditiva la define el concepto de singularidad.

**Definición 8.6 (Singularidad)** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $\nu$  una función completamente aditiva definida sobre  $\mathcal{A}$ . Se dice que  $\nu$  es singular respecto de  $\mu$  si

$$\exists A \in \mathcal{A}, \text{ tal que } \mu(A) = 0 \text{ y } \nu(B \cap A^c) = 0, \quad \forall B \in \mathcal{A}.$$

Lo que la definición viene a afirmar es que  $\nu$  es distinta de cero solamente sobre un conjunto que tiene medida nula respecto  $\mu$ .

Como fácilmente se comprueba partir de las definiciones de continuidad absoluta y singularidad, si  $\nu$  es a la vez absolutamente continua y singular respecto de la misma medida  $\mu$ , entonces es idénticamente nula sobre  $\mathcal{A}$ .

## 8.4. Descomposición de Lebesgue. Teorema de Radon-Nikodym

Esta sección recoge el resultado fundamental del capítulo, la llamada *descomposición de Lebesgue* que permite, bajo condiciones no demasiado exigentes, escribir cualquier función completamente aditiva definida en un espacio de medida como suma de dos funciones, una absolutamente continua y otra singular respecto de la medida del espacio.

**Teorema 8.5 (Descomposición de Lebesgue)** *Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $\nu$  una función completamente aditiva definida sobre  $\mathcal{A}$ , con  $\mu$  y  $\nu$   $\sigma$ -finitas. Existen  $\nu_c$  y  $\nu_s$ , completamente aditivas y únicas,  $\nu_c$  absolutamente continua y  $\nu_s$  singular respecto de  $\mu$  y tales que  $\nu = \nu_c + \nu_s$ . Existe además  $g$ , medible y finita, tal que*

$$\nu_c(E) = \int_E g d\mu, \quad \forall E \in \mathcal{A}.$$

**Demostración.-** La  $\sigma$ -finitud de  $\mu$  y  $\nu$  nos permite expresar  $\Omega$  como unión numerable disjunta de conjuntos sobre los que ambas son finitas. Podemos por tanto suponer, sin pérdida de generalidad, que ambas son finitas sobre  $\Omega$ . Como por el teorema de la descomposición de Jordan  $\nu$  puede expresarse como diferencia de dos medidas, estableceremos el teorema suponiendo que  $\nu$  es también no negativa.

Sea  $\mathcal{F} = \{f \geq 0, \text{ medible y tal que } \int_E f d\mu \leq \nu(E), \forall E \in \mathcal{A}\}$ , familia que no es vacía pues al menos contiene la función  $f \equiv 0$ . Sea  $L = \sup\{\int_\Omega f d\mu, f \in \mathcal{F}\}$ , entonces  $L \leq \nu(\Omega) < +\infty$  y podremos encontrar una sucesión  $\{f_n, n \geq 1\} \subset \mathcal{F}$  verificando  $\lim_n \int_\Omega f_n d\mu = L$ , a partir de la cual definimos una nueva sucesión,  $g_n = \max_{j \leq n} f_j$ ,  $n \geq 1$ , que será monótona creciente con  $g = \lim_n g_n$ . Como  $f_n \leq g_n$ ,

$$\int_\Omega f_n d\mu \leq \int_\Omega g_n d\mu \leq \int_\Omega g d\mu, \quad \forall n,$$

y de aquí

$$L \leq \int_\Omega g d\mu. \quad (8.2)$$

Pero  $g$  es un elemento de  $\mathcal{F}$ . En efecto, dado  $n$  y  $k \leq n$ , definimos  $A_k = \{\omega; f_k(\omega) = g_n(\omega)\}$  y  $B_k = A_k - \bigcup_{i=1}^{k-1} A_i$ , constituyendo estos últimos una partición medible finita de  $\Omega$  que nos permite obtener

$$\int_E g_n d\mu = \sum_{k=1}^n \int_{E \cap B_k} g_n d\mu = \sum_{k=1}^n \int_{E \cap B_k} f_n d\mu \leq \sum_{k=1}^n \nu(E \cap B_k) = \nu(E),$$

y por el teorema de la convergencia monótona de Lebesgue  $\int_E g d\mu \leq \nu(E)$  y entonces  $\int_\Omega g d\mu \leq L$ , que junto con (8.2) supone la igualdad. Esta  $g$ , que podemos suponer finita por serlo casi por todas partes, cumple el enunciado del teorema y nos permite definir para cualquier  $E \in \mathcal{A}$

$$\nu_c(E) = \int_E g d\mu,$$

que es absolutamente continua respecto de  $\mu$ .

Para completar la descomposición del enunciado definamos

$$\nu_s = \nu - \nu_c,$$

que será completamente aditiva y no negativa porque  $g \in \mathcal{F}$ . La comprobación de su singularidad respecto de  $\mu$  demostrará el teorema. Sea, para cada  $n$ ,  $\nu_n = \nu_s - \frac{1}{n}\mu$  y  $D_n$  su descomposición de Hahn por lo que  $\nu_n(E \cap D_n) \leq 0$  y  $\nu_n(E \cap D_n^c) \geq 0$ ,  $\forall E \in \mathcal{A}$ . Si  $D = \bigcap_n D_n$ , como  $\nu_n(E \cap D) \leq 0$  se deduce que  $\nu_s(E \cap D) \leq \frac{1}{n}\mu(E \cap D)$ ,  $\forall n$  y de aquí  $\nu_s(E \cap D) = 0$ ,  $\forall E \in \mathcal{A}$ .

Si  $\mu(D^c) = 0$ , la singularidad de  $\nu_s$  estará comprobada, pero como  $D^c = \bigcup_n D_n^c$  verificaremos solamente que  $\mu(D_n^c) = 0$ ,  $\forall n$  y para ello,

$$\int_E (g + \frac{1}{n} \mathbf{1}_{D_n^c}) d\mu \leq \nu_c(E) + \nu_s(E \cap D_n^c) \leq \nu_c(E) + \nu_s(E) = \nu(E),$$

lo que supone que  $(g + \frac{1}{n} \mathbf{1}_{D_n^c}) \in \mathcal{F}$ ,  $\forall n$  y  $L \geq \int_\Omega (g + \frac{1}{n} \mathbf{1}_{D_n^c}) d\mu$ , que conduce a  $\mu(D_n^c) = 0$ ,  $\forall n$ .

La unicidad de la descomposición es sencilla de comprobar.  $\square$

**Observación 8.2** *La función  $g$  obtenida en el teorema no es única, pero cualquier otra función con estas características será igual casi por todas partes a ella.*

*Obsérvese que de la demostración no se desprende la integrabilidad de  $g$  que solo ocurrirá cuando  $\nu_c$  sea finita. Lo que siempre está garantizada es la existencia de su integral.*

Una situación particularmente interesante se produce cuando  $\nu$  es absolutamente continua respecto de  $\mu$ . El correspondiente corolario tiene nombre propio.

**Corolario 8.2 (Teorema de Radon-Nikodym)** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida y  $\nu$  una función completamente aditiva definida sobre  $\mathcal{A}$ , con  $\mu$  y  $\nu$   $\sigma$ -finitas. Si  $\nu$  es absolutamente continua respecto de  $\mu$ , existe entonces una función  $g$ , medible y finita, tal que

$$\nu(E) = \int_E g d\mu, \quad \forall E \in \mathcal{A}.$$

A  $g$  se la denomina *derivada de Radon-Nikodym de  $\nu$  respecto de  $\mu$* .

El Teorema de Lebesgue y su corolario, el Teorema de Radon-Nikodym, nos permiten caracterizar aquellas funciones de conjunto que vienen definidas mediante integración. El siguiente teorema recoge dicha caracterización.

**Teorema 8.6** Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida con  $\mu$   $\sigma$ -finita y  $\nu$  una función definida sobre  $\mathcal{A}$ . Existirá una función  $g$ , medible y finita, tal que

$$\nu(E) = \int_E g d\mu, \quad \forall E \in \mathcal{A},$$

sí y sólo sí  $\nu$  es completamente aditiva,  $\sigma$ -finita y absolutamente continua respecto de  $\mu$ .

**Demostración.-** Si  $\nu$  verifica las condiciones del enunciado, Radon-Nikodym asegura la existencia de semejante  $g$ . Para ver la implicación contraria bastará probar que la función que (8.6) define es  $\sigma$ -finita, porque la aditividad completa ya ha sido probada en el capítulo 4 y la continuidad absoluta respecto de  $\mu$  es inmediata.

Como  $\mu$  es  $\sigma$ -finita, existe una sucesión  $\{A_n\}_{n \geq 1} \subset \mathcal{A}$  tal que  $\Omega = \bigcup_{n \geq 1} A_n$  y  $\mu(A_n) < +\infty$ ,  $\forall n$ . Para cada entero  $m > 0$  definimos  $B_m = \{\omega; |g(\omega)| \leq m\}$  y como  $g$  es finita,  $\Omega = \bigcup_{m \geq 1} B_m$ . La intersección de ambas sucesiones dará lugar a una nueva sucesión que recubre también a  $\Omega$  y para cuyos elementos

$$\nu(A_n \cap B_m) = \int_{A_n \cap B_m} g d\mu \leq \int_{A_n} m d\mu = m\mu(A_n) < +\infty, \quad \forall n \text{ y } \forall m. \square$$

## 8.5. Aplicación a $(\mathcal{R}, \mathcal{B}, \lambda)$

En esta sección nos vamos a ocupar de un caso particular de gran trascendencia en Teoría de la Probabilidad. Nuestro espacio medible será ahora  $(\mathcal{R}, \mathcal{B})$  y sobre él consideramos dos medidas, la de Lebesgue,  $\lambda$ , y una medida de Lebesgue-Stieltjes,  $\mu_F$ , obtenida a partir de una función de distribución  $F$  acotada y absolutamente continua. Recordemos la definición de continuidad absoluta para las funciones reales.

**Definición 8.7 (Continuidad absoluta)** Decimos que  $F$  es absolutamente continua si dado  $\epsilon > 0$ , existe  $\delta > 0$ , tal que para cada familia finita disjunta de intervalos  $\{I_i = ]a_i, b_i], i = 1, \dots, k\}$ ,

$$\sum_{i=1}^k (F(b_i) - F(a_i)) < \epsilon \text{ si } \sum_{i=1}^k (b_i - a_i) < \delta. \quad (8.3)$$

Si escribimos  $I = \bigcup_{i=1}^k I_i$ , (8.3) puede también escribirse de la forma

$$\mu_F(I) = \sum_{i=1}^k \mu_F(I_i) < \epsilon \text{ si } \lambda(I) = \sum_{i=1}^k \lambda(I_i) < \delta. \quad (8.4)$$

Esta nueva forma de escribir (8.3) nos permite establecer la equivalencia entre ambas formas de continuidad absoluta.

**Teorema 8.7** Si  $F$  es una función de distribución acotada y  $\mu_F$  la medida de Lebesgue-Stieltjes asociada,  $F$  es absolutamente continua si y sólo si  $\mu_F$  es absolutamente continua respecto de la medida de Lebesgue,  $\lambda$ .

**Demostración.-** Veamos las dos implicaciones:

1. La continuidad absoluta de  $\mu_F$  y el Teorema (8.4) hacen que se verifique (8.4), lo que supone la continuidad absoluta de  $F$ .
2. Sea ahora  $F$  absolutamente continua. Para  $A \in \mathcal{B}$ , con  $\lambda(A) = 0$ , por las condiciones de regularidad de  $\lambda$ , dado  $\delta > 0$ , existe  $J$  tal que  $A \subset J$ ,  $J = \bigcup_{i \geq 1} I_i$ , con los  $I_i = ]a_i, b_i]$ , disjuntos y tal que  $\lambda(J - A) \leq \delta/2$ . Entonces  $\lambda(J) = \sum_{i \geq 1} \lambda(I_i) \leq \delta/2$ , desigualdad válida para cualquier suma finita que, por la continuidad absoluta de  $F$ , nos permite escribir  $\sum_{i=1}^n \mu_F(I_i) \leq \epsilon$ ,  $\forall n$ . En consecuencia  $\mu_F(A) \leq \mu_F(J) \leq \epsilon$ ,  $\forall \epsilon$ , y de aquí  $\mu_F(A) = 0$ .  $\square$

Estos resultados adquieren para nosotros su mayor relevancia cuando los utilizamos en un contexto probabilístico, concretamente en el espacio  $(\mathcal{R}, \mathcal{B}, \mathcal{P})$ . Vimos en capítulo 2 que una medida de probabilidad no es más que una medida de Lebesgue-Stieltjes definida a partir de  $F$ , su función de distribución asociada. La combinación con el teorema de Radon-Nikodym de estos resultados nos permite caracterizar aquellas medidas de probabilidad sobre  $\mathcal{R}$  que poseen densidad. La caracterización la damos en términos de la función de distribución asociada.

**Teorema 8.8** Sea  $F$  una función de distribución de probabilidad,  $F$  será absolutamente continua si y sólo si su  $\mathcal{P}$  asociada posee densidad respecto de la medida de Lebesgue,  $\lambda$ .



La demostración es inmediata utilizando el teorema de Radon-Nidodym y el teorema anterior.

**Observación 8.3** *Este resultado es también válido si sustituimos  $\mathcal{P}$  por una medida de Lebesgue-Stieltjes cuya  $F$  sea acotada.*



# Bibliografía

- [1] L. Breiman. *Probability*. Addison-Wesley, Reading, MA, 1968.
- [2] P. Billingsley. *Convergence of Probability Measures*. John Wiley, New York, 1968.
- [3] P. Billingsley. *Probability and Measure*. John Wiley, New York, 1986.
- [4] K. Chung. *A Course in Probability Theory* 2nd ed. Academic Press, New York, 1974.
- [5] W. Feller. *An Introduction to Probability Theory and Its Applications*. Vol. 2. John Wiley, New York, 1966.
- [6] W. Feller. *An Introduction to Probability Theory and Its Applications*. Vol. 1. 3rd. ed. John Wiley, New York, 1968.
- [7] B.V. Gnedenko. *The Theory of Probability*. Chelsea, New York, 1962.
- [8] P. Halmos. *Measure Theory*. Van Nostrand, Princeton, NJ, 1950.
- [9] O. Kallenberg. *Foundations of Modern Probability*. Springer Verlag, New York, 1997.
- [10] M. Loève. *Probability Theory*. Springer Verlag, New York, 1977-78.
- [11] A. Rényi. *Probability Theory*. North-Holland, Amsterdam, 1970.
- [12] G.R. Shorack. *Probability for Statisticians*. Springer Verlag, New York, 2000.