

Procesos estocásticos

Ernesto Mordecki

22 de junio de 2023

Índice general

1. Medida e integración para procesos estocásticos	3
1.1. Medida y probabilidad: Motivación	3
1.1.1. Ejemplo	3
1.1.2. Ejemplo	3
1.1.3. Ejemplo	4
1.2. Probabilidades finitamente aditivas	4
1.3. Álgebras y σ -álgebras: propiedades y ejemplos	10
1.4. Medidas en \mathbf{R}	10
1.5. Ejemplos de medidas de probabilidad en \mathbf{R}	14
1.5.1. Medidas discretas en \mathbf{R}	16
1.5.2. Medidas absolutamente continuas en \mathbf{R}	17
1.5.3. Medidas singulares	17
1.6. Medidas en \mathbf{R}^n	19
1.6.1. Ejemplos	22
1.6.2. Medidas absolutamente continuas en \mathbf{R}^n	23
1.7. Medidas de probabilidad en \mathbf{R}^∞	24
1.8. Variables aleatorias	28
1.8.1. Variables aleatorias y límites	30
1.9. Vectores aleatorios	32
1.10. Esperanza matemática. Integral de Lebesgue	34
1.11. Tres teoremas de convergencia	38
1.12. Cambio de variable en la integral de Lebesgue	39
2. Leyes de los grandes números	43
2.1. Ley débil de los grandes números	43
2.2. Ejercicios de Leyes débiles de los grandes números	53
2.3. Ley fuerte de los grandes números	54
2.4. Ejercicios de Ley fuerte de los grandes números	54

3. Funciones características	55
3.1. Definiciones y primeras propiedades	55
3.2. Fórmula de inversión. Teorema de unicidad	61
3.3. Teoremas de Helly	64
3.4. Relación entre la convergencia de distribuciones y de funciones características	68
3.5. Ejercicios	70
4. Teorema central del límite	73
4.1. Teorema de Lindeberg–Lévy	73
4.2. Teorema de Lindeberg	76
4.3. Teorema de Lyapunov	82
4.4. Ejercicios	83
5. Martingalas	87
5.1. Esperanza condicional	88
5.2. Propiedades de la esperanza condicional	92
5.3. Martingalas	96
5.4. Teorema del muestreo opcional	101
5.5. Convergencia casi segura de Martingalas	106
5.6. Integrabilidad uniforme y convergencia en media de martingalas	109
6. El Proceso de Wiener	113
6.1. Vectores gaussianos	113
6.2. Definición del Proceso de Wiener	115
6.3. Construcción del proceso de Wiener	116
6.4. Propiedades de las trayectorias del Proceso de Wiener	120
6.5. La distribución del máximo del proceso de Wiener	124
7. Parada óptima para el proceso de Wiener	128
7.1. El problema de parada óptima: Ejemplo Guía	129

Capítulo 1

Medida e integración para procesos estocásticos

1.1. Medida y probabilidad: Motivación

Queremos extender los modelos de probabilidad finitos.

1.1.1. Ejemplo

Sea un conjunto finito

$$\{\omega_1, \dots, \omega_n\},$$

Si tenemos números positivos p_1, \dots, p_n tales que $p_1 + \dots + p_n = 1$, podemos definir para un subconjunto \mathbf{A} de Ω

$$\mathbf{P}(\mathbf{A}) = \sum_{k: \omega_k \in \mathbf{A}} p_k. \quad (1.1)$$

Esto nos da un modelo de probabilidad finito. La propiedad más importante que se verifica es que

$$\mathbf{P}(\mathbf{A} + \mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}), \quad (1.2)$$

donde escribimos $\mathbf{A} + \mathbf{B}$ para la unión de los conjuntos *disjuntos* \mathbf{A} y \mathbf{B} .

1.1.2. Ejemplo

Un caso particular del ejemplo anterior consiste en poner una probabilidad en el conjunto de tiras de ceros y unos. Más específicamente

$$\Omega = \{\omega = (a_1, \dots, a_n): a_k = 0 \text{ ó } 1, k = 1, \dots, n\}.$$

Tenemos que el cardinal de Ω es $|\Omega| = 2^n$. Si $0 < p < 1$ asignamos probabilidades

$$P(\omega) = p^\mu(1-p)^{n-\mu}$$

donde $\mu - \mu(\omega)$ indica la cantidad de unos de la tira ω . Observemos que

$$\sum_{\omega \in \Omega} \mathbf{P}(\omega) = \sum_{k=0}^n \mathbf{P}(\mu(\omega) = k) = \sum_{k=0}^n C_k^n p^k (1-p)^{n-k} = 1,$$

donde la última igualdad es el binomio de Newton.

1.1.3. Ejemplo

Consideremos ahora el conjunto de las sucesiones de ceros y unos (es decir, las “tiras infinitas” en el ejemplo anterior).

$$\Omega = \{\omega = (a_1, a_2, \dots): a_n = 0 \text{ ó } 1, n = 1, 2, \dots\}.$$

En este caso Ω es un conjunto infinito, así que no podemos utilizar la definición (1.1). Si quisieramos hacerlo asignando la misma probabilidad a cada sucesión, ésta debería ser cero, pero las probabilidades de conjuntos serían nulas. Observemos que en este caso tenemos una biyección con $[0, 1]$, dada por la fórmula

$$x = \frac{a_1}{2} + \frac{a_2}{2^2} + \dots, \quad x \in [0, 1].$$

Podríamos pensar entonces que asignar una probabilidad en el conjunto de sucesiones es equivalente a asignar una probabilidad en el conjunto $[0, 1]$. Aquí ocurre lo mismo, no es posible pensar en asignarle probabilidades a los puntos, la fórmula (1.1) se extiende a lo sumo a una cantidad numerable de sumandos, lo que no constituye todos los conjuntos de interés. La estrategia entonces es asignar probabilidad a subconjuntos de Ω y no a puntos. Precisamos entonces dotar a Ω de una estructura de subconjuntos adecuada para las operaciones necesarias, como por ejemplo que se verifique la propiedad (1.2).

1.2. Probabilidades finitamente aditivas

Definición 1. Una familia \mathcal{F} de subconjuntos de Ω es un álgebra cuando se verifica

- (a) $\Omega \in \mathcal{F}$

(b) Si $\mathbf{A} \in \mathcal{F}$, $\mathbf{B} \in \mathcal{F}$ tenemos $\mathbf{A} \cup \mathbf{B} \in \mathcal{F}$.

(c) Si $\mathbf{A} \in \mathcal{F}$ entonces $\overline{\mathbf{A}} \in \mathcal{F}$.

Definición 2. Dada un álgebra \mathcal{F} de subconjuntos de Ω , una medida finitamente aditiva es una función $\mu: \mathcal{F} \rightarrow [0, \infty]$ que verifica

$$\mu(\mathbf{A} + \mathbf{B}) = \mu(\mathbf{A}) + \mu(\mathbf{B}), \quad (1.3)$$

para todo par de conjuntos \mathbf{A} , \mathbf{B} disjuntos en el álgebra. Si $\mu(\Omega) < \infty$ decimos que la medida finitamente aditiva es finita, si $\mu(\Omega) = 1$, decimos que es una probabilidad finitamente aditiva.

Definición 3. Una terna $(\Omega, \mathcal{F}, \mathbf{P})$ es un modelo de probabilidad extendido cuando

(a) Ω es un conjunto de puntos,

(b) \mathcal{F} es un álgebra,

(c) \mathbf{P} es una probabilidad finitamente aditiva.

Es necesario considerar uniones más generales que las de la propiedad (6.15), en particular uniones numerables de conjuntos. Eso motiva la siguiente definición.

Definición 4. Una familia \mathcal{F} de subconjuntos de Ω es una σ -álgebra (se lee sigma álgebra cuando es un álgebra, que además verifica

(b*) Si $\mathbf{A}_n \in \mathcal{F}$ para todo $n = 1, 2, \dots$, entonces

$$\bigcup_{n=1}^{\infty} \mathbf{A}_n \in \mathcal{F}.$$

Tenemos la siguiente definición equivalente.

Definición 5. Una familia \mathcal{F} de subconjuntos de Ω es una σ -álgebra cuando:

(a) $\Omega \in \mathcal{F}$,

(b) Si $\mathbf{A}_n \in \mathcal{F}$ para todo $n = 1, 2, \dots$, entonces

$$\bigcup_{n=1}^{\infty} \mathbf{A}_n \in \mathcal{F}.$$

(c) Si $\mathbf{A} \in \mathcal{F}$ entonces $\overline{\mathbf{A}} \in \mathcal{F}$.

Ejercicio 1. (\star) Demostrar las dos definiciones de σ -álgebra son equivalentes.

Definición 6. El espacio Ω con la σ -álgebra \mathcal{F} se denota espacio medible (Ω, \mathcal{F}) .

Definición 7. (a) Una medida μ definida en un álgebra \mathcal{F} se denomina σ -aditiva, o simplemente una medida, cuando para toda sucesión de conjuntos $\mathbf{A}_1, \mathbf{A}_2, \dots$ disjuntos dos a dos de \mathcal{F} , y tales que $\sum \mathbf{A}_n \in \mathcal{F}$, se verifica

$$\mu \left(\sum_{n=1}^{\infty} \mathbf{A}_n \right) = \sum_{n=1}^{\infty} \mu(\mathbf{A}_n).$$

(b) Una medida σ -aditiva en un álgebra que verifica $\mathbf{P}(\Omega) = 1$ se llama una medida de probabilidad, o sencillamente, una probabilidad.

Propiedad 1. Para cualquier suceso \mathbf{A} se tiene $\mathbf{P}(\overline{\mathbf{A}}) = 1 - \mathbf{P}(\mathbf{A})$.

Demostración. Por definición de $\overline{\mathbf{A}}$ (suceso contrario al suceso \mathbf{A}), tenemos $\overline{\mathbf{A}} = \Omega \setminus \mathbf{A}$. De aquí resulta $\mathbf{A}\overline{\mathbf{A}} = \emptyset$. Como por el axioma II tenemos $\mathbf{P}(\mathbf{A} \cup \overline{\mathbf{A}}) = \mathbf{P}(\Omega) = 1$, aplicando el axioma III concluimos, que $1 = \mathbf{P}(\Omega) = \mathbf{P}(\mathbf{A} \cup \overline{\mathbf{A}}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\overline{\mathbf{A}})$. \square

Propiedad 2. El suceso imposible tiene probabilidad nula: $\mathbf{P}(\emptyset) = 0$.

Demostración. Esta igualdad se obtiene de la propiedad anterior, si consideramos $\mathbf{A} = \Omega$ y aplicamos el axioma II. \square

Propiedad 3. Si $\mathbf{A} \subset \mathbf{B}$, entonces $\mathbf{P}(\mathbf{A}) \leq \mathbf{P}(\mathbf{B})$.

Demostración. Como $\mathbf{A} \subset \mathbf{B}$, tenemos $\mathbf{B} = \mathbf{A} \cup (\mathbf{B} \setminus \mathbf{A})$. Es claro que $\mathbf{B} \setminus \mathbf{A} = \mathbf{B}\overline{\mathbf{A}}$ es un suceso, además los sucesos \mathbf{A} y $\mathbf{B} \setminus \mathbf{A}$ son incompatibles. Por los axiomas III y I tenemos $\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B} \setminus \mathbf{A}) \geq \mathbf{P}(\mathbf{A})$. \square

Propiedad 4. Para cualquier suceso \mathbf{A} se tiene $0 \leq \mathbf{P}(\mathbf{A}) \leq 1$.

Demostración. En virtud del axioma I es suficiente demostrar la segunda desigualdad, la cual se deduce inmediatamente de la propiedad anterior, por la inclusión $\mathbf{A} \subset \Omega$ y el axioma II. \square

Propiedad 5. Para sucesos \mathbf{A} y \mathbf{B} arbitrarios vale la igualdad

$$\mathbf{P}(\mathbf{A} \cup \mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{A} \cap \mathbf{B}). \quad (1.4)$$

Demostración. Es claro que $\mathbf{A} = \mathbf{AB} \cup \mathbf{A}\bar{\mathbf{B}}$ y $\mathbf{B} = \mathbf{AB} \cup \bar{\mathbf{A}}\mathbf{B}$. Como los sucesos en las dos sumas anteriores son incompatibles, por el axioma III, resulta

$$\mathbf{P}(\mathbf{A}) = \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{A}\bar{\mathbf{B}}), \quad \mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\bar{\mathbf{A}}\mathbf{B}).$$

Tenemos también $\mathbf{A} \cup \mathbf{B} = \mathbf{AB} \cup \mathbf{A}\bar{\mathbf{B}} \cup \bar{\mathbf{A}}\mathbf{B}$, donde a la derecha se suman tres sucesos incompatibles dos a dos. Aplicando nuevamente el axioma III y sustituyendo, resulta

$$\begin{aligned} \mathbf{P}(\mathbf{A} \cup \mathbf{B}) &= \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{A}) - \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{AB}) \\ &= \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{A} \cap \mathbf{B}), \end{aligned}$$

que es la igualdad buscada. \square

Observación. Si los sucesos \mathbf{A} y \mathbf{B} son incompatibles, entonces $\mathbf{P}(\mathbf{AB}) = 0$, y de la fórmula (1.4) se obtiene la igualdad ya conocida $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B})$.

Observación. En forma análoga, no es difícil demostrar, que para tres sucesos \mathbf{A}, \mathbf{B} y \mathbf{C} arbitrarios, tiene lugar la igualdad

$$\begin{aligned} \mathbf{P}(\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}) &= \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) + \mathbf{P}(\mathbf{C}) \\ &\quad - \mathbf{P}(\mathbf{AB}) - \mathbf{P}(\mathbf{AC}) - \mathbf{P}(\mathbf{BC}) + \mathbf{P}(\mathbf{ABC}). \end{aligned}$$

Es posible también demostrar una fórmula general: para sucesos arbitrarios $\mathbf{A}_1, \dots, \mathbf{A}_n$ vale la igualdad

$$\begin{aligned} \mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) &= \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(\mathbf{A}_i \mathbf{A}_j) + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(\mathbf{A}_i \mathbf{A}_j \mathbf{A}_k) \\ &\quad - \dots + (-1)^{n+1} \mathbf{P}(\mathbf{A}_1 \dots \mathbf{A}_n). \end{aligned}$$

Propiedad 6. *Dados n sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ arbitrarios, tiene lugar la desigualdad*

$$\mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) \leq \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i).$$

Demostración. Para $n = 2$ esta desigualdad se obtiene de (1.4). Para $n > 2$, es fácil obtener el resultado anterior, mediante la aplicación sucesiva de la desigualdad $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) \leq \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B})$. \square

Propiedad 7. *Sean $\mathbf{A}_1, \dots, \mathbf{A}_n$ sucesos incompatibles dos a dos, y tales que alguno de ellos ocurre. Entonces $\sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) = 1$.*

Demostración. Tenemos $\mathbf{A}_i \mathbf{A}_j = \emptyset$ cuando $i \neq j$, para $i, j = 1, \dots, n$. Además $\bigcup_{i=1}^n \mathbf{A}_i = \Omega$. De los axiomas II y III obtenemos

$$1 = \mathbf{P}(\Omega) = \mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) = \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i),$$

concluyendo la demostración. \square

Propiedad 8. *Sea $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \mathbf{A}_3 \supset \dots$ una sucesión de sucesos, y designemos $\mathbf{A} = \bigcap_{i=1}^{\infty} \mathbf{A}_i$. Entonces, existe el $\lim_n \mathbf{P}(\mathbf{A}_n)$, y es igual a $\mathbf{P}(\mathbf{A})$.*

Demostración. Para cada n , tenemos

$$\mathbf{A}_n = \bigcup_{k=n}^{\infty} (\mathbf{A}_k \setminus \mathbf{A}_{k+1}) \cup \mathbf{A}.$$

Como los sucesos que aparecen a la derecha son incompatibles dos a dos, utilizando el axioma III, obtenemos

$$\mathbf{P}(\mathbf{A}_n) = \sum_{k=n}^{\infty} \mathbf{P}(\mathbf{A}_k \setminus \mathbf{A}_{k+1}) + \mathbf{P}(\mathbf{A}). \quad (1.5)$$

Si aquí tomamos $n = 1$, resulta

$$\sum_{k=1}^{\infty} \mathbf{P}(\mathbf{A}_k \setminus \mathbf{A}_{k+1}) \leq \mathbf{P}(\mathbf{A}_1) \leq 1.$$

De la convergencia de la serie a la izquierda en la fórmula anterior, surge que

$$\sum_{k=n}^{\infty} \mathbf{P}(\mathbf{A}_k \setminus \mathbf{A}_{k+1}) \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

Tomando límite en ambos lados de la igualdad (1.5), obtenemos

$$\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_n) = \mathbf{P}(\mathbf{A}),$$

concluyendo la demostración. \square

El siguiente resultado tiene utilidad práctica

Teorema 1. *Sea \mathbf{P} una medida de probabilidad finitamente aditiva definida en un álgebra \mathcal{F} . Las cuatro siguientes afirmaciones son equivalentes:*

(1) \mathbf{P} es σ -aditiva (es decir, \mathbf{P} es una probabilidad).

(2) \mathbf{P} es continua por abajo, es decir, dada la sucesión creciente de conjuntos de \mathcal{F}

$$\mathbf{A}_1 \subset \mathbf{A}_2 \subset \cdots, \text{ con } \bigcup_{n=1}^{\infty} \mathbf{A}_n \in \mathcal{F},$$

se verifica

$$\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_n) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \mathbf{A}_n\right).$$

(3) \mathbf{P} es continua por arriba, es decir, dada la sucesión decreciente de conjuntos de \mathcal{F}

$$\mathbf{A}_1 \supset \mathbf{A}_2 \supset \cdots, \text{ con } \bigcap_{n=1}^{\infty} \mathbf{A}_n \in \mathcal{F},$$

se verifica

$$\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_n) = \mathbf{P}\left(\bigcap_{n=1}^{\infty} \mathbf{A}_n\right).$$

(4) \mathbf{P} es continua en \emptyset , el conjunto vacío. Es decir, dada la sucesión decreciente al vacío de conjuntos de \mathcal{F}

$$\mathbf{A}_1 \supset \mathbf{A}_2 \supset \cdots, \text{ con } \bigcap_{n=1}^{\infty} \mathbf{A}_n = \emptyset,$$

se verifica

$$\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_n) = 0.$$

Demostración. Ver Shiryaev. \square

Por último en esta sección presentamos la axiomática de la teoría de la probabilidad enunciada por A. N. Kolmogorov en 1933.

Definición 8. Una espacio de probabilidad es una terna $(\Omega, \mathcal{F}, \mathbf{P})$, donde

- (a) Ω es un conjunto de puntos ω , llamado espacio de probabilidad.
- (b) \mathcal{F} es una σ -álgebra,
- (c) \mathbf{P} es una probabilidad en \mathcal{F} .

1.3. Álgebras y σ -álgebras: propiedades y ejemplos

Sea Ω un espacio de probabilidad. Las familias de subconjuntos de Ω dadas por

$$\mathcal{F}_* = \{\emptyset, \Omega\}, \quad \mathcal{F}^* = \{\mathbf{A} : \mathbf{A} \subset \Omega\},$$

son ambas álgebras y σ -álgebras, conocidas como *triviales*. Cuando Ω es un espacio finito, la σ -álgebra \mathcal{F}^* es listable, y se utiliza para definir espacios de probabilidad. Otro ejemplo sencillo de σ -álgebra es cuando se da $\mathbf{A} \subset \Omega$, definiendo

$$\mathcal{F}_{\mathbf{A}} = \{\mathbf{A}, \overline{\mathbf{A}}, \emptyset, \Omega\}.$$

Consideremos ahora una partición numerable $\mathcal{D} = \{\mathbf{D}_n : n = 1, 2, \dots\}$ del espacio Ω , es decir,

$$\Omega = \sum_{n=1}^{\infty} \mathbf{D}_n, \quad \mathbf{D}_i \cap \mathbf{D}_j = \emptyset \ (i \neq j).$$

Es claro ver que

$$\mathcal{F} = \{\text{uniones finitas y numerables de elementos de } \mathcal{D}\}$$

es una σ -álgebra, la mínima que contiene a todos los elementos de \mathcal{D} .

Lema 1. *Dada una familia arbitraria \mathcal{E} de subconjuntos de Ω , existe una mínima σ -álgebra, designada $\sigma(\mathcal{E})$, que contienen todos los subconjuntos de \mathcal{E} .*

Demostración. No es difícil ver que

$$\sigma(\mathcal{E}) = \bigcap \{\mathcal{G} : \mathcal{G} \text{ es } \sigma\text{-álgebra y } \mathcal{E} \subset \mathcal{G}\}$$

es una intersección no vacía (porque \mathcal{F}^* es un de los factores de la intersección) y es una σ -álgebra. Es claro además que es la mínima. \square

A la σ -álgebra mínima cuya existencia acabamos de demostrar se le llama *σ -álgebra generada por \mathcal{E}* .

1.4. Medidas en \mathbf{R}

El objetivo de esta sección es la construcción de medidas en la recta real. Consideremos entonces el conjunto $\mathbf{R} = \{x : -\infty < x < \infty\}$ de los números reales, y los intervalos de la forma

$$(a, b] = \{x \in \mathbf{R} : a < x \leq b\}.$$

para todo $-\infty \leq a < b \leq \infty$, donde entendemos que $(a, \infty] = (a, \infty)$, y la familia de las uniones disjuntas y finitas de intervalos como los anteriores, es decir¹:

$$\mathcal{F} = \left\{ \mathbf{A} \subset \mathbf{R} : \mathbf{A} = \sum_{k=1}^n (a_i, b_i], a_i < b_i \ (i = 1, \dots, n), n = 1, 2, \dots \right\}.$$

No es difícil verificar que \mathcal{F} es un álgebra. Observemos también que \mathcal{F} no es una σ -álgebra, dado que, por ejemplo, $\cup_{n=1}^{\infty} (0, 1 - 1/n] = (0, 1)$ no es un conjunto de \mathcal{F} .

Definición 9. La mínima σ -álgebra generada por \mathcal{F} se denomina σ -álgebra de Borel, y se denota mediante $\mathcal{B}(\mathbf{R})$. Los conjuntos de esta σ -álgebra son los borelianos.

Ejercicio 2. (\star) Verificar que, para $a < b$ reales, los conjuntos de la forma

$$\{a\}, (a, b), [a, b], [a, b), (-\infty, b), (-\infty, b], (a, \infty)$$

son borelianos.

En algunos casos es conveniente trabajar en la *recta real ampliada*, es decir, el conjunto

$$\overline{\mathbf{R}} = \{x : -\infty \leq x \leq \infty\}$$

y construir análogamente la σ -álgebra de Borel, denotada $\mathcal{B}(\overline{\mathbf{R}})$ en este caso.

Consideremos ahora una medida de probabilidad \mathbf{P} definida en $\mathcal{B}(\mathbf{R})$, la σ -álgebra de Borel de la recta real. Podemos definir la función $F : \mathbf{R} \rightarrow \mathbf{R}$, mediante²

$$F(x) = \mathbf{P}(-\infty, x]. \quad (1.6)$$

La función F se llama *función de distribución*, y verifica las siguientes propiedades.

Propiedad 1. Si $a < b$, entonces $F(b) - F(a) = \mathbf{P}(a < b]$, y la función F resulta no decreciente.

Demostración. Aplicando la aditividad de \mathbf{P} , tenemos

$$\mathbf{P}(-\infty, b] = \mathbf{P}(-\infty, a] + \mathbf{P}(a, b].$$

Como $\mathbf{P}(A) \geq 0$ se deduce que F es no decreciente. \square

¹Cuando aparece una suma de conjuntos, se supone además que éstos son disjuntos

²Aquí utilizamos la notación $\mathbf{P}(-\infty, x]$ en vez de la formalmente correcta $\mathbf{P}((-\infty, x])$.

Propiedad 2. *Se tiene $\lim_{x \rightarrow +\infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$.*

Demostración. Consideremos primero una sucesión $x_n \nearrow \infty$. Como tenemos

$$\bigcup_{n=1}^{\infty} (-\infty, x_n] = \mathbf{R},$$

por continuidad de la probabilidad, se verifica

$$F(x_n) = \mathbf{P}(-\infty, x_n] \nearrow \mathbf{P}(\mathbf{R}) = 1.$$

Ahora, si dada una sucesión $y_n \searrow -\infty$. Como tenemos

$$\bigcap_{n=1}^{\infty} (-\infty, y_n] = \emptyset,$$

y, por continuidad de la probabilidad en el vacío, se verifica

$$F(y_n) = \mathbf{P}(-\infty, y_n] \searrow \mathbf{P}(\emptyset) = 0.$$

□

Propiedad 3. *La función de distribución $F(x)$ es continua por la derecha y tiene límite a la izquierda.*

Demostración. Sea x un real arbitrario. Consideremos una sucesión numérica x_1, x_2, \dots decreciente y convergente a x . Es decir $x_1 > x_2 > \dots$, $x_n \rightarrow x$ ($n \rightarrow \infty$). Sea $\mathbf{A}_n = (x, x_n]$. Es claro que $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \dots$, y que $\bigcap_{n=1}^{\infty} \mathbf{A}_n = \emptyset$. Por continuidad de la probabilidad en el vacío, existe el $\lim_n \mathbf{P}(\mathbf{A}_n) = \mathbf{P}(\emptyset) = 0$. Entonces, aplicando la propiedad 5, obtenemos $\mathbf{P}(\mathbf{A}_n) = F(x_n) - F(x) \rightarrow 0$ ($n \rightarrow \infty$), lo que concluye la demostración. El límite por la izquierda es análogo. □

El claro que la función F es a priori un objeto más sencillo que la medida \mathbf{P} . Lo interesante es que partir de una función F que verifique las propiedades anteriores siempre se puede construir una medida de probabilidad en los borelianos, que además es la única que verifica (1.6). Dicho resultado se basa el otro fundamental de la teoría de la medida, el *teorema de Carathéodory*, que enunciamos a continuación.

Teorema 2 (Extensión de medidas). *Sea Ω un conjunto. Sea \mathcal{F} un álgebra de subconjuntos de Ω , y sea $\mathcal{B} = \sigma(\mathcal{F})$ la mínima σ -álgebra engendrada por \mathcal{F} . Sea μ_0 una medida σ -aditiva definida en el álgebra \mathcal{F} . Existe una medida μ definida en la σ -álgebra \mathcal{B} , que es la única que verifica*

$$\mu(\mathbf{A}) = \mu_0(\mathbf{A}), \quad \text{para todo conjunto } \mathbf{A} \in \mathcal{F}.$$

Sigue ahora el resultado que construye una medida de probabilidad a partir de una función de distribución.

Teorema 3. *Sea $F: \mathbf{R} \rightarrow \mathbf{R}$ una función de distribución (es decir, una función que verifica las propiedades (5), (2) y (3)). Existe una medida de probabilidad definida en $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, que es la única que verifica*

$$F(b) - F(a) = \mathbf{P}(a, b], \quad \text{para todo } -\infty \leq a < b \leq \infty$$

Demostración. Consideremos el álgebra de las uniones finitas de intervalos semi-abiertos, es decir

$$\mathcal{F} = \left\{ \mathbf{A} \subset \mathbf{R} : \mathbf{A} = \sum_{k=1}^n (a_k, b_k], a_k < b_k \ (k = 1, \dots, n), n = 1, 2, \dots \right\}.$$

Definamos una medida \mathbf{P}_0 en éste álgebra mediante

$$\mathbf{P}_0(\mathbf{A}) = \sum_{k=1}^n (F(b_k) - F(a_k)).$$

No es difícil ver que \mathbf{P}_0 es aditiva en el álgebra. Vamos a aplicar el Teorema 2 de Charathéodory, para lo que tenemos que verificar que \mathbf{P}_0 es σ -aditiva en \mathcal{F} . Aplicando el Teorema 1, sabemos que es suficiente verificar que \mathbf{P}_0 es continua en \emptyset . Sea entonces una sucesión $\{A_n\}$ decreciente de conjuntos de \mathcal{F} con intersección vacía, es decir

$$\mathbf{A}_1 = \sum_{k=1}^{n_1} (a_k^1, b_k^1] \supset \mathbf{A}_2 = \sum_{k=1}^{n_2} (a_k^2, b_k^2] \supset \dots, \quad \bigcap_{n=1}^{\infty} \mathbf{A}_n = \emptyset.$$

Supongamos primero que existe $N \in \mathbf{N}$ tal que $[-N, N] \supset \mathbf{A}_1$. Sea $\varepsilon > 0$. Elegimos para cada n existe un conjunto $\mathbf{B}_n \in \mathcal{F}$ que verifica³

$$\frac{[\mathbf{B}_n] \subset \mathbf{A}_n, \quad \mathbf{P}_0(\mathbf{A}_n - \mathbf{B}_n)}{2^n \text{ @.}}$$

Como $\bigcap_{n=1}^{\infty} \mathbf{A}_n = \emptyset$ tenemos $\bigcap_{n=1}^{\infty} [\mathbf{B}_n] = \emptyset$. Como además $[\mathbf{B}_n] \subset [-N, N]$ que es un conjunto compacto en la recta real, podemos⁴ elegir n_0 tal que

³ $[\mathbf{A}]$ denota la clausura de \mathbf{A} .

⁴Utilizamos una forma equivalente de la definición de conjunto compacto, que se puede demostrar tomando complementos.

$\bigcap_{n=1}^{n_0} [\mathbf{B}_n] = \emptyset$. Tenemos entonces

$$\begin{aligned} \mathbf{P}_0(\mathbf{A}_{n_0}) &= \mathbf{P}_0\left(\mathbf{A}_{n_0} - \bigcap_{n=1}^{n_0} \mathbf{B}_n\right) + \mathbf{P}_0\left(\bigcap_{n=1}^{n_0} [\mathbf{B}_n]\right) \\ &= \mathbf{P}_0\left(\mathbf{A}_{n_0} - \bigcap_{n=1}^{n_0} \mathbf{B}_n\right) \\ &\leq \mathbf{P}_0\left(\bigcap_{n=1}^{n_0} \mathbf{A}_n - \mathbf{B}_n\right) \leq \sum_{n=1}^{n_0} \mathbf{P}_0(\mathbf{A}_n - \mathbf{B}_n) < \varepsilon. \end{aligned}$$

lo que demuestra la continuidad en \emptyset , en este caso. Veamos ahora el caso en el \mathbf{A}_1 no es acotado. Dado $\varepsilon > 0$, aplicando la propiedad 2 existe $N \in \mathbf{N}$ tal que

$$\mathbf{P}_0(\overline{-N, N}) < \frac{\varepsilon}{2}.$$

Para ese N , por la demostración anterior, existe n_0 tal que

$$\mathbf{P}_0(\mathbf{A}_{n_0} \cap (-N, N]) < \frac{\varepsilon}{2}.$$

Por la monotonía de \mathbf{P}_0 , tenemos que si $n \geq n_0$, se verifica

$$\mathbf{P}_0(\mathbf{A}_n) \leq \mathbf{P}_0(\mathbf{A}_{n_0}) \leq \mathbf{P}_0(\mathbf{A}_{n_0} \cap (-N, N]) + \mathbf{P}_0(\overline{-N, N}) < \varepsilon,$$

concluyendo la demostración del teorema. \square

Observación. Con exactamente los mismos argumentos podemos construir medidas finitas, en el caso en que tengamos una función F que es no decreciente, continua a la derecha con límite a la izquierda, y que $F(\infty) - F(-\infty) < \infty$.

También se puede construir una medida en la recta real si F es no decreciente, continua a la derecha con límite a la izquierda, pero $F(\infty) - F(-\infty) = \infty$, como es ejemplo de la función $F(x) = x$. En este caso también se puede construir una medida en \mathbf{R} (que en el caso de $F(x) = x$ es la *medida de Lebesgue*, mediante el teorema anterior y un argumento de aproximación (ver Shiryaev pp.158-159.)

1.5. Ejemplos de medidas de probabilidad en \mathbf{R}

El teorema de la clase anterior nos dice que para construir una medida de probabilidad en $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ puede hacerse a partir de una función de distribución. Veamos algunos ejemplos y algunas clases de distribuciones.

Ejemplo 1. Consideremos la función de distribución

$$F(x) = \begin{cases} 0, & \text{cuando } x < 0, \\ x, & \text{cuando } 0 \leq x \leq 1 \\ 1, & \text{cuando } 1 < x. \end{cases}$$

La probabilidad que se obtiene, denominada λ , se denomina *medida de Lebesgue en* $[0, 1]$, y es la única que verifica $\lambda(a, b) = b - a$ para $0 \leq a < b \leq 1$. Observemos que la medida de un punto es nula, porque

$$\lambda(\{a\}) = \lim_{n \rightarrow \infty} \lambda\left(a - \frac{1}{n}, a\right] = \lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

Podemos demostrar también que la medida de cualquier conjunto numerable es cero. Por ejemplo $\lambda(\mathbf{Q}) = 0$. Como la medida de los conjuntos $(-\infty, 0)$ y $(0, \infty)$ es nula, podemos restringir el espacio de probabilidad considerando $\Omega = [0, 1]$. La σ -álgebra que consideramos es

$$\mathcal{B}([0, 1]) = \{\mathbf{A} \cap [0, 1] : \mathbf{A} \in \mathcal{B}(\mathbf{R})\}.$$

Ejemplo 2. La construcción anterior nos permite responder a la pregunta inicial, relativa a cómo construir una medida en el espacio $\Omega = \{0, 1\}^N$ de las sucesiones de ceros y unos. La biyección que tenemos

$$x \mapsto \frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \dots$$

La idea es “transferir” la medida desde el intervalo $[0, 1)$ al conjunto de sucesiones. Identifiquemos los conjuntos que establece esta biyección. Consideremos entonces una tira e_1, \dots, e_N de N ceros y unos y el conjunto de todas las sucesiones que comienzan con esta tira. Tenemos

$$\{\omega \in \Omega : a_1 = e_1, \dots, a_N = e_N\} = \bigcap_{n=1}^N \bigcup_{j=0}^{2^{n-1}-1} \left[\frac{2j + e_n}{2^n}, \frac{2j + 1 + e_n}{2^n} \right). \quad (1.7)$$

Esto nos da por ejemplo

$$\{\omega \in \Omega : a_1 = 1\} = \left[\frac{1}{2}, 1 \right), \quad \{\omega \in \Omega : a_1 = 0, a_2 = 1\} = \left[\frac{1}{4}, \frac{1}{2} \right).$$

La conclusión de (1.7) es que a una tira de largo N corresponde un intervalo de la forma $[k/2^N, (k+1)/2^N)$, que tiene medida $1/2^N$. Podemos entonces asignar

$$\mathbf{P}(\{\omega \in \Omega : a_1 = e_1, \dots, a_N = e_N\}) = \frac{1}{2^N}. \quad (1.8)$$

De aquí se concluye que la medida en el espacio Ω de sucesiones de ceros y unos es la misma que la que se obtiene aplicando el Teorema de Charathéodory a la medida en el álgebra en Ω de los conjuntos que son uniones finitas de conjuntos de la forma

$$\{\omega \in \Omega: a_1 = e_1, \dots, a_N = e_N\}.$$

con la medida finitamente aditiva definida en (1.8).

Ejercicio 3. (★★) Construir directamente la medida \mathbf{P} que verifica (1.8) aplicando el teorema de Charathéodory.

Ejercicio 4. (★★★) Es claro que la medida que estamos construyendo corresponde a una sucesión de ensayos de Bernoulli independientes y equilibrados ($p = 1/2$). Investigar que ocurre si queremos construir una medida que corresponda a ensayos de Bernoulli no equilibrados ($p \neq 1/2$). Sugerencia: Construir la función de distribución en \mathbf{R} obtenida a partir de la biyección anterior, comenzando por algunos puntos racionales.

1.5.1. Medidas discretas en \mathbf{R}

Dada una sucesión $\{x_n\}$ en la recta real y una sucesión $\{p_n\}$ de números positivos tales que $\sum_n p_n = 1$ podemos construir una función $F: \mathbf{R} \rightarrow \mathbf{R}$ de la forma

$$F(x) = \sum_{n: x_n \leq x} p_n. \quad (1.9)$$

No es difícil ver que verifica las propiedades que definen una función de distribución: es claramente no decreciente, por convergencia de series se verifica $F(-\infty) = 0$ y $F(\infty) = 1$, y es continua por la derecha y tiene límite por la izquierda. Ésto último es claro cuando x_n es un punto aislado, dado que la función F es constante en pequeños entornos a la derecha y a la izquierda de x_n . En el caso en que x_n es un punto de acumulación de la sucesión, el argumento es más fino, y se basa en el hecho de que suficientemente cerca las sumas que aparecen son restos de series convergentes, que por lo tanto tienden a cero. Estas funciones de distribución se llaman *discretas*. Tenemos entonces una medida de probabilidad en la recta. Denotando $\mathbf{D} = \{x_n\}$ el recorrido de la sucesión, tenemos

$$\mathbf{P}(\{x_n\}) = \Delta F(x_n) := F(x_n) - F(x_{n-}), \quad \mathbf{P}(\mathbf{D}) = 1.$$

Ejemplo 3. Consideremos la función de distribución

$$F(x) = \begin{cases} 0, & \text{cuando } x < a, \\ 1, & \text{cuando } a \leq x. \end{cases}$$

La probabilidad que se denomina *delta de Dirac* en el punto a , y es usual designarla mediante δ_a . No es difícil ver que esta probabilidad es discreta, y concentra toda la probabilidad en el punto a , es decir

$$\delta_1(\mathbf{A}) = \begin{cases} 1, & \text{si } a \in \mathbf{A}, \\ 0, & \text{si } a \notin \mathbf{A}. \end{cases}$$

1.5.2. Medidas absolutamente continuas en \mathbf{R}

Consideremos una función $f: \mathbf{R} \rightarrow [0, \infty)$ integrable según Riemman en toda la recta real, y que verifica $\int_{-\infty}^{\infty} f(x)dx = 1$. Podemos construir una función $F: \mathbf{R} \rightarrow \mathbf{R}$ mediante la fórmula

$$F(x) = \int_{-\infty}^x f(t)dt. \quad (1.10)$$

Aquí es sencillo ver, a partir de la fórmula anterior que F verifica la definición de función de distribución, resultando además continua en todos los puntos. Las funciones de distribución que verifican (1.30) se llaman *absolutamente continuas*. La función f se llama *densidad*. Si tomamos

$$f(x) = \mathbf{1}_{[0,1]}(x)$$

integrando obtenemos el modelo del Ejemplo 1, que resulta entonces tener una distribución absolutamente continua. Otro ejemplo consiste en considerar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

que se llama *densidad normal* o también *densidad gaussiana* con parámetros μ y σ .

1.5.3. Medidas singulares

Es interesante la pregunta si las dos clases anteriores de funciones de distribución son exhaustivas, es decir, si toda medida de probabilidad en la recta corresponde a alguna de estas dos clases. En primer lugar observemos que el conjunto de puntos de discontinuidad de una función de distribución es a lo sumo numerable:

Propiedad 4. *Una función de distribución tiene una cantidad finita o numerable de puntos de discontinuidad.*

Demostración. Como es acotada y monótona, la función F tiene:

- a lo sumo un salto de magnitud h , con $h > 1/2$,
- a lo sumo dos saltos de magnitud h , con $1/2 \geq h > 1/3$,
- a lo sumo tres saltos de magnitud h , con $1/3 \geq h > 1/4$,
- \vdots
- a lo sumo m saltos de magnitud h , con $1/m \geq h > 1/(m+1)$,
- \vdots

En consecuencia, el conjunto de los puntos de discontinuidad de la función $F(x)$ es finito o numerable, dado que sus elementos se pueden numerar. \square

Consideremos entonces una función de distribución F arbitraria y denotamos $\mathbf{D} = \{x_n\}$ el conjunto de sus puntos de discontinuidad, y

$$p_n = \Delta F(x_n).$$

Veamos que si $\sum_n p_n = 1$ tenemos una función de distribución discreta. En efecto, dado un par $a < b$, de la construcción de la sucesión $\{x_n\}$ obtenemos

$$F(b) - F(a) \geq \sum_{n: x_n \in (a,b)} p_n.$$

Si la desigualdad anterior no es una igualdad para algún par $a < b$, obtenemos $F(\infty) - F(-\infty) > 1$, y de la igualdad anterior tomando límite si $a \rightarrow -\infty$ obtenemos (1.32).

En caso de que $0 < \sum_n p_n < 1$, podemos definir las funciones no decrecientes y continuas a la derecha y con límite a la izquierda, mediante

$$F_d(x) = \sum_{n: x_n \leq x} p_n, \quad F_c(x) = F(x) - F_d(x).$$

Es claro que F_c es continua. La distribución $F = F_c + F_d$ en este caso es una *mezcla* de una distribución continua y otra discreta.

Por último si no hay puntos de discontinuidad, entonces F es continua. Tenemos entonces una primer respuesta negativa a la pregunta inicial: dada

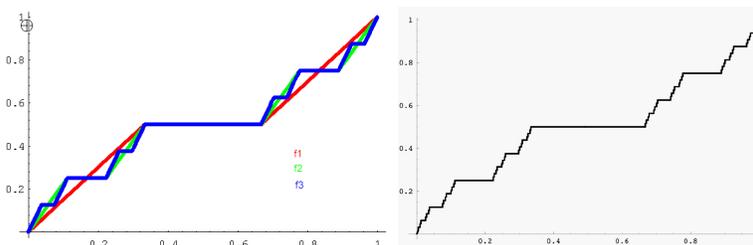
una distribución discreta G , una absolutamente continua H , y un número $0 < \lambda < 1$, la mezcla

$$F(x) = \lambda G(x) + (1 - \lambda)H(x), \quad (1.11)$$

es una distribución que no es ni discreta ni absolutamente continua.

La segunda pregunta es entonces si todas las distribuciones continuas son absolutamente continuas, es nos permitiría afirmar que todas las distribuciones son de la forma (1.11). La respuesta a esa pregunta es negativa, como muestra el siguiente ejemplo.

Ejemplo 4 (Función de Cantor). Ver Shiryaev pp. 156-158. A la izquierda



las primeras tres funciones de la sucesión cuyo límite es la función de Cantor, y a la derecha una función con n grande

1.6. Medidas en \mathbf{R}^n

Nuestro interés es construir ahora medidas en

$$\Omega = \mathbf{R}^n = \{x = (x_1, \dots, x_n) : x_i \in \mathbf{R}, i = 1, \dots, n\},$$

el producto cartesiano de \mathbf{R} consigo mismo n veces. Comenzamos construyendo la σ -álgebra de Borel en Ω .

La forma más directa en este caso, que es análoga a la utilizada en \mathbf{R} , es la siguiente. Consideremos los conjuntos de la forma

$$I = (a_1, b_1] \times \dots \times (a_n, b_n] = \{x : a_i < x_i \leq b_i, i = 1, \dots, n.\} \quad (1.12)$$

que llamamos *rectángulos*, que es el producto cartesiano de n intervalos semi-abiertos $I_i = (a_i, b_i]$ ($i = 1, \dots, n$).

No es difícil ver que la unión finita de rectángulos disjuntos, denotada \mathcal{F} , es un álgebra. La σ -álgebra de Borel en \mathbf{R}^n , denotada mediante $\mathcal{B}(\mathbf{R}^n)$ es entonces la mínima σ -álgebra generada por \mathcal{F} , es decir $\sigma(\mathcal{F})$.

Una construcción alternativa de la σ -álgebra de Borel es la que se obtiene procediendo a la construcción estándar de la sigma álgebra de Borel en un espacio producto. En ese caso se considera la clase de los rectángulos de lados borelianos, es decir los subconjuntos de \mathbf{R}^n de la forma $B = B_1 \times \cdots \times B_n$ donde cada B_i es un boreliano de \mathbf{R} . La σ -álgebra producto es la mínima que contiene a todos los rectángulos de lados borelianos, que es denotada

$$\mathcal{F} = \mathcal{B}(\mathbf{R}) \otimes \cdots \otimes \mathcal{B}(\mathbf{R}),$$

(donde hay n factores). Como los rectángulos de la forma (1.12) son rectángulos de lados borelianos, es claro que $\mathcal{B}(\mathbf{R}^n) \subset \mathcal{F}$. En realidad estas dos σ -álgebras son iguales.

Demostración. Veamos que $\mathbf{B}_1 \times \mathbf{B}_2 \in \mathcal{B}(\mathbf{R}^2)$, con $\mathbf{B}_1, \mathbf{B}_2$ borelianos en \mathbf{R} . Como

$$\mathbf{B}_1 \times \mathbf{B}_2 = (\mathbf{B}_1 \times \mathbf{R}) \cap (\mathbf{R} \times \mathbf{B}_2).$$

es suficiente ver que $\mathbf{B} \times \mathbf{R} \in \mathcal{B}(\mathbf{R}^2)$ para cualquier boreliano \mathbf{B} . Consideremos la siguiente familia de conjuntos en \mathbf{R} .

$$\mathcal{E} = \{\mathbf{A} \subset \mathbf{R} : \mathbf{A} \times \mathbf{R} \in \mathcal{B}(\mathbf{R}^2)\}$$

Es claro que los intervalos de la forma $I = (a, b] \in \mathcal{E}$. Veamos además que \mathcal{E} es una σ -álgebra:

- (a) Tenemos $\mathbf{R} \in \mathcal{E}$ porque $\mathbf{R}^2 \in \mathcal{B}(\mathbf{R}^2)$.
- (b) Sea ahora $\{\mathbf{A}_n\} \subset \mathcal{E}$. Sabemos que $\mathbf{A}_n \times \mathbf{R} \in \mathcal{B}(\mathbf{R}^2)$ para todo n , es decir $\cup_n (\mathbf{A}_n \times \mathbf{R}) \in \mathcal{B}(\mathbf{R}^2)$. Pero

$$\cup_n (\mathbf{A}_n \times \mathbf{R}) = (\cup_n \mathbf{A}_n) \times \mathbf{R},$$

por lo que $\cup_n \mathbf{A}_n \in \mathcal{E}$.

- (c) La demostración es similar, la igualdad clave en este caso es

$$\overline{\mathbf{A}} \times \mathbf{R} = \overline{\mathbf{A} \times \mathbf{R}}$$

Eso concluye que \mathcal{E} es una σ -álgebra que contiene a los intervalos, por lo tanto contiene a todos los borelianos, es decir, $\mathbf{B} \times \mathbf{R} \in \mathcal{B}(\mathbf{R}^2)$, concluyendo la demostración. \square

Ejercicio 5. (\star) Generalizar la demostración anterior a \mathbf{R}^n .

Veamos ahora como construir medidas en el espacio medible $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$. Seguimos un argumento similar al recorrido en \mathbf{R} . Consideremos entonces una probabilidad \mathbf{P} en $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$, consideremos para cada $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ el cuadrante de la forma

$$(-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_n],$$

y la función $F: \mathbf{R}^n \rightarrow \mathbf{R}$ definida mediante

$$F(x) = \mathbf{P}(-\infty, x]. \quad (1.13)$$

Veamos como se calcula los incrementos de F . Si $n = 2$, y tenemos $(a, b] = (a_1, b_1] \times (a_2, b_2]$, denotando $c = (a_1, b_2)$ y $d = (a_2, b_1)$, podemos escribir la unión de rectángulos

$$(-\infty, b] = (-\infty, a_1] \times (a_2, b_2] + (a_1, b_1] \times (-\infty, a_2] + (-\infty, a] + (a, b].$$

Basado en esta descomposición y en (1.13), obtenemos

$$\mathbf{P}(a, b] = F(b) + F(a) - F(a_1, b_2) - F(a_2, b_1). \quad (1.14)$$

La parte izquierda de la fórmula anterior se denomina *incremento* de F en $(a, b]$ y puede obtenerse de la siguiente forma. Definamos los operadores Δ_{a_i, b_i} asociado a un intervalo $(a_i, b_i]$ para $i = 1, 2$, mediante las fórmulas

$$\begin{aligned} \Delta_{a_1, b_1} F(x, y) &= F(b_1, y) - F(a_1, y), \\ \Delta_{a_2, b_2} F(x, y) &= F(x, b_2) - F(x, a_2). \end{aligned}$$

Con esta notación, cuyo objetivo es generalizar la definición de incremento a \mathbf{R}^n , podemos escribir la fórmula (1.14) como

$$\mathbf{P}(a, b] = F(b) + F(a) - F(a_1, b_2) - F(a_2, b_1) = \Delta_{a_1, b_1} \Delta_{a_2, b_2} F(x, y).$$

Se puede demostrar por inducción en n , que a partir de (1.13), en el caso general del espacio $\Omega = \mathbf{R}^n$ obtenemos

$$\mathbf{P}(a, b] = \Delta_{a_1, b_1} \cdots \Delta_{a_n, b_n} F(x).$$

De aquí obtenemos que si F se define a partir de la fórmula (1.13), entonces

$$\Delta_{a, b} F(x) := \Delta_{a_1, b_1} \cdots \Delta_{a_n, b_n} F(x) \geq 0. \quad (1.15)$$

Observamos también, por la continuidad de la medida \mathbf{P} que si definimos $a = (a_1, \dots, a_n) \leq b = (b_1, \dots, b_n)$ cuando $a_i \leq b_i$ para todo $i = 1, \dots, n$, tenemos

$$F(x_n) \rightarrow F(x), \quad \text{cuando } x_n \searrow x. \quad (1.16)$$

Por último observemos que

$$\lim_{x \rightarrow \infty} F(x) = 1, \quad \lim_{x_i \rightarrow -\infty} F(x) = 0, \quad (i = 1, \dots, n). \quad (1.17)$$

Una función $F: \mathbf{R}^n \rightarrow \mathbf{R}$ que verifica las propiedades (1.15), (1.16) y (1.17) se llama *función de distribución en \mathbf{R}^n* .

Con el teorema de extensión de medidas, y con argumentos análogos a los del teorema en \mathbf{R} , podemos demostrar el siguiente resultado.

Teorema 4. *Sea F una función de distribución en \mathbf{R}^n . Existe una única medida de probabilidad en $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ tal que*

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

donde $\Delta_{a,b}$ es el incremento iterado definido en (1.15).

1.6.1. Ejemplos

Ejemplo 5. Medidas producto. Consideremos medidas de probabilidad, digamos

$$(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P}), \quad (\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{Q}).$$

Sabemos construir un espacio producto $\mathbf{R} \times \mathbf{R} = \mathbf{R}^2$ y hemos construido la σ -álgebra producto $\mathcal{B}(\mathbf{R}) \otimes \mathcal{B}(\mathbf{R}) = \mathcal{B}(\mathbf{R}^2)$, la σ -álgebra de Borel. Para construir una medida producto consideramos las distribuciones asociadas a cada modelo

$$F_{\mathbf{P}}(x) = \mathbf{P}(-\infty, x], \quad F_{\mathbf{Q}}(x) = \mathbf{Q}(-\infty, x].$$

No es difícil ver que

$$F(x, y) = F_{\mathbf{P}}(x)F_{\mathbf{Q}}(y), \quad (1.18)$$

es una función de distribución en \mathbf{R}^2 , por lo que de la aplicación del teorema 4 deducimos la existencia de una medida, designada $\mathbf{P} \times \mathbf{Q}$ en $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2))$ que llamamos medida producto. Es sencillo de ver a partir de (1.14) y (1.18), que para dos intervalos $I = (a, b]$ $J = (c, d]$

$$(\mathbf{P} \times \mathbf{Q})(I \times J) = \mathbf{P}(I)\mathbf{Q}(J).$$

Ejercicio 6. (**) Demostrar que

$$(\mathbf{P} \times \mathbf{Q})(\mathbf{A} \times \mathbf{B}) = \mathbf{P}(\mathbf{A})\mathbf{Q}(\mathbf{B}).$$

para \mathbf{A}, \mathbf{B} borelianos en \mathbf{R} .

Ejemplo 6. Medidas discretas en \mathbf{R}^n . Dada una sucesión $\{x_n\}$ en \mathbf{R}^n y una sucesión $\{p_n\}$ de números positivos tales que $\sum_n p_n = 1$ podemos construir una función $F: \mathbf{R}^n \rightarrow \mathbf{R}$ de la forma

$$F(x) = \sum_{n: x_n \leq x} p_n. \quad (1.19)$$

No es difícil ver que verifica las propiedades que definen una función de distribución. La medida construida a partir del teorema 4 se denomina *medida discreta en \mathbf{R}^n* .

Ejercicio 7. (★★) Se considera la medida discreta en \mathbf{R}^2 concentrada en los puntos $\{x_1 = (1, 2), x_2 = (2, 1)\}$, con probabilidades $p_1 = 1/2, p_2 = 1/2$. Demostrar que esta medida no es una medida producto.

1.6.2. Medidas absolutamente continuas en \mathbf{R}^n

Consideremos una función $f: \mathbf{R}^n \rightarrow [0, \infty)$ integrable según Riemman en toda \mathbf{R}^n , y que verifica

$$\int_{\mathbf{R}^n} f(x) dx = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1,$$

que llamamos *densidad*. Podemos construir una función $F: \mathbf{R}^n \rightarrow \mathbf{R}$ mediante la fórmula

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(t_1, \dots, t_n) dt_1 \dots dt_n = \int_{(-\infty, x]} f(t) dt. \quad (1.20)$$

Aquí es sencillo ver, a partir de la fórmula anterior que F verifica la definición de función de distribución, resultando además continua en todos los puntos.

La *distribución uniforme en $[0, 1]^n$* se obtiene considerando la densidad

$$f(x) = \mathbf{1}_{[0,1]^n}(x)$$

integrando obtenemos la distribución, que tiene la forma

$$F(x) = x_1 \times \cdots \times x_n = F_1(x_1) \times \cdots \times F(x_n)$$

que resulta ser el producto de las distribuciones unidimensionales uniformes en $[0, 1]$, y por lo tanto una medida producto.

Ejemplo 7. Dado un vector $\mu \in \mathbf{R}^n$ y una matriz B definida positiva, la densidad dada por

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} e^{-\frac{1}{2}(x-\mu)^t B^{-1}(x-\mu)} \quad (1.21)$$

da lugar a la *distribución normal n -dimensional*.

Ejercicio 8. Demostrar que la densidad definida por la fórmula (1.21) verifica $\int_{\mathbf{R}^n} f(x) dx = 1$.

1.7. Medidas de probabilidad en \mathbf{R}^∞

Nuestro espacio de probabilidad en este caso es el de las sucesiones reales, es decir

$$\Omega = \{\omega: \omega = (x_1, x_2, \dots), x_n \in \mathbf{R}, n = 1, 2, \dots\},$$

que puede verse como el producto cartesiano de \mathbf{R} una cantidad numerable de veces, es decir \mathbf{R}^∞ . Tenemos que construir una σ -álgebra en este espacio. La idea es construir una clase suficientemente rica de conjuntos pero a la vez suficientemente sencilla. La propuesta central es considerar *cilindros*, es decir, conjuntos con restricciones en un número finito de coordenadas. Tenemos tres opciones para construir cilindros:

1. Aquellos cuya base es un rectángulo (producto de intervalos) en \mathbf{R}^n ($n = 1, 2, \dots$), es decir, dado $\mathbf{I} = I_1 \times \dots \times I_n$

$$\mathbf{I} \times \mathbf{R}^\infty = \{x \in \mathbf{R}^\infty: (x_1, \dots, x_n) \in \mathbf{I}\}.$$

2. Aquellos cuya base es un producto de borelianos reales, es decir

$$\mathbf{B}_1 \times \dots \times \mathbf{B}_n \times \mathbf{R}^\infty = \{x \in \mathbf{R}^\infty: (x_1, \dots, x_n) \in \mathbf{B}_1 \times \dots \times \mathbf{B}_n\}$$

3. Aquellos cuya base es un boreliano \mathbf{B}^n de \mathbf{R}^n , es decir

$$\mathbf{B}^n \times \mathbf{R}^\infty = \{x \in \mathbf{R}^\infty: (x_1, \dots, x_n) \in \mathbf{B}^n\}.$$

Se generan entonces tres σ -álgebras, digamos \mathcal{B}_1 , \mathcal{B}_2 y \mathcal{B}_3 que son las mínimas que contienen a estas tres clases de cilindros respectivamente. Es claro, dado que el producto de intervalos es un producto de borelianos reales, que a la vez es un boreliano en \mathbf{R}^n , que

$$\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{B}_3.$$

No es difícil ver que estas tres σ -álgebras son en realidad iguales. Basta para eso ver que, dado un boreliano \mathbf{B}^n en \mathbf{R}^n para un n arbitrario, tenemos $\mathbf{B}^n \times \mathbf{R}^\infty \in \mathcal{B}_1$.

Esto se verifica probando que la familia

$$\mathcal{E} = \{\mathbf{A} \subset \mathbf{R}^n : \mathbf{A} \times \mathbf{R}^\infty \in \mathcal{B}_1\}$$

es una σ -álgebra que contiene los productos de intervalos $I_1 \times \cdots \times I_n$ y por lo tanto contiene a \mathbf{B}^n , lo que significa que $\mathbf{B} \times \mathbf{R}^\infty \in \mathcal{B}_1$, es decir $\mathbf{B}^n \times \mathbf{R}^\infty \in \mathcal{B}_1$. La σ -álgebra resultante se denota entonces $\mathcal{B}(\mathbf{R}^\infty)$ y se denomina σ -álgebra de borel.

Consideremos ahora una medida de probabilidad \mathbf{P} definida en el espacio medible $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$. A partir de esta probabilidad, para cada n podemos definir una probabilidad \mathbf{P}_n en $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ de la siguiente forma: dado un boreliano \mathbf{B}^n de $\mathcal{B}(\mathbf{R}^n)$ definimos

$$\mathbf{P}_n(\mathbf{B}^n) = \mathbf{P}(\mathbf{B}^n \times \mathbf{R}^\infty).$$

Tenemos entonces una sucesión de medidas de probabilidad $\mathbf{P}_1, \mathbf{P}_2, \dots$ definidas en $\mathbf{R}, \mathbf{R}^2, \dots$ que verifican la siguiente *condición de compatibilidad*:

$$\mathbf{P}_{n+1}(\mathbf{B}^n \times \mathbf{R}) = \mathbf{P}_n(\mathbf{B}^n) \text{ para todo } \mathbf{B}^n \in \mathcal{B}(\mathbf{R}^n). \quad (1.22)$$

Lo que es interesante es que esta condición es suficiente para la construcción de una medida en \mathbf{R}^∞ , como prueba el siguiente *Teorema de Kolomogorov de extensión de medidas en $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$* :

Teorema 5. *Dada una sucesión de espacios de probabilidad $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P}_1)$, $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2), \mathbf{P}_2)$, \dots tal que las medidas de probabilidad respectivas verifican la condición de compatibilidad (1.22), existe una medida de probabilidad en $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$, que es la única que verifica*

$$\mathbf{P}_n(\mathbf{B}^n) = \mathbf{P}(\mathbf{B}^n \times \mathbf{R}^\infty) \text{ para todo } \mathbf{B}^n \in \mathcal{B}(\mathbf{R}^n), n = 1, 2, \dots \quad (1.23)$$

Demostración. La demostración se basa en el Teorema de Carathéodory, construyendo una medida en el álgebra de las uniones de cilindros de la forma $\mathbf{I}^n \times \mathbf{R}^\infty$, y demostrando que es continua en el conjunto vacío.

Comencemos observando que la representación de los cilindros no es única, porque podemos agregar \mathbf{R} como factor a la base una cantidad arbitraria finita de veces. Veamos entonces que la familia

$$\mathcal{F} = \left\{ \widehat{\mathbf{C}}^n = \sum_{k=1}^n (\mathbf{I}^{\ell_k} \times \mathbf{R}^\infty) \right\}.$$

es un álgebra. Es claro que $\emptyset \in \mathcal{B}(\mathbf{R}^n)$, luego $\emptyset \in \mathcal{F}$. Dados dos cilindros $\mathbf{A}^m \times \mathbf{R}^\infty$, $\mathbf{B}^n \times \mathbf{R}^\infty$, donde (digamos) $m < n$, tenemos el conjunto

$$\mathbf{D}^n = (\mathbf{A}^m \times \mathbf{R}^{n-m}) \cup \mathbf{B}^n,$$

de forma que

$$(\mathbf{A}^m \times \mathbf{R}^\infty) \cup (\mathbf{B}^n \times \mathbf{R}^\infty) = \mathbf{D}^n \times \mathbf{R}^\infty.$$

Análogamente, como

$$\overline{\mathbf{D}^n \times \mathbf{R}^\infty} = \overline{\mathbf{D}^n} \times \mathbf{R}^\infty$$

e $\overline{\mathbf{D}^n}$ es una unión finita de rectángulos, verificamos que \mathcal{F} es un álgebra.

Definimos ahora la medida \mathbf{P} en el álgebra de la siguiente forma: dado un cilindro de la forma $\mathbf{A}^n \times \mathbf{R}^\infty$ asignamos

$$\mathbf{P}(\mathbf{A}^n \times \mathbf{R}^\infty) = \mathbf{P}_n(\mathbf{A}^n),$$

es decir, a cada cilindro le asignamos la probabilidad de su base. No es difícil ver que esta asignación es correcta, si tenemos un mismo cilindro que se representa con bases de distinta dimensión.

Veamos ahora que es aditiva: dados dos cilindros disjuntos, podemos suponer que sus bases tienen la misma dimensión y por lo tanto son disjuntas (de no serlo, no lo serían los cilindros). Designémoslos entonces por $\mathbf{A}^n \times \mathbf{R}^\infty$ y $\mathbf{B}^n \times \mathbf{R}^\infty$, tenemos

$$\begin{aligned} \mathbf{P}((\mathbf{A}^n \times \mathbf{R}^\infty) + (\mathbf{B}^n \times \mathbf{R}^\infty)) &= \mathbf{P}((\mathbf{A}^n + \mathbf{B}^n) \times \mathbf{R}^\infty) = \mathbf{P}_n(\mathbf{A}^n + \mathbf{B}^n) \\ &= \mathbf{P}_n(\mathbf{A}^n) + \mathbf{P}_n(\mathbf{B}^n) = \mathbf{P}(\mathbf{A}^n \times \mathbf{R}^\infty) + \mathbf{P}(\mathbf{B}^n \times \mathbf{R}^\infty). \end{aligned}$$

Veamos ahora la continuidad en el vacío. Sea entonces una sucesión decreciente de cilindros

$$\widehat{\mathbf{A}}^n = \mathbf{A}^n \times \mathbf{R}^\infty,$$

con bases \mathbf{A}^n rectangulares y tales que $\bigcap_{n=1}^\infty \widehat{\mathbf{A}}^n = \emptyset$. Supongamos por absurdo que

$$\lim_n \mathbf{P}(\widehat{\mathbf{A}}^n) = \delta > 0.$$

Es claro que para cada n podemos encontrar un rectángulo compacto $\mathbf{C}^n \subset \mathbf{A}^n$, tal que

$$\mathbf{P}^n(\mathbf{A}^n \setminus \mathbf{C}^n) \leq \frac{\delta}{2^{n+1}}.$$

Ahora, como los \mathbf{C}^n no necesariamente son decrecientes, consideramos

$$\mathbf{D}^n = \bigcap_{k=1}^n \mathbf{C}^k, \quad \widehat{\mathbf{D}}^n = \mathbf{D}^n \times \mathbf{R}^n,$$

Es claro que

$$\bigcap_{n=1}^{\infty} \widehat{\mathbf{D}}^n = \bigcap_{n=1}^{\infty} \widehat{\mathbf{C}}^n = \bigcap_{n=1}^{\infty} \widehat{\mathbf{A}}^n = \emptyset \quad (1.24)$$

Por otra parte

$$\begin{aligned} \mathbf{P}(\widehat{\mathbf{A}}^n \setminus \widehat{\mathbf{D}}^n) &= \mathbf{P}\left(\widehat{\mathbf{A}}^n \cap \overline{\left(\bigcap_{k=1}^n \widehat{\mathbf{C}}^k\right)}\right) = \mathbf{P}\left(\widehat{\mathbf{A}}^n \cap \left(\bigcup_{k=1}^n \overline{\widehat{\mathbf{C}}^k}\right)\right) \\ &= \mathbf{P}\left(\bigcup_{k=1}^n \left(\widehat{\mathbf{A}}^n \cap \overline{\widehat{\mathbf{C}}^k}\right)\right) \\ &\leq \sum_{k=1}^n \mathbf{P}\left(\widehat{\mathbf{A}}^n \setminus \widehat{\mathbf{C}}^k\right) \leq \sum_{k=1}^n \mathbf{P}\left(\widehat{\mathbf{A}}^k \setminus \widehat{\mathbf{C}}^k\right) \leq \frac{\delta}{2}. \end{aligned}$$

De aquí resulta que $\mathbf{P}(\widehat{\mathbf{D}}^n) \geq \delta/2 > 0$. Veamos que esto contradice (1.24). En efecto, como los $\widehat{\mathbf{D}}^n$ tienen medida positiva son no vacíos, es decir, existe una sucesión $\widehat{x}^n \in \widehat{\mathbf{D}}^n$.

Para cada n elegimos, en forma secuencial, una subsucesión de la obtenida anteriormente de forma que las primeras n coordenadas de \widehat{x}^n converjan en \mathbf{R}^n a un punto de \mathbf{D}^n . La sucesión diagonal es convergente a un límite \widehat{x}^0 cuyas primeras n coordenadas están en \mathbf{D}^n , para cada n , es decir $\widehat{x}^0 \in \widehat{\mathbf{D}}^n$ para todo n , lo que contradice (1.24).

Demostremos así la continuidad en el vacío, y la aplicación del Teorema de Caratédory nos dice que existe una medida \mathbf{P} definida en $\mathcal{B}(\mathbf{R}^\infty)$, que verifica (1.23) para todos los cilindros con base rectangular. Como eso caracteriza una medida en \mathbf{R}^n , la condición se verifica para todos los borelianos. \square

Ejemplo 8. Consideremos un espacio de probabilidad $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P})$ y la sucesión $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P}_1), (\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2), \mathbf{P}_2), \dots$, donde

$$\mathbf{P}_n = \mathbf{P} \times \dots \times \mathbf{P}$$

es la medida producto n factores iguales a \mathbf{P} . Veamos que esta sucesión de medidas verifica la condición de compatibilidad (1.22). En efecto, como $\mathbf{P}(\mathbf{R}) = 1$, tenemos

$$\mathbf{P}_{n+1}(\mathbf{B}_1 \times \dots \times \mathbf{B}_n \times \mathbf{R}) = \mathbf{P}(\mathbf{B}_1) \cdots \mathbf{P}(\mathbf{B}_n) \mathbf{P}(\mathbf{R}) = \mathbf{P}_n(\mathbf{B}_1 \times \dots \times \mathbf{B}_n) \quad (1.25)$$

Ahora las medidas en \mathbf{R}^n

$$\mathbf{Q}_1(\mathbf{B}) = \mathbf{P}_{n+1}(\mathbf{B} \times \mathbf{R}), \quad \mathbf{Q}_2(\mathbf{B}) = \mathbf{P}_n(\mathbf{B}),$$

coinciden en los productos de borelianos, y por lo tanto en los borelianos de \mathbf{R}^n , por la unicidad del teorema de construcción de medidas en \mathbf{R}^n , lo que indica que (1.25) se verifica para todo boreliano de \mathbf{R}^n . Podemos aplicar entonces el teorema 5, para deducir la existencia de una medida que denominamos *medida producto en \mathbf{R}^∞* . Esta medida es la que corresponde a una *sucesión de variables aleatorias independientes e idénticamente distribuídas* que es un tema que veremos mas adelante.

Ejemplo 9. En el ejemplo anterior, consideremos la medida de probabilidad en \mathbf{R} , dada por

$$\mathbf{P} = p\delta_0 + (1 - p)\delta_1.$$

donde δ_a es la medida de Dirac en el punto a .

Ejercicio 9. (**) Verificar que la aplicación del teorema 5 genera una medida \mathbf{P} en el espacio de sucesiones de ceros y unos, que, dada una tira e_1, \dots, e_n con $e_i = 0$ ó 1 ($i = 1, \dots, n$) asigna, al conjunto de sucesiones que comienzan por e_1, \dots, e_n , la probabilidad $p^{\#1}(1 - p)^{\#0}$.

1.8. Variables aleatorias

Consideremos (Ω, \mathcal{F}) un espacio medible y $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

Definición 10. Una función real $X: \Omega \rightarrow \mathbf{R}$ se denomina variable aleatoria (función medible en el lenguaje del análisis) cuando

$$\{\omega: X(\omega) \in \mathbf{B}\} \in \mathcal{F} \text{ para todo boreliano } \mathbf{B} \in \mathcal{B}(\mathbf{R}).$$

Equivalentemente, $X^{-1}(B) \in \mathcal{F} \forall B \in \mathcal{B}(\mathbf{R})$.

Ejemplo 10. Dada una partición finita $\mathbf{A}_1, \dots, \mathbf{A}_n$ de Ω , donde $\mathbf{A}_i \in \mathcal{F}$, la función que toma una cantidad finita de valores, dada por

$$X(\omega) = \sum_{k=1}^n x_k \mathbf{1}_{\mathbf{A}_k}(\omega),$$

se denomina *variable aleatoria simple*.

Ejercicio 10. (*) Verificar que X es una variable aleatoria.

Lema 2. Sea \mathcal{E} una familia de conjuntos en \mathbf{R} tal que $\sigma(\mathcal{E}) = \mathcal{B}(\mathbf{R})$. Dada $X: \Omega \rightarrow \mathbf{R}$, es suficiente que $X^{-1}(\mathbf{E}) \in \mathcal{F} \forall \mathbf{E} \in \mathcal{E}$ para que X sea variable aleatoria.

Demostración. Sea la familia

$$\mathcal{D} = \{\mathbf{D} \subset \mathbf{R}: X^{-1}(\mathbf{D}) \in \mathcal{F}\}.$$

Sabemos por hipótesis que $\mathcal{E} \subset \mathcal{D}$. Veamos que \mathcal{D} es una σ -álgebra. En efecto: (a) $X^{-1}(\mathbf{R}) = \Omega \in \mathcal{F}$. Tenemos además, cuando $\{\mathbf{A}_n\} \subset \mathcal{D}$, que

$$X^{-1}(\cup_n \mathbf{A}_n) = \cup_n X^{-1}(\mathbf{A}_n) \in \mathcal{F},$$

luego $\cup_n \mathbf{A}_n \in \mathcal{D}$. Análogamente, si $\mathbf{A} \in \mathcal{D}$, tenemos que

$$X^{-1}(\overline{\mathbf{A}}) = \overline{X^{-1}(\mathbf{A})} \in \mathcal{F},$$

y $\overline{\mathbf{A}} \in \mathcal{D}$. Luego \mathcal{D} es σ -álgebra. Resulta

$$\mathcal{B}(\mathbf{R}) = \sigma(\mathcal{E}) \subset \sigma(\mathcal{D}) = \mathcal{D},$$

por lo que todos los borelianos están en \mathcal{D} , y verifican $X^{-1}(\mathbf{B}) \in \mathcal{F}$. \square

Ejercicio 11. ($\star\star$) Demostrar que para una función cualquiera $X: \Omega \rightarrow \mathbf{R}$ tenemos

$$X^{-1}(\cup_\alpha \mathbf{A}_\alpha) = \cup_\alpha X^{-1}(\mathbf{A}_\alpha),$$

donde α pertenece a un conjunto no necesariamente numerable. Enunciar y demostrar el análogo para la intersección de conjuntos

Ejemplo 11. Consideramos $(\Omega, \mathcal{F}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$. Una función $f: \mathbf{R}^n \rightarrow \mathbf{R}$ tal que $f^{-1}(\mathbf{B}) \in \mathcal{B}(\mathbf{R}^n) \forall \mathbf{B} \in \mathcal{B}(\mathbf{R})$ se llama *función boreliana*. Veamos que si f es una función continua, entonces es boreliana. En efecto, por el teorema anterior, considerando $\mathcal{E} = \{(a, b)\}$ la familia de intervalos abiertos, sabemos que $f^{-1}(a, b)$ es un conjunto abierto en \mathbf{R}^n . Como un conjunto abierto es una unión numerable de bolas abiertas, tenemos que es boreliano en \mathbf{R}^n . Luego, el lema anterior nos dice que f es medible. En particular, son medibles las funciones $f: \mathbf{R} \rightarrow \mathbf{R}$ con fórmulas respectivas

$$|x|, \quad x^n, \quad x^+ = \max(x, 0), \quad x^- = \max(0, -x),$$

y las funciones $f: \mathbf{R}^2 \rightarrow \mathbf{R}$ con fórmulas

$$f(x, y) = x + y, \quad f(x, y) = xy.$$

Ejercicio 12. (\star) Dadas una variable aleatoria $X: \Omega \rightarrow \mathbf{R}$ y una función real y medible $f: \mathbf{R} \rightarrow \mathbf{R}$, demostrar que $f \circ X$ es una variable aleatoria.

Definición 11. *Dados un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$, y una variable aleatoria X , definimos*

- (a) *La función de distribución de la variable aleatoria X , designada $F_X: \mathbf{R} \rightarrow \mathbf{R}$, mediante*

$$F_X(x) = \mathbf{P}(\{\omega: X(\omega) \leq x\}) \text{ para todo } x \in \mathbf{R}.$$

- (a) *La distribución de probabilidad de la variable aleatoria X , a la medida de probabilidad \mathbf{R} , designada $\mathbf{P}_X: \mathcal{B}(\mathbf{R}) \rightarrow \mathbf{R}$, mediante*

$$\mathbf{P}_X(\mathbf{B}) = \mathbf{P}(\{\omega: X(\omega) \in \mathbf{B}\}) \text{ para todo boreliano } \mathbf{B} \in \mathcal{B}(\mathbf{R})$$

Observemos en primer lugar que F_X es efectivamente una función de distribución (es decir, no decreciente, con límites cero y uno en menos y más infinito, y continúa a la derecha con límite a la izquierda), y que se verifica

$$\mathbf{P}_X(a, b] = F_X(b) - F_X(a).$$

Como consecuencia, la medida \mathbf{P}_X es la que se obtiene a partir de la construcción vista anteriormente a partir de F_X .

Observemos también que, si tenemos una distribución F en \mathbf{R} , que nos da una medida \mathbf{P} también en \mathbf{R} , el esquema anterior puede construirse considerando $\Omega = \mathbf{R}$ y la variable aleatoria identidad: $X(x) = x$.

1.8.1. Variables aleatorias y límites

Consideremos una sucesión de variables aleatorias $\{X_n\}$ definidas en un espacio medible (Ω, \mathcal{F}) . Comencemos recordando la definición de límite superior y límite inferior.

Definición 12. *Sea $\{a_n\}$ una sucesión numérica en \mathbf{R} . Definimos el límite superior de la sucesión, que denotamos \limsup , y el límite inferior de la sucesión, que denotamos \liminf , mediante*

$$\limsup_n a_n = \inf_n \left(\sup_{k \geq n} a_k \right), \quad \liminf_n a_n = \sup_n \left(\inf_{k \geq n} a_k \right).$$

Ejercicio 13. $(\star\star)$ Demostrar que si una sucesión $\{a_n\}$ verifica

$$\limsup_n a_n = \liminf_n a_n = a,$$

entonces la sucesión tiene límite a (con la definición usual $\forall \varepsilon \exists N$ de límite).

Para incluir en nuestro estudio los límites infinitos, definimos la recta real ampliada y su σ -álgebra de Borel mediante

$$\bar{\mathbf{R}} = \{\mathbf{R}, -\infty, \infty\}, \quad \mathcal{B}(\bar{\mathbf{R}}) = \sigma(\mathcal{B}(\mathbf{R}), -\infty, \infty).$$

Definimos ahora variables aleatorias en $\bar{\mathbf{R}}$.

Definición 13. Una función real $X: \Omega \rightarrow \bar{\mathbf{R}}$ se denomina variable aleatoria extendida cuando $X^{-1}(\mathbf{B}) \in \mathcal{F} \forall \mathbf{B} \in \mathcal{B}(\bar{\mathbf{R}})$.

Ejercicio 14. (★) (a) Demostrar que $\mathbf{A} \subset \bar{\mathbf{R}} \in \mathcal{B}(\bar{\mathbf{R}})$ si y solo si $\mathbf{A} \cap \mathbf{R} \in \mathcal{B}(\mathbf{R})$. (b) Demostrar que si una variable aleatoria extendida toma únicamente valores reales, entonces es una variable aleatoria.

Volviendo a las variables aleatorias, no es difícil establecer que

Proposición 1. Dada una sucesión $\{X_n\}$ de variables aleatorias extendidas, las funciones

$$(a) \sup_n \{X_n\}, \quad (b) \inf_n \{X_n\}, \quad (c) \limsup_n \{X_n\}, \quad (d) \liminf_n \{X_n\},$$

son variables aleatorias extendidas.

Demostración. (a) se obtiene porque

$$\left\{ \sup_n X_n > x \right\} = \cup_n \{X_n > x\}$$

aplicando el Lema 2. (b) es análogo. (c) y (d) siguen de las definiciones de límite superior e inferior. \square

Por último, tenemos que el límite de una sucesión de variables aleatorias es una variable aleatoria.

Proposición 2. Sea una sucesión $\{X_n\}$ de variables aleatorias extendidas, tal que existe para todo $\omega \in \Omega$ el límite $X(\omega) = \lim_n X_n(\omega)$. Entonces X es una variable aleatoria extendida.

Demostración. Tenemos que $X = \limsup_n X_n$, y aplicamos el resultado anterior. \square

Presentamos por último un resultado que permite aproximar variables aleatorias extendidas mediante una sucesión de variables aleatorias simples.

Teorema 6 (De aproximación). (a) *Dada una variable aleatoria extendida X , existe una sucesión de variables aleatorias simples $\{X_n\}$ que verifican $|X_n| \leq X$, y $X_n \rightarrow X$ en Ω .*

(b) *Si además $X \geq 0$, entonces existe una sucesión de variables aleatorias simples y no negativas $\{X_n\}$ que verifican $X_n \nearrow X$ en Ω .*

Demostración. Comenzamos demostrando (b). No es difícil ver que la sucesión construída mediante

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{\left\{\frac{k-1}{2^n} \leq X(\omega)\right\}} + n \mathbf{1}_{\left\{\frac{n}{2^n} \leq X(\omega)\right\}}$$

verifica la tesis. Para demostrar (a), aproximamos X^+ y X^- como en (b), por sucesiones $\{X_n^+\}$ y $\{X_n^-\}$ respectivamente. Tenemos $0 \leq X_n^\pm \nearrow X^\pm$ en Ω . Definimos $X_n = X_n^+ - X_n^-$, que resulta ser simple. Tenemos

$$\begin{aligned} |X_n| &= X_n^+ + X_n^- \leq X^+ + X^- = |X|, \\ X_n &= X_n^+ - X_n^- \rightarrow X^+ - X^- = X, \end{aligned}$$

lo que concluye la demostración. \square

1.9. Vectores aleatorios

Consideremos (Ω, \mathcal{F}) un espacio medible y $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

Definición 14. *Una función $X = (X_1, \dots, X_n): \Omega \rightarrow \mathbf{R}^n$ se denomina vector aleatorio cuando $X^{-1}(B) \in \mathcal{F} \forall B \in \mathcal{B}(\mathbf{R}^n)$.*

Proposición 3. *Una función $X = (X_1, \dots, X_n)$ es un vector aleatorio si y solo si cada una de sus coordenadas X_k ($k = 1, \dots, n$) es una variable aleatoria.*

Demostración. Supongamos que X es un vector aleatorio. Como la función proyección $\pi_k: \mathbf{R}^n \rightarrow \mathbf{R}$ es continua, $X_k = \pi_k(X)$ es variable aleatoria.

Recíprocamente, si sabemos que las coordenadas son variables aleatorias, consideramos

$$\mathcal{D} = \{\mathbf{D} \subset \mathbf{R}^n: X^{-1}(\mathbf{D}) \in \mathcal{F}\}.$$

No es difícil ver que \mathcal{D} es una σ -álgebra. Además, dados $\mathbf{B}_1, \dots, \mathbf{B}_n$ borelianos reales, tenemos

$$X_1(\mathbf{B}_1) \cap \dots \cap X_n(\mathbf{B}_n) = X^{-1}(\mathbf{B}_1 \times \dots \times \mathbf{B}_n).$$

Esto muestra que $\mathbf{B}_1 \times \dots \times \mathbf{B}_n \in \mathcal{D}$, por lo que \mathcal{D} contiene a todos los borelianos, concluyendo la demostración. \square

Definición 15. Una sucesión $\{X_n\}$ de variables aleatorias se denomina proceso estocástico.

Definición 16. Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ y un vector aleatorio $X = (X_1, \dots, X_n): \Omega \rightarrow \mathbf{R}^n$.

(a) La función $F_X: \mathbf{R}^n \rightarrow \mathbf{R}$ definida mediante

$$F_X(x) = \mathbf{P}(\{\omega: X(\omega) \leq x\})$$

se denomina función de distribución del vector X .

(b) La medida $\mathbf{P}_X: \mathcal{B}(\mathbf{R}^n) \rightarrow \mathbf{R}$ definida mediante

$$\mathbf{P}_X(\mathbf{B}) = \mathbf{P}(\{\omega: X(\omega) \in \mathbf{B}\})$$

se denomina distribución de probabilidad del vector X .

Definición 17. Las variables aleatorias X_1, \dots, X_n se denominan independientes cuando se verifica

$$\mathbf{P}(X_1 \in \mathbf{B}_1, \dots, X_n \in \mathbf{B}_n) = \mathbf{P}(X_1 \in \mathbf{B}_1) \cdots \mathbf{P}(X_n \in \mathbf{B}_n) \quad (1.26)$$

para cualquier n -úpla de borelianos reales $\mathbf{B}_1, \dots, \mathbf{B}_n$.

La independencia de variables aleatorias es una propiedad clave en la teoría de la probabilidad, y se corresponde con la estructura de medidas producto antes estudiada.

Proposición 4. Las siguientes afirmaciones son equivalentes:

(a) Para cualquier n -úpla de borelianos reales $\mathbf{B}_1, \dots, \mathbf{B}_n$, se verifica (1.26).

(b) La función de distribución del vector X se factoriza como producto de las distribuciones de sus variables aleatorias coordenadas, es decir, para todo $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, tenemos

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \quad (1.27)$$

(c) La distribución de probabilidad del vector X es la medida producto de las respectivas distribuciones de probabilidad de sus variables aleatorias coordenadas, es decir

$$\mathbf{P}_X = \mathbf{P}_{X_1} \times \cdots \times \mathbf{P}_{X_n}.$$

Demostración. (a) \Rightarrow (b). Dado $x = (x_1, \dots, x_n)$, aplicando (1.26) a los borelianos $\mathbf{B}_k = (-\infty, x_k]$ ($k = 1, \dots, n$) obtenemos

$$\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbf{P}(X_1 \leq x_1) \cdots \mathbf{P}(X_n \leq x_n)$$

El primer término de esta igualdad es $F_X(x)$, y el segundo $F_{X_1}(x_1) \cdots F_{X_n}(x_n)$, concluyendo (1.27).

(a) \Leftrightarrow (c) Consideremos $\mathbf{B} = \mathbf{B}_1 \times \cdots \times \mathbf{B}_k$. Recordando que $\mathbf{P}_X(\mathbf{B}) = \mathbf{P}(X \in \mathbf{B})$ y que $\mathbf{P}_{X_k}(\mathbf{B}_k) = \mathbf{P}(X_k \in \mathbf{B}_k)$ ($k = 1, \dots, n$), la fórmula (1.26) es equivalente a

$$\mathbf{P}_X(\mathbf{B}_1 \times \cdots \times \mathbf{B}_n) = \mathbf{P}_{X_1}(\mathbf{B}_1) \cdots \mathbf{P}_{X_n}(\mathbf{B}_n). \quad (1.28)$$

Como los recángulos de lados borelianos generan la σ -álgebra de Borel en \mathbf{R}^n , esta fórmula (1.28) es equivalente a decir que la medida \mathbf{P}_X en \mathbf{R}^n es la medida producto de $\mathbf{P}_{X_1}, \dots, \mathbf{P}_{X_n}$.

(b) \Rightarrow (c) Considerando los conjuntos $\mathbf{B}_k = (-\infty, x_k]$ ($k = 1, \dots, n$), la fórmula (1.27) se puede escribir como

$$\mathbf{P}_X((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = \mathbf{P}_{X_1}(-\infty, x_1] \cdots \mathbf{P}_{X_n}(-\infty, x_n].$$

Las siguientes dos medidas definidas en \mathbf{R}

$$\begin{aligned} \mu(\mathbf{A}) &= \mathbf{P}_X(\mathbf{A} \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]), \\ \nu(\mathbf{A}) &= \mathbf{P}_{X_1}(\mathbf{A}) \mathbf{P}_{X_2}(-\infty, x_2] \cdots \mathbf{P}_{X_n}(-\infty, x_n] \end{aligned}$$

coinciden en las semirrectas $(-\infty, x_1]$, que generan la σ -álgebra de Borel, por lo tanto son iguales. Tenemos entonces, para cualquier boreliano real \mathbf{B}_1 , que

$$\begin{aligned} \mathbf{P}_X(\mathbf{B}_1 \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]) \\ = \mathbf{P}_{X_1}(\mathbf{B}_1) \mathbf{P}_{X_2}(-\infty, x_2] \cdots \mathbf{P}_{X_n}(-\infty, x_n]. \end{aligned}$$

Un procedimiento inductivo demuestra entonces (1.28), que es equivalente a (c). \square

1.10. Esperanza matemática. Integral de Lebesgue

Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$. Vimos que las variables aleatorias son funciones medibles en el lenguaje del análisis real. Estudiamos ahora el concepto de *integral de Lebesgue* de una función medible X , denominado *esperanza matemática*, también llamado *valor esperado*, o simplemente *esperanza*, en el contexto de la probabilidad. El proceso de definición es mediante aproximación por variables aleatorias simples.

Definición 18. La esperanza matemática de una variable aleatoria simple

$$X = \sum_{k=1}^n x_k \mathbf{1}_{\mathbf{A}_k}$$

se denota $\mathbf{E}X$ y es el número real

$$\mathbf{E}X = \sum_{k=1}^n x_k \mathbf{P}(\mathbf{A}_k).$$

Observación. Las notaciones usuales para la integral de Lebesgue, que coincide con la esperanza matemática, es

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

Utilizaremos indistintamente cualquiera de las notaciones.

No es difícil ver que la definición no depende de la representación elegida para la variable simple. Se verifica la siguiente propiedad de *monotonía*.

Proposición 5. Sean $X \leq Y$ variables aleatorias simples. Luego $\mathbf{E}X \leq \mathbf{E}Y$.

Demostración. Tenemos

$$X = \sum_{k=1}^n x_k \mathbf{1}_{\mathbf{A}_k}, \quad Y = \sum_{j=1}^m y_j \mathbf{1}_{\mathbf{B}_j}.$$

Podemos representar a X, Y de la siguiente manera:

$$X = \sum_{k,j=1}^{n,m} c_{k,j} \mathbf{1}_{\mathbf{A}_k \mathbf{B}_j}, \quad Y = \sum_{k,j=1}^{n,m} d_{k,j} \mathbf{1}_{\mathbf{A}_k \mathbf{B}_j}.$$

donde $c_{k,j} = x_k$, $d_{k,j} = y_j$. Es claro que $c_{k,j} \leq d_{k,j}$. Entonces

$$\mathbf{E}X = \sum_{k,j=1}^{n,m} c_{k,j} \mathbf{P}(\mathbf{A}_k \mathbf{B}_j) \leq \sum_{k,j=1}^{n,m} d_{k,j} \mathbf{P}(\mathbf{A}_k \mathbf{B}_j) = \mathbf{E}Y.$$

concluyendo la demostración. \square

Definición 19. Dada la variable aleatoria $X: \Omega \rightarrow \mathbf{R}$, $X \geq 0$, definimos

$$\mathbf{E}X = \lim_n \mathbf{E}X_n,$$

donde $\{X_n\}$ es una sucesión de variables aleatorias simples tales que $0 \leq X_n \nearrow X$.

Lema 3. Sean $\{X_n\}, Y$ variables aleatorias simples, tales que

$$0 \leq X_n \nearrow X \geq Y.$$

Entonces

$$\lim_n \mathbf{E} X_n \geq \mathbf{E} Y.$$

Demostración. \square

Corolario 1. La definición de esperanza matemática no depende de la sucesión de variables aleatorias simples elegida, es decir, si $\{Y_n\}$ son simples, y se verifica $0 \leq Y_n \nearrow X$, entonces

$$\mathbf{E} Y_n \nearrow \mathbf{E} X.$$

Proposición 6. Sea $X \geq 0$ variable aleatoria. Tenemos

$$\mathbf{E} X = \sup\{\mathbf{E} \varphi : \varphi \text{ es variable aleatoria simple, } 0 \leq \varphi \leq X\}.$$

(donde el conjunto del cual tomamos el supremo no es vacío porque $\varphi \equiv 0$ es variable aleatoria simple).

Propiedad 5. Dadas las variables aleatorias $0 \leq X \leq Y$, tenemos

$$0 \leq \mathbf{E} X \leq \mathbf{E} Y.$$

Demostración. Es inmediata a partir de la proposición 6, dado que

$$\{\varphi \text{ simple, } 0 \leq \varphi \leq X\} \subset \{\varphi \text{ simple, } 0 \leq \varphi \leq Y\}.$$

\square

Definición 20. Decimos que la variable aleatoria $X: \mathbf{R}$ es integrable, o que tiene esperanza finita, cuando la variable aleatoria $|X|$ (que es no negativa) tiene esperanza finita, es decir, cuando

$$\mathbf{E} |X| < \infty.$$

Como $0 \leq X^+ \leq |X|$, y $0 \leq X^- \leq |X|$, si $\mathbf{E} |X|$ es finita, también lo son $\mathbf{E} (X^+)$ y $\mathbf{E} (X^-)$.

Definición 21. Dada X integrable, definimos su esperanza matemática o la integral de Lebesgue de X , mediante la fórmula

$$\int_{\Omega} X d\mathbf{P} = \mathbf{E} X := \mathbf{E} (X^+) - \mathbf{E} (X^-).$$

La propiedad 5 nos permite probar que, dado $\mathbf{A} \in \mathcal{F}$, si X es integrable, también lo es $\mathbf{1}_{\mathbf{A}}X$.

Definición 22. Dada la variable aleatoria integrable $X: \mathbf{R}$ y el suceso $\mathbf{A} \in \mathcal{F}$, definimos la integral de X en \mathbf{A} mediante

$$\int_{\mathbf{A}} X d\mathbf{P} := \int_{\Omega} \mathbf{1}_{\mathbf{A}} X d\mathbf{P}.$$

Veamos algunas propiedades de la integral recién definida.⁵

Propiedad 6. Si X es integrable, también lo es cX , y vale $\mathbf{E}(cX) = c \mathbf{E} X$.

Propiedad 7. Si $X \leq Y$ son integrables, $\mathbf{E} X \leq \mathbf{E} Y$.

Propiedad 8. Si X es integrable, tenemos $|\mathbf{E} X| \leq \mathbf{E} |X|$.

Propiedad 9. Si X, Y son itegrables, también lo es $X + Y$, y vale

$$\mathbf{E}(X + Y) = \mathbf{E} X + \mathbf{E} Y.$$

El siguiente grupo de propiedades se verifican en un conjunto de probabilidad total.

Definición 23. Diremos que dos variables aleatorias X, Y son iguales casi seguramente, cuando

$$\mathbf{P}(X = Y) = 1.$$

En ese caso escribimos $X = Y$ c.s.

Propiedad 10. Si $X = 0$ c.s. entonces $\mathbf{E} X = 0$.

Propiedad 11. Si X es integrable y vale $X = Y$ c.s., entonces Y es integrable y vale $\mathbf{E} X = \mathbf{E} Y$.

Propiedad 12. Si $X \geq 0$ y vale $\mathbf{E} X = 0$ entonces $X = 0$ c.s.

Propiedad 13. Si X, Y son integrables, y vale $\int_{\mathbf{A}} X d\mathbf{P} \leq \int_{\mathbf{A}} Y d\mathbf{P}$ para todo $\mathbf{A} \in \mathcal{F}$, entonces $X \leq Y$ c.s.

⁵Ver las demostraciones en "Probability" de A. Shiryaev, pp 183-185.

1.11. Tres teoremas de convergencia

Teorema 7 (Convergencia monótona). *Sea $\{X_n\}$ una sucesión de variables aleatorias no negativas, creciente, y convergente a X en todo $\omega \in \Omega$. Entonces $\mathbf{E} X_n \nearrow \mathbf{E} X$.*

Demostración. Como $X_n \leq X_{n+1}$, por monotonía de la esperanza, existe $\lim_n \mathbf{E} X_n = \alpha$ (que puede ser infinito). Además, como $X_n \leq X$, obtenemos $\mathbf{E} X_n \leq \mathbf{E} X$, que implica $\alpha \leq \mathbf{E} X$.

Veamos que efectivamente $\alpha = \mathbf{E} X$. Sea entonces $0 < c < 1$ y consideremos una variable aleatoria simple $\varphi \leq X$. La sucesión

$$\mathbf{A}_n = \{\omega : X_n(\omega) \geq c\varphi(\omega)\}$$

es creciente a Ω , es decir $\cup_n \mathbf{A}_n = \Omega$. Tenemos

$$\mathbf{E} X_n \geq \mathbf{E}(X_n \mathbf{1}_{\mathbf{A}_n}) \geq c \mathbf{E}(\varphi \mathbf{1}_{\mathbf{A}_n}).$$

Ejercicio 15. Demostrar que $\mathbf{E}(\varphi \mathbf{1}_{\mathbf{A}_n}) \rightarrow \mathbf{E} \varphi$. Sugerencia: aplicar la definición de función simple y la continuidad de la probabilidad.

Con el ejercicio anterior, tomando límite a ambos miembros, tenemos

$$\alpha \geq c \mathbf{E} \varphi.$$

Tomando supremo en φ , obtenemos $\alpha \geq c \mathbf{E} X$, como c es arbitrario, concluimos la demostración. \square

Corolario 2. *Sea $\{X_n\}$ una sucesión de variables aleatorias no negativas. Entonces*

$$\mathbf{E} \sum_n X_n = \sum_n \mathbf{E} X_n.$$

Teorema 8 (Lema de Fatou). *Sea $\{X_n\}$ una sucesión de variables aleatorias no negativas. Entonces*

$$\mathbf{E} \left(\liminf_n X_n \right) \leq \liminf_n \mathbf{E} X_n.$$

Demostración. Como $Z_n = \inf_{k \geq n} X_k \leq X_k$ para todo $k \geq n$, tenemos $\mathbf{E} \inf_{k \geq n} X_k \leq \mathbf{E} X_k$, de donde

$$\mathbf{E} \left(\inf_{k \geq n} X_k \right) \leq \inf_{k \geq n} \mathbf{E} X_k.$$

Ahora, Z_n es creciente, y tiene límite $\sup_n Z_n$. Aplicando convergencia monótona

$$\begin{aligned}\mathbf{E} \left(\liminf_n X_n \right) &= \mathbf{E} \left(\sup_n \inf_{k \geq n} X_k \right) = \mathbf{E} \left(\sup_n Z_n \right) = \sup_n \mathbf{E} Z_n \\ &= \sup_n \mathbf{E} \inf_{k \geq n} X_k \leq \sup_n \inf_{k \geq n} \mathbf{E} X_k = \liminf_n \mathbf{E} X_n,\end{aligned}$$

concluyendo la demostración. \square

Ejercicio 16. Encontrar un ejemplo en donde en el Lema de Fatou se verifique la desigualdad estricta.

Teorema 9 (Convergencia Dominada). *Sea $\{X_n\}$ una sucesión de variables aleatorias convergentes a X . Supongamos que existe Y integrable, tal que $|X_n| \leq Y$ para todo $\omega \in \Omega$. Entonces*

$$(a) \lim_n \mathbf{E} X_n = \mathbf{E} X, \quad (b) \lim_n \mathbf{E} |X_n - X| = 0.$$

Demostración. Observemos primero que (b) implica (a), ya que

$$|\mathbf{E} X_n - \mathbf{E} X| = |\mathbf{E}(X_n - X)| \leq \mathbf{E} |X_n - X|.$$

Para ver (b), observamos que $|X| \leq Y$, por lo que X es integrable, y además $2Y - |X_n - X|$ es no negativa, y tiene límite $2Y$. Apliquemos entonces el Lema de Fatou:

$$\begin{aligned}\mathbf{E}(2Y) &= \mathbf{E} \liminf_n (2Y - |X_n - X|) \leq \liminf_n \mathbf{E}(2Y - |X_n - X|) \\ &= \mathbf{E}(2Y) + \liminf_n (-\mathbf{E} |X_n - X|) = \mathbf{E}(2Y) - \limsup_n \mathbf{E} |X_n - X|.\end{aligned}$$

De aquí

$$\limsup_n \mathbf{E} |X_n - X| = 0.$$

Como $\mathbf{E} |X_n - X| \geq 0$, deducimos (b), concluyendo la demostración. \square

1.12. Cambio de variable en la integral de Lebesgue

Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$, y una variable aleatoria X con distribución de probabilidad \mathbf{P}_X y función de distribución F_X .

Teorema 10 (Cambio de variable). (a) *Para cualquier función no negativa $g: \mathbf{R} \rightarrow \mathbf{R}$ tenemos la identidad*

$$\mathbf{E} g(X) = \int_{\Omega} g(X(\omega)) d\mathbf{P} = \int_{\mathbf{R}} g(x) d\mathbf{P}_X. \quad (1.29)$$

(b) *Si la variable aleatoria $g(X)$ es integrable, vale también la fórmula (1.29).*

Demostración. Veamos la fórmula (1.29) para características de conjuntos en \mathbf{R} :

$$\int_{\mathbf{R}} \mathbf{1}_{\mathbf{A}}(x) d\mathbf{P}_X = \mathbf{P}_X(\mathbf{A}) = \mathbf{P}(X \in \mathbf{A}) = \int_{\Omega} \mathbf{1}_{\mathbf{A}}(X(\omega)) d\mathbf{P}.$$

La linealidad de la integral nos permite obtener (1.29) para funciones simples en \mathbf{R} de la forma $g(x) = \sum_{k=1}^n x_k \mathbf{1}_{\mathbf{A}_k}(x)$. A su vez, la función medible y no negativa $g(x)$ se aproxima por una sucesión creciente de funciones simples. El teorema de convergencia monótona nos permite concluir la demostración de (a). El enunciado (b) se obtiene tomando la parte positiva y la negativa de $g(X)$ y aplicando (a). \square

Definición 24. *Dada una función de distribución F en \mathbf{R} , la integral de una función medible g con respecto a la medida construida en \mathbf{R} a partir de F se denomina integral de Lebesgue-Stieltjes con respecto a F , y se designa*

$$\int_{\mathbf{R}} g(x) dF(x) = \int_{\mathbf{R}} g(x) d\mathbf{P}_X.$$

Con esta nueva definición, el teorema anterior se escribe

$$\int_{\Omega} g(X(\omega)) d\mathbf{P} = \int_{\mathbf{R}} g(x) dF_X(x).$$

Recordemos que, a partir de la función $F(x) = x$ se construye la medida de Lebesgue, que denotamos m . La teoría de la integración con respecto a la medida de Lebesgue es análoga a la realizada para medidas finitas. En particular valen los tres teoremas de convergencia. Tenemos así una nueva definición de variables aleatorias absolutamente continuas:

Definición 25. *La variable aleatoria X (o su función de distribución F_X) se dice absolutamente continua, cuando existe una función medible $f_X: \mathbf{R} \rightarrow [0, \infty)$, que verifica $\int_{-\infty}^{\infty} f(x) dx = 1$, y tal que*

$$F_X(x) = \int_{\mathbf{R}} \mathbf{1}_{(-\infty, x]} f_X(t) m(dt).$$

La función f_X se llama densidad de la variable aleatoria X .

Como la medida de Lebesgue no asigna medida a los puntos, es decir, $m(\{x\}) = 0$, tenemos

$$\int_{\mathbf{R}} \mathbf{1}_{(-\infty, x]} f_X(t) m(dt) = \int_{\mathbf{R}} \mathbf{1}_{(-\infty, x)} f_X(t) m(dt)$$

y utilizamos la notación

$$F_X(x) = \int_{-\infty}^x f_X(t) m(dt). \quad (1.30)$$

Proposición 7. (a) *Dadas una variable aleatoria X con densidad f_X , y una función medible $g(x)$ tal que $g(X)$ es integrable, se cumple*

$$\mathbf{E} g(X) = \int_{\Omega} g(X(\omega)) d\mathbf{P} = \int_{\mathbf{R}} g(x) f_X(x) m(dx). \quad (1.31)$$

(b) *Cuando la variable aleatoria X es discreta, y toma los valores en el conjunto $\{x_1, x_2, \dots\}$ con probabilidades $\{p_1, p_2, \dots\}$, entonces*

$$\mathbf{E} g(X) = \int_{\Omega} g(X(\omega)) d\mathbf{P} = \sum_n g(x_n) p_n. \quad (1.32)$$

Demostración. Comenzamos por (a). Si $g(x) = \mathbf{1}_{(a,b]}$, la variable aleatoria $g(X)$ es simple, y

$$\mathbf{E} g(X) = \mathbf{P}_X(a, b] = F_X(b) - F_X(a) = \int_a^b f_X(x) m(dx) = \int_{\mathbf{R}} g(x) f_X(x) m(dx),$$

donde aplicamos (1.30). Esto quiere decir que la igualdad anterior vale para conjuntos borelianos en \mathbf{R} , por unicidad de la extensión de medidas. Por linealidad lo extendemos a funciones simples, y por convergencia monótona a funciones no negativas, pasando al caso general considerando $g(X)^+$ y $g(X)^-$.

En el caso (b), la variable aleatoria X se representa como

$$X = \sum_{n=1}^{\infty} x_n \mathbf{1}_{\mathbf{A}_n},$$

donde $p_n = \mathbf{P}(\mathbf{A}_n)$. Si la variable X toma una cantidad finita de valores, lo mismo ocurre con $g(X)$, y la fórmula (1.32) es la definición de integral de

una variable aleatoria simple. En caso de que toma una cantidad numerable de valores, tenemos

$$X_N = \sum_{n=1}^N x_n \mathbf{1}_{\mathbf{A}_n} \nearrow \sum_{n=1}^{\infty} x_n \mathbf{1}_{\mathbf{A}_n}.$$

Si $g(x) \geq 0$, tenemos $g(X_N) \nearrow g(X)$, y aplicando convergencia monótona, tenemos

$$\mathbf{E} g(X) = \lim_N \mathbf{E} g(X_N) = \lim_N \sum_{n=1}^N g(x_n) p_n = \sum_{n=1}^{\infty} g(x_n) p_n.$$

La demostración concluye tomando parte positiva y negativa de $g(X)$. \square

Capítulo 2

Leyes de los grandes números

2.1. Ley débil de los grandes números

Se denomina *ley de los grandes números* a cualquier proposición que establece, bajo determinadas condiciones, la convergencia en probabilidad o casi segura de los promedios aritméticos de una cantidad creciente a infinito de variables aleatorias. Si tenemos convergencia en probabilidad, decimos *ley débil de los grandes números*, si tenemos convergencia casi segura, *ley fuerte de los grandes números*. En esta sección, estudiaremos diversas formas de la ley débil de los grandes números.

Teorema 11 (Chebishev). *Consideremos una sucesión de variables aleatorias X_1, X_2, \dots , independientes dos a dos, y con esperanzas matemáticas a_1, a_2, \dots . Supongamos que se cumple la condición*

$$\text{var } X_n \leq C \text{ para cada } n = 1, 2, \dots,$$

donde C es una constante arbitraria. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i \xrightarrow{\mathbf{P}} 0. \quad (2.1)$$

Es importante el caso particular en el que las esperanzas de las variables aleatorias son todas iguales, digamos $\mathbf{E} X_n = a$ para todo $n = 1, 2, \dots$. En esta situación tenemos $\sum_{i=1}^n a_i/n = a$, y la convergencia en (2.1) se transforma en

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} a.$$

Demostración. Consideremos para cada $n = 1, 2, \dots$ la variable aleatoria $Z_n = \sum_{i=1}^n X_i/n$. Es claro que $\mathbf{E} Z_n = \sum_{i=1}^n a_i/n$. Como las variables aleatorias son independientes dos a dos, tenemos $\mathbf{var} \sum_{i=1}^n X_i = \sum_{i=1}^n \mathbf{var} X_i \leq nC$, Aplicamos ahora la desigualdad de Chebishev, para obtener

$$\begin{aligned} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i \right| \geq \varepsilon \right) &= \mathbf{P} (|Z_n - \mathbf{E} Z_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{var} Z_n \\ &= \frac{1}{n^2 \varepsilon^2} \mathbf{var} \sum_{i=1}^n X_i \rightarrow 0, \end{aligned} \quad (2.2)$$

si $n \rightarrow \infty$, dado que $\sum_{i=1}^n \mathbf{var} X_i \leq nC$. Esto concluye la demostración. \square

El teorema de Chebishev da fundamento a la *regla de la media aritmética*, utilizada en el procesamiento de resultados de mediciones: para la estimación del valor de una constante física a , desconocida, mediante los resultados de n mediciones de su magnitud, se recomienda tomar el promedio aritmético de estas mediciones. Veamos la fundamentación. Supongamos que X_1, \dots, X_n son los resultados de las n mediciones de esta constante a . Por cuanto las mediciones habitualmente se acompañan de errores, y no podemos predecir el resultado de la medición siguiente, consideramos que el resultado de la i -ésima ($i = 1, \dots, n$) medición de la constante a es una variable aleatoria $X_i = a + \Delta_i$, donde Δ_i es el error que se comete en la i -ésima medición. Suponemos que

$$\mathbf{E} \Delta_i = 0, \quad \mathbf{var} \Delta_i = \sigma^2 \quad (i = 1, \dots, n), \quad (2.3)$$

donde el valor de σ^2 puede ser desconocido. Las condiciones en (2.3) son equivalentes a la condiciones

$$\mathbf{E} X_i = a, \quad \mathbf{var} X_i = \sigma^2 \quad (i = 1, \dots, n). \quad (2.4)$$

La primer condición en (2.4) se interpreta como la ausencia de error sistemático en las mediciones, la segunda significa que todas las mediciones se llevan a cabo con la misma precisión. Si se supone además que las variables aleatorias X_1, \dots, X_n son independientes dos a dos, es aplicable el teorema de Chebishev con $a_1 = \dots = a_n = a$ y la constante $C = \sigma^2$, de acuerdo al cual obtenemos

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} a.$$

En consecuencia, el promedio aritmético de los resultados de las mediciones de una constante física a converge en probabilidad al valor de esta constante;

en la terminología de la estadística matemática se dice que el promedio aritmético de las mediciones es un *estimador consistente* de la constante desconocida a .

Veamos otro corolario del teorema de Chebishev, demostrando, que de este teorema se deduce el teorema de Bernoulli de la sección 2.4.

Consideremos entonces una serie de n experimentos independientes, con dos resultados posibles cada uno (éxito y fracaso), y probabilidad de éxito igual a p en cada experimento ($0 < p < 1$). Sea μ la cantidad de éxitos en n experimentos. Veamos como se formula el teorema de Bernoulli por medio de variables aleatorias. Para ésto, sea X_1, X_2, \dots una sucesión de variables aleatorias, cada una de las cuales toma el valor 1 con probabilidad p (si ocurre un éxito) y el valor 0 con probabilidad $q = 1 - p$ (si ocurre un fracaso). Tenemos $\mathbf{E} X_i = p$, $\mathbf{var} X_i = pq \leq 1$ para cada $i = 1, 2, \dots$. Además $\mu = \sum_{i=1}^n X_i$, porque la suma contiene tantos sumandos iguales a uno como éxitos ocurren en los primeros n experimentos, siendo nulos los sumandos restantes. Las variables aleatorias X_1, X_2, \dots son independientes dos a dos y cumplen la condición $\mathbf{var} X_i \leq 1$ ($i = 1, 2, \dots$), por lo que es aplicable el teorema de Chebishev, y de (2.1) se obtiene $\mu/n \xrightarrow{\mathbf{P}} p$. La última afirmación significa que la frecuencia de éxitos en n experimentos, converge en probabilidad a p , la probabilidad de éxito en un experimento.

Teorema 12 (Markov). *Consideremos una sucesión de variables aleatorias X_1, X_2, \dots con esperanzas matemáticas a_1, a_2, \dots . Supongamos que se cumple la condición*

$$\frac{1}{n^2} \mathbf{var} \sum_{i=1}^n X_i \rightarrow 0, \quad (2.5)$$

si $n \rightarrow \infty$. Entonces tiene lugar la convergencia en (2.1), es decir

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i \xrightarrow{\mathbf{P}} 0.$$

Demostración. Al igual que en la demostración del teorema de Chebishev, aplicamos la desigualdad de Chebishev con la misma elección de la variable aleatoria Z_n . Tenemos, como antes, $\mathbf{E} Z_n = \sum_{i=1}^n a_i/n$. Por la condición (2.5), tenemos

$$\mathbf{var} Z_n = \frac{1}{n^2} \mathbf{var} \sum_{i=1}^n X_i \rightarrow 0 \text{ si } n \rightarrow \infty.$$

Esto muestra que la acotación y el límite en la fórmula (2.2) tienen lugar, concluyendo la demostración. \square

El teorema de Markov es una generalización del teorema de Chebishev. La tesis en ambos teoremas es la misma, pero la hipótesis en el teorema de Markov es más general que en teorema de Chebishev. Si se cumplen las hipótesis del teorema de Chebishev tenemos $\text{var} \sum_{i=1}^n X_i = \sum_{i=1}^n \text{var} X_i \leq nC$, dada la independencia dos a dos de las variables aleatorias X_1, X_2, \dots y la condición (11), por lo que la condición (2.5) se cumple de manera evidente.

Es importante observar que en el teorema de Markov no figura supuesto alguno sobre la *independencia* de las variables aleatorias consideradas. Ésto permite su aplicación para la demostración de la ley de los grandes números (es decir, la demostración de la afirmación (2.1)) para sucesiones de variables aleatorias dependientes.

Veamos como obtener, con la ayuda del teorema de Markov, una ley de los grandes números para una sucesión estacionaria.

Decimos que una sucesión de variables aleatorias X_1, X_2, \dots es una *sucesión estacionaria* (más precisamente, una *sucesión débilmente estacionaria*), si la esperanza de las variables aleatorias $\mathbf{E} X_n$ es constante para todo n , existen los momentos de segundo orden para todo n , y además, la esperanza $\mathbf{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$.

Ejemplo 12. Consideremos una sucesión de variables aleatorias $\{X_n\}$ no correlacionadas dos a dos, y tal que $\mathbf{E} X_n = 0$, $\text{var} X_n = 1$ para todo $n = 1, 2, \dots$. Veamos que es estacionaria.

Sea $\rho(X_n, X_m)$ el coeficiente de correlación entre X_n y X_m , es decir

$$\rho(X_n, X_m) = \frac{\mathbf{E}(X_n X_m) - \mathbf{E} X_n \mathbf{E} X_m}{\sqrt{\text{var} X_n} \sqrt{\text{var} X_m}}.$$

En vista de nuestros supuestos $\rho(X_n, X_m) = \mathbf{E}(X_n X_m) = 0$ si $n \neq m$, y en el caso $m = n$ obtenemos $\rho(X_n, X_n) = \mathbf{E} X_n^2 = 1$. En consecuencia

$$\mathbf{E}(X_n X_m) = \begin{cases} 0, & \text{si } n - m \neq 0, \\ 1, & \text{si } n - m = 0, \end{cases}$$

es decir, la esperanza $\mathbf{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$. De aquí concluimos que la sucesión de variables aleatorias considerada es estacionaria.

Ejemplo 13. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuídas. Supongamos que existen $\mathbf{E} X_n = a$ y $\text{var} X_n = \sigma^2$. Esta sucesión de variables aleatorias es estacionaria.

En efecto, si $m = n$, tenemos $\mathbf{E} X_n^2 = \text{var} X_n + (\mathbf{E} X_n)^2 = \sigma^2 + a^2$. Si $m \neq n$, como la correlación es nula, es nula la covarianza, y tenemos

$\mathbf{E} X_n X_m = \mathbf{E} X_n \mathbf{E} X_m = a^2$ En conclusión

$$\mathbf{E}(X_n X_m) = \begin{cases} a^2, & \text{si } n - m \neq 0, \\ \sigma^2 + a^2, & \text{si } n - m = 0, \end{cases}$$

y $\mathbf{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$.

Teorema 13. *Consideremos una sucesión estacionaria de variables aleatorias X_1, X_2, \dots , con $\mathbf{E} X_n = a$. Supongamos que $\rho(X_n, X_m) \rightarrow 0$ cuando $|n - m| \rightarrow \infty$, donde $\rho(X_n, X_m)$ es el coeficiente de correlación entre las variables aleatorias X_n y X_m . Entonces*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} a.$$

Demostración. Es suficiente demostrar que nuestras hipótesis implican la condición (2.5) del teorema de Markov.

Como la sucesión es estacionaria $\mathbf{E} X_n^2$ no depende de n , por lo que $\text{var } X_n = \mathbf{E} X_n^2 - (\mathbf{E} X_n)^2 = \mathbf{E} X_n^2 - a^2$ tampoco depende de n . Pongamos entonces $\sigma^2 = \text{var } X_n$. Por las propiedades de la covarianza, tenemos

$$\begin{aligned} \text{var} \sum_{i=1}^n X_i &= \sum_{i=1}^n \text{var } X_i + 2 \sum_{1 \leq i < k \leq n} \text{cov}(X_i, X_k) \\ &= n\sigma^2 + 2\sigma^2 \sum_{1 \leq i < k \leq n} \rho(X_i, X_k) = n\sigma^2 + 2\sigma^2(T_1 + T_2) \end{aligned}$$

donde

$$T_1 = \sum_{\substack{1 \leq i < k \leq n \\ |i-k| < M}} \rho(X_i, X_k), \quad T_2 = \sum_{\substack{1 \leq i < k \leq n \\ |i-k| \geq M}} \rho(X_i, X_k),$$

y M es una constante positiva.

Sea $\varepsilon > 0$ arbitrario. Existe M tal que se verifica la desigualdad $|\rho(X_i, X_k)| < \varepsilon$ si $|i-k| \geq M$, dado que $|\rho(X_i, X_k)| \rightarrow 0$ si $|i-k| \rightarrow \infty$. Entonces $|T_2| \leq \varepsilon n^2$. Para acotar la suma T_1 aplicamos la desigualdad $|\rho(X_n, X_m)| \leq 1$, y obtenemos $|T_1| \leq (2M + 1)n$. En conclusión,

$$\text{var} \sum_{i=1}^n X_i \leq n\sigma^2 + 2\sigma^2((2M + 1)n + \varepsilon n^2)$$

de donde, dividiendo por n^2 , obtenemos

$$\frac{1}{n^2} \text{var} \sum_{i=1}^n X_i \leq \frac{\sigma^2}{n} + 2\sigma^2 \left(\frac{2M + 1}{n} + \varepsilon \right).$$

Como ε es arbitrario, esta última acotación conduce a la condición (2.5) en el teorema de Markov, lo que concluye la demostración. \square

Aplicando la desigualdad de Chebishev es posible obtener una demostración del *teorema de Weierstrass* del análisis real, de acuerdo al cual, para cualquier función continua $f(x)$ definida en el intervalo $[0, 1]$, existe una sucesión de polinomios $\{P_n(x)\}$ tal que $P_n(x) \rightarrow f(x)$ uniformemente en el intervalo $[0, 1]$, cuando $n \rightarrow \infty$.

Teorema 14 (Bernstein). *Consideremos una función continua $f(x)$ definida en el intervalo $[0, 1]$. Sea*

$$B_n(x; f) = \sum_{k=0}^n f(k/n) \binom{n}{k} x^k (1-x)^{n-k}.$$

Entonces $B_n(x; f) \rightarrow f(x)$ uniformemente, en el intervalo $[0, 1]$, cuando $n \rightarrow \infty$. Las funciones $B_n(x; f)$ son polinomios de grado n , que se denominan polinomios de Bernstein.

Demostración. Consideremos la diferencia

$$f(x) - B_n(x; f) = \sum_{k=0}^n [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k} = T_1 + T_2,$$

donde

$$T_1 = \sum_{k: |\frac{k}{n} - x| < \delta} [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k},$$

$$T_2 = \sum_{k: |\frac{k}{n} - x| \geq \delta} [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k}.$$

Primero obtenemos una acotación para el término T_1 . Como $f(x)$ es una función continua en el intervalo $[0, 1]$, dado $\varepsilon > 0$ existe $\delta > 0$ tal que $|x - y| < \delta$ implica $|f(x) - f(y)| < \varepsilon$. Entonces

$$|T_1| \leq \sum_{k: |\frac{k}{n} - x| < \delta} |f(x) - f(k/n)| \binom{n}{k} x^k (1-x)^{n-k}$$

$$\leq \varepsilon \sum_{k: |\frac{k}{n} - x| < \delta} \binom{n}{k} x^k (1-x)^{n-k} \leq \varepsilon (x + (1-x))^n = \varepsilon. \quad (2.6)$$

La acotación del término T_2 , se obtiene aplicando la desigualdad de Chebishev en un esquema de Bernoulli. Consideremos entonces una serie de n experimentos independientes con dos resultados posibles cada uno: éxito con probabilidad x , y fracaso con probabilidad $1 - x$. Si μ designa la cantidad de éxitos en n experimentos, dado $\delta > 0$, tenemos

$$\mathbf{P}\left(\left|\frac{\mu}{n} - x\right| \geq \delta\right) = \sum_{k: \left|\frac{k}{n} - x\right| \geq \delta} \binom{n}{k} x^k (1-x)^{n-k}.$$

Como $\mathbf{E}(\mu/n) = x$, $\mathbf{var}(\mu/n) = x(1-x)/n$, y $|f(x)| \leq C$ donde C es una constante, aplicando la desigualdad de Chebishev, tenemos

$$\begin{aligned} |T_2| &\leq 2C \sum_{k: \left|\frac{k}{n} - x\right| \geq \delta} \binom{n}{k} x^k (1-x)^{n-k} \\ &= 2C \mathbf{P}\left(\left|\frac{\mu}{n} - x\right| \geq \delta\right) \leq 2C \frac{x(1-x)}{n\delta^2} \leq \frac{C}{2n\delta^2} < \varepsilon \end{aligned} \quad (2.7)$$

(donde usamos que $x(1-x) \leq 1/4$), si n es suficientemente grande. En conclusión, como las acotaciones en (2.6) y (2.7) son independientes de x tenemos que, para n suficientemente grande, vale

$$\sup_{0 \leq x \leq 1} |f(x) - B_n(x; f)| \leq |T_1| + |T_2| < 2\varepsilon,$$

concluyendo la demostración. \square

Como conclusión de esta sección presentamos diferentes condiciones para la validez de la ley débil de los grandes números. Estamos particularmente interesados en debilitar las condiciones en los momentos de las variables aleatorias (como por ejemplo la condición (11) en el teorema de Chebishev, o la condición (2.5) en el teorema de Markov), mas aún en obtener resultados sin condiciones de momentos. Con este fin asumiremos, por una parte, que las variables aleatorias X_1, X_2, \dots están *idénticamente distribuidas*, y por otra que son independientes dos a dos.

Teorema 15. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuidas, con función de distribución $V(x)$. Supongamos que se verifican las dos siguientes condiciones:*

$$n \mathbf{P}(|X_1| \geq n) \rightarrow 0, \quad (2.8)$$

$$\int_{\{|x| < n\}} x dV(x) \rightarrow 0, \quad (2.9)$$

cuando $n \rightarrow \infty$. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} 0. \quad (2.10)$$

Antes de comenzar la demostración observemos que, si la condición de independencia dos a dos se sustituye por la condición de independencia mutua de las variables X_1, \dots, X_n , para cada $n = 1, 2, \dots$, entonces las condiciones (2.8) y (2.9) no solo son suficientes sino también necesarias. Este último resultado es consecuencia de un teorema más general, obtenido por Kolmogorov, para sucesiones de variables aleatorias independientes, no necesariamente idénticamente distribuidas, en el que no se asume ninguna condición de existencia de momentos de estas variables aleatorias¹.

Demostración. Consideremos para cada $n = 1, 2, \dots$ las variables aleatorias $X_{ni} = X_i/n$ ($1 \leq i \leq n$). El límite en (2.10), en términos de estas nuevas variables, es

$$\sum_{i=1}^n X_{ni} \xrightarrow{\mathbf{P}} 0. \quad (2.11)$$

Consideremos ahora las variables aleatorias truncadas, definidas mediante

$$\bar{X}_{ni} = \begin{cases} X_{ni}, & \text{si } |X_{ni}| < 1, \\ 0, & \text{si } |X_{ni}| \geq 1. \end{cases}$$

Estas variables aleatorias están acotadas, por lo que existen sus momentos de cualquier orden. Para calcular los dos primeros momentos, observamos que $\bar{X}_{ni} = g(X_i/n)$, donde $g(x) = x \mathbf{1}_{\{|x| < 1\}}$. Aplicando entonces la identidad (??), tenemos

$$\mathbf{E} \bar{X}_{ni} = \mathbf{E} g(X_i/n) = \int_{-\infty}^{\infty} g(y/n) dV(y) = \frac{1}{n} \int_{\{|y| < n\}} y dV(y), \quad (2.12)$$

$$\mathbf{E} \bar{X}_{ni}^2 = \int_{-\infty}^{\infty} (g(y/n))^2 dV(y) = \frac{1}{n^2} \int_{\{|y| < n\}} y^2 dV(y). \quad (2.13)$$

Demostremos ahora que de la condición (2.8), se obtiene que

$$\frac{1}{n} \int_{\{|x| < n\}} x^2 dV(x) \rightarrow 0 \quad (n \rightarrow \infty). \quad (2.14)$$

En efecto, recordando que $k^2 \leq k(k+1) = 2 \sum_{j=1}^k j$, tenemos

$$\int_{\{|x| < n\}} x^2 dV(x) = \sum_{k=1}^n \int_{\{k-1 \leq |x| < k\}} x^2 dV(x)$$

¹Ver por ejemplo §4.4. en V.V. Petrov, Limit Theorems of Probability Theory, (1995)

$$\begin{aligned}
&\leq \sum_{k=1}^n k^2 \mathbf{P}(k-1 \leq |X_1| < k) \\
&\leq \sum_{k=1}^n (2 \sum_{j=1}^k j) \mathbf{P}(k-1 \leq |X_1| < k) = \\
&= 2 \sum_{j=1}^n j \mathbf{P}(j-1 \leq |X_1| < n) \leq 2 \sum_{j=1}^n j \mathbf{P}(j-1 \leq |X_1|).
\end{aligned}$$

Es sencillo demostrar que si $\{b_j\}$ es una sucesión numérica, tal que $b_j \rightarrow 0$ ($j \rightarrow \infty$), entonces $(\sum_{j=1}^n b_j)/n \rightarrow 0$ ($n \rightarrow \infty$). Si ponemos $b_j = j \mathbf{P}(j-1 \leq |X_1|)$, tenemos

$$b_j = j \mathbf{P}(j-1 \leq |X_1|) = \frac{j}{j-1} (j-1) \mathbf{P}(j-1 \leq |X_1|) \rightarrow 0 \quad (j \rightarrow \infty),$$

en vista de (2.8). Por ésto

$$\frac{1}{n} \int_{\{|x|<n\}} x^2 dV(x) \leq \frac{2}{n} \sum_{j=1}^n b_j \rightarrow 0 \quad (n \rightarrow \infty),$$

y obtenemos (2.14).

Aplicamos ahora la desigualdad de Chebishev. Teniendo en cuenta (2.13) y (2.14), obtenemos

$$\begin{aligned}
\mathbf{P} \left(\left| \sum_{i=1}^n \bar{X}_{ni} - \sum_{i=1}^n \mathbf{E} \bar{X}_{ni} \right| \geq \delta \right) &\leq \frac{1}{\delta^2} \mathbf{var} \sum_{i=1}^n \bar{X}_{ni} = \frac{1}{\delta^2} \sum_{i=1}^n \mathbf{var} \bar{X}_{ni} \\
&\leq \frac{1}{\delta^2} \sum_{i=1}^n \mathbf{E} \bar{X}_{ni}^2 = \frac{1}{\delta^2 n} \int_{\{|x|<n\}} x^2 dV(x) \rightarrow 0,
\end{aligned}$$

para $\delta > 0$ cuando $n \rightarrow \infty$. En los cálculos anteriores utilizamos la propiedades de variables aleatorias independientes dos a dos. De esta forma, obtenemos

$$\sum_{i=1}^n \bar{X}_{ni} - \sum_{i=1}^n \mathbf{E} \bar{X}_{ni} \xrightarrow{\mathbf{P}} 0.$$

En vista de la fórmula (2.12) y la condición (2.9), tenemos

$$\sum_{i=1}^n \mathbf{E} \bar{X}_{ni} = \int_{\{|x|<n\}} x dV(x) \rightarrow 0 \quad (n \rightarrow \infty).$$

Por esto concluimos que $\sum_{i=1}^n \bar{X}_{ni} \xrightarrow{\mathbf{P}} 0$. Resta entonces demostrar, que

$$\sum_{i=1}^n X_{ni} - \sum_{i=1}^n \bar{X}_{ni} \xrightarrow{\mathbf{P}} 0,$$

de donde se obtiene (2.11).

Veamos esta última condición. Para todo $\delta > 0$ tenemos

$$\begin{aligned} \mathbf{P}\left(\left|\sum_{i=1}^n X_{ni} - \sum_{i=1}^n \bar{X}_{ni}\right| \geq \delta\right) &\leq \mathbf{P}\left(\sum_{i=1}^n X_{ni} \neq \sum_{i=1}^n \bar{X}_{ni}\right) \\ &\leq \mathbf{P}\left(\bigcup_{i=1}^n \{X_{ni} \neq \bar{X}_{ni}\}\right) \leq \sum_{i=1}^n \mathbf{P}(X_{ni} \neq \bar{X}_{ni}) = \sum_{i=1}^n \mathbf{P}(|X_{ni}| \geq 1) \\ &= \sum_{i=1}^n \mathbf{P}(|X_i| \geq n) = n \mathbf{P}(|X_1| \geq n) \rightarrow 0 \end{aligned}$$

por la condición (2.8). Queda demostrado entonces (2.10). \square

Veamos ahora un corolario del teorema 15, de sencilla formulación.

Teorema 16. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuidas. Supongamos que existe la esperanza matemática $\mathbf{E} X_1$. Entonces*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} \mathbf{E} X_1.$$

Demostración. Demostremos primero el teorema 16 en el caso particular en el que $\mathbf{E} X_1 = 0$, verificando las condiciones en el teorema 15. Sea $V(x)$ la función de distribución de la variable aleatoria X_1 . La condición (2.9) se cumple en forma evidente. Para verificar la condición (2.8), veamos que

$$\mathbf{P}(|X_1| \geq n) = \int_{\{|x| \geq n\}} dV(x) \leq \frac{1}{n} \int_{\{|x| \geq n\}} |x| dV(x),$$

por ésto,

$$n \mathbf{P}(|X_1| \geq n) \leq \int_{\{|x| \geq n\}} |x| dV(x) \rightarrow 0$$

cuando $n \rightarrow \infty$, porque $\int_{-\infty}^{\infty} |x| dV(x) = \mathbf{E} |X_1| < \infty$. De esta forma se verifican todas las hipótesis del teorema 15, concluyendo la demostración en este caso particular.

Supongamos ahora que $\mathbf{E} X_1 = a \neq 0$. Consideremos la sucesión de variables aleatorias $\{Y_n\}$, donde $Y_n = X_n - a$ ($n = 1, 2, \dots$). Como $\{X_n\}$ es una sucesión de variables aleatorias independientes dos a dos, también lo es la sucesión $\{Y_n\}$. Además $\mathbf{E} Y_n = \mathbf{E} X_n - a = 0$. Según demostramos en la primera parte, tiene lugar la convergencia, $\sum_{i=1}^n Y_i/n \xrightarrow{\mathbf{P}} 0$, que claramente equivale a $\sum_{i=1}^n X_i/n \xrightarrow{\mathbf{P}} a$, lo que concluye la demostración. \square

El teorema 16 fue obtenido por A. Ya. Jínchin, bajo la hipótesis más restrictiva de independencia mutua, en vez de independencia dos a dos, de las variables aleatorias X_1, \dots, X_n .

En el caso en que las variables aleatorias sean idénticamente distribuidas, el teorema de Chebishev conduce a un resultado mas débil que el teorema 16, dado que en el teorema de Chebishev se exige la existencia de varianzas, y no únicamente la existencia de esperanzas matemáticas.

2.2. Ejercicios de Leyes débiles de los grandes números

17. Sea $\{X_n\}$ una sucesión de variables aleatorias, independientes dos a dos, tales que

$$\mathbf{P}(X_n = 2^n) = \mathbf{P}(X_n = -2^n) = 2^{-2n-1}, \quad \mathbf{P}(X_n = 0) = 1 - 2^{-2n},$$

para cada n . ¿Es aplicable la ley de los grandes números a esta sucesión?

18. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes dos a dos, tales que

$$\mathbf{P}(X_n = n^{1/4}) = \mathbf{P}(X_n = -n^{1/4}) = \mathbf{P}(X_n = 0) = 1/3,$$

para cada n . Demostrar que para esta sucesión es aplicable la ley de los grandes números.

19. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tales que

$$\mathbf{P}(X_n = n^\gamma) = \mathbf{P}(X_n = -n^\gamma) = 1/2,$$

para cada n , donde $\gamma < 1/2$. ¿Es aplicable la ley de los grandes números a esta sucesión?

20. Sea $\{X_n\}$ una sucesión de variables aleatorias, tal que $\mathbf{E} X_n = 0$, $\mathbf{E} |X_n| = 1/n$ para cada n . Demostrar que para esta sucesión es aplicable la ley de los grandes números.

21. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tales que

$$\mathbf{P}(X_n = n) = \mathbf{P}(X_n = -n) = 1/2$$

para cada n . ¿Es aplicable la ley de los grandes números a esta sucesión?

2.3. Ley fuerte de los grandes números

Es la sección 3 del capítulo IV de la segunda edición del libro “Probability” de A.N. Shiryaev (Springer, 1996), páginas 388-395.

2.4. Ejercicios de Ley fuerte de los grandes números

1. Probar que $\mathbf{E}(X^2) < \infty$ si y solo si $\sum_{n=1}^{\infty} n \mathbf{P}(|X| \geq n)$ converge.
2. Suponiendo que X_1, X_2, \dots son independientes e idénticamente distribuidas, mostrar que si $\mathbf{E}|X_1|^\alpha < \infty$ para algún $0 < \alpha < 1$, entonces, con $S_n = X_1 + \dots + X_n$, tenemos

$$\frac{S_n}{n^{1/\alpha}} \rightarrow 0 \text{ (P-c.s.)},$$

y si $\mathbf{E}|X_1|^\beta < \infty$ para algún $1 < \beta < 2$, entonces

$$\frac{S_n - n \mathbf{E} X_1}{n^{1/\beta}} \rightarrow 0 \text{ (P-c.s.)}$$

3. Construir una sucesión de variables aleatorias que verifique la ley débil de los grandes números pero que no verifique la ley fuerte de los grandes números.
4. Utilizar el desarrollo del Ejemplo 2: El método de Monte Carlo (pg. 394) para dar una aproximación de π .

Capítulo 3

Funciones características

Una función característica es una función que toma valores complejos y tiene argumento real. Definida a partir de una variable aleatoria X , *caracteriza* la distribución $F(x)$ de esta variable aleatoria. Las funciones características son especialmente adecuadas para el estudio de la convergencia débil de variables aleatorias independientes, y serán utilizadas para demostrar teoremas centrales del límite. Dadas dos variables aleatorias X e Y , definimos la *variable aleatoria compleja* $Z = X + iY$, donde $i = \sqrt{-1}$ es la unidad imaginaria. Si las variables aleatorias X, Y tienen esperanzas respectivas $\mathbf{E}X, \mathbf{E}Y$, definimos la esperanza matemática de la variable aleatoria compleja Z mediante $\mathbf{E}Z = \mathbf{E}X + i\mathbf{E}Y$. No es difícil verificar (tomando partes real e imaginaria), que si a, b son dos números complejos, se tiene $\mathbf{E}(aZ + b) = a\mathbf{E}Z + b$; y que si Z_1, Z_2 son dos variables aleatorias complejas, se tiene $\mathbf{E}(Z_1 + Z_2) = \mathbf{E}Z_1 + \mathbf{E}Z_2$. Si $z = a + ib$ es un número complejo, designamos $\bar{z} = a - ib$ el *complejo conjugado* de z .

3.1. Definiciones y primeras propiedades

Consideremos una variable aleatoria X definida en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$. Llamamos *función característica* de la variable aleatoria X a la función $f(t)$, definida para todo t real, mediante la igualdad

$$f(t) = \mathbf{E}e^{itX}. \quad (3.1)$$

La fórmula (3.1) es equivalente a

$$f(t) = \mathbf{E} \cos tX + i \mathbf{E} \operatorname{sen} tX. \quad (3.2)$$

Como las variables aleatorias $\cos tX$ y $\sin tX$ están acotadas para todo t real, sus esperanzas matemáticas existen. Por ésto, la función característica de una variable aleatoria arbitraria X está correctamente definida para todo t real.

De acuerdo a la definición de esperanza matemática, tenemos

$$f(t) = \int_{\Omega} e^{itX} d\mathbf{P}. \quad (3.3)$$

Si la variable aleatoria X tiene función de distribución $F(x)$, aplicando la identidad (??) obtenemos

$$f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x). \quad (3.4)$$

Consideremos ahora los dos tipos más importantes de distribuciones. Si la variable aleatoria X tiene distribución discreta, y toma los valores x_1, x_2, \dots con probabilidades p_1, p_2, \dots respectivamente, entonces

$$f(t) = \sum_k e^{itx_k} p_k, \quad (3.5)$$

como se obtiene de aplicar cualquiera de las fórmulas (3.1), (3.2), ó (3.3).

Si X tiene distribución absolutamente continua, con densidad dada por $p(x)$, entonces

$$f(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx, \quad (3.6)$$

como se obtiene de aplicar (3.4).

Calculemos las funciones características de variables aleatorias con distribuciones de ambos tipos, en los casos mas importantes.

Ejemplo 14. Supongamos que la variable aleatoria X tiene distribución degenerada, es decir, existe una constante c tal que $\mathbf{P}(X = c) = 1$. Aplicando la fórmula (3.5), tenemos

$$f(t) = \mathbf{E} e^{itX} = e^{itc}.$$

En particular, si $\mathbf{P}(X = 0) = 1$, tenemos $f(t) = 1$ para todo t real.

Ejemplo 15. Sea X una variable aleatoria con distribución binomial con parámetros (n, p) . Aplicando (3.5), obtenemos

$$f(t) = \sum_{m=0}^n e^{itm} \binom{n}{m} p^m q^{n-m} = \sum_{m=0}^n \binom{n}{m} (pe^{it})^m q^{n-m} = (pe^{it} + q)^n,$$

donde $q = 1 - p$.

Ejemplo 16. Si X tiene distribución de Poisson con parámetro $\lambda > 0$, aplicando la fórmula (3.5), obtenemos

$$f(t) = \mathbf{E} e^{itX} = \sum_{m=0}^{\infty} e^{itm} \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{(e^{it}\lambda)^m}{m!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

Ejemplo 17. Consideremos una variable aleatoria X con distribución normal estándar, con densidad dada por $p(x) = e^{-x^2/2}/\sqrt{2\pi}$. Aplicando la fórmula (3.6), tenemos

$$f(t) = \mathbf{E} e^{itX} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx-x^2/2} dx.$$

Para calcular la integral anterior derivamos con respecto de t , y obtenemos

$$f'(t) = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{itx-x^2/2} dx = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} d(-e^{-x^2/2}).$$

Luego de integrar por partes, resulta

$$f'(t) = \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{itx-x^2/2} dx = -t f(t).$$

En consecuencia, $(\ln f(t))' = -t$, $\ln f(t) = -t^2/2 + C$. Como $f(0) = 1$, obtenemos que $C = 0$, y en conclusión

$$f(t) = e^{-t^2/2}.$$

Si X es una variable aleatoria con distribución normal con parámetros (a, σ) , entonces, como veremos en el ejemplo 19, su función característica está dada por

$$f(t) = e^{iat-\sigma^2 t^2/2}.$$

Ejemplo 18. Sea X una variable aleatoria con distribución exponencial con parámetro $\alpha > 0$. La densidad de esta variable aleatoria está dada por $p(x) = \alpha e^{-\alpha x}$ si $x \geq 0$, $p(x) = 0$ si $x < 0$. Por ésto, aplicando la fórmula (3.6), tenemos

$$f(t) = \mathbf{E} e^{itX} = \alpha \int_0^{\infty} e^{(it-\alpha)x} dx = \frac{\alpha}{\alpha - it}.$$

Estudiemos ahora las propiedades que verifica la función característica $f(t)$ de una variable aleatoria X con función de distribución $F(x)$. Las dos primeras propiedades son evidentes.

Propiedad 1. Se tiene $f(0) = 1$.

Propiedad 2. Se tiene $|f(t)| \leq 1$ para cualquier $t \in \mathbf{R}$.

Propiedad 3. La función característica $f(t)$ es uniformemente continua en la recta real.

Demostración. La demostración, basada en la fórmula (3.3), se adapta sin dificultad si tomamos como definición cualquiera de las fórmulas (3.4), (3.5) ó (3.6).

Sean t, h reales arbitrarios. De (3.3), tenemos

$$f(t+h) - f(t) = \int_{\Omega} e^{itX} (e^{ihX} - 1) d\mathbf{P}. \quad (3.7)$$

Utilizamos la acotación $|e^{iu} - 1| \leq |u|$, válida para todo u real, porque

$$|e^{iu} - 1| = \left| \int_0^u e^{ix} dx \right| \leq \int_0^{|u|} |e^{ix}| dx = |u|.$$

Sea $\varepsilon > 0$, arbitrario. Existe un real A que verifica: A y $-A$ son puntos de continuidad de $F(x)$; $1 - F(A) < \varepsilon/8$; $F(-A) < \varepsilon/8$. Tomando valor absoluto a ambos lados en (3.7), y designando $\mathbf{B} = \{\omega \in \Omega: |X(\omega)| \geq A\}$, obtenemos

$$\begin{aligned} |f(t+h) - f(t)| &\leq \int_{\Omega} |e^{ihX} - 1| d\mathbf{P} \leq 2 \int_{\mathbf{B}} d\mathbf{P} + \int_{\mathbf{B}^c} |e^{ihX} - 1| d\mathbf{P}, \\ &\leq 2\mathbf{P}(\mathbf{B}) + \int_{\mathbf{B}^c} |hX| d\mathbf{P} \leq 2\mathbf{P}(|X| \geq A) + A|h| \\ &\leq 2(1 - F(A) + F(-A)) + Ah \leq \frac{\varepsilon}{2} + Ah < \varepsilon, \end{aligned}$$

si tomamos $h < \varepsilon/(2A)$. Como la acotación es independiente de t , esto concluye la demostración. \square

Propiedad 4. Consideremos la variable aleatoria $Y = aX + b$, donde a, b son constantes. Entonces, la función característica $g(t)$ de la variable aleatoria Y verifica $g(t) = e^{ibt} f(at)$, donde $f(t)$ es la función característica de la variable aleatoria X .

Demostración. Utilizando la propiedad $e^{i(\alpha+\beta)} = e^{i\alpha} e^{i\beta}$ y aplicando la definición (3.2), tenemos

$$g(t) = \mathbf{E} e^{itY} = \mathbf{E} e^{it(aX+b)} = \mathbf{E} (e^{ibt} e^{iatX}) = e^{ibt} \mathbf{E} e^{iatX} = e^{ibt} f(at),$$

lo que concluye la demostración. \square

Ejemplo 19. Calculemos la función característica $g(t)$ de una variable aleatoria Y con función de distribución normal con parámetros (a, σ) .

Es claro que la variable aleatoria $X = (Y - a)/\sigma$ tiene distribución normal con parámetros $(0, 1)$, y por lo tanto función característica $f(t) = e^{-t^2/2}$, como vimos en el ejemplo 17. Aplicando la propiedad anterior, obtenemos

$$g(t) = e^{iat} f(\sigma t) = e^{iat - \sigma^2 t^2 / 2}.$$

Propiedad 5. Consideremos una variable aleatoria X para la cual existe $\alpha_k = \mathbf{E}(X^k)$, el momento de orden k , para algún natural $k \geq 1$. Entonces, su función característica $f(t)$ tiene derivadas continuas, para todo t real, hasta el orden k inclusive. Además

$$f^{(m)}(0) = i^m \alpha_m \quad (1 \leq m \leq k), \quad (3.8)$$

donde $\alpha_m = \mathbf{E}(X^m)$.

Demostración. Derivando formalmente m veces, con respecto de t , bajo el signo de integración en la fórmula (3.4), obtenemos

$$f^{(m)}(t) = \int_{-\infty}^{\infty} (ix)^m e^{itx} dF(x). \quad (3.9)$$

Es claro que como existe el momento de orden $m \leq k$ (proposición ??), tenemos

$$\left| \int_{-\infty}^{\infty} (ix)^m e^{itx} dF(x) \right| \leq \int_{-\infty}^{\infty} |x|^m dF(x) < \infty.$$

Esto permite demostrar que la derivación bajo el signo de integral es válida, y obtener la fórmula (3.9). Sustituyendo $t = 0$ en (3.9) se obtiene (3.8). \square

Dada una variable aleatoria X , si para algún natural $k \geq 1$ existe $\alpha_k = \mathbf{E}(X^k)$, el momento de orden k de la variable aleatoria, aplicando la propiedad 5, el desarrollo de Taylor, y la igualdad $f(0) = 1$, se obtiene, que

$$f(t) = 1 + \sum_{m=1}^k \frac{\alpha_m}{m!} (it)^m + o(|t|^k) \quad (t \rightarrow 0). \quad (3.10)$$

En la demostración de la próxima propiedad, utilizamos el resultado siguiente.

Lema 4. Consideremos dos variables aleatorias independientes X, Y , dos funciones reales $u(x), v(x)$, definidas en la recta real, y supongamos que existen $\mathbf{E}u(X)$ y $\mathbf{E}v(Y)$. Entonces, tiene lugar la identidad

$$\mathbf{E}(u(X)v(Y)) = \mathbf{E}u(X) \mathbf{E}v(Y).$$

Demostración. Veremos la demostración en el caso en que ambas variables tienen distribución discreta, y en el caso en que tienen distribución absolutamente continua.

Supongamos primero que X toma los valores x_1, x_2, \dots , con probabilidades p_1, p_2, \dots , respectivamente; Y toma los valores y_1, y_2, \dots , con probabilidades q_1, q_2, \dots , respectivamente. Aplicando la proposición ??, obtenemos que $\mathbf{P}(X = x_k, Y = y_j) = \mathbf{P}(X = x_k)\mathbf{P}(Y = y_j) = p_k q_j$ ($k, j = 1, 2, \dots$). Por ésto, tenemos

$$\begin{aligned}\mathbf{E}(u(X)v(Y)) &= \sum_{k,j} u(x_k)v(y_j) \mathbf{P}(X = x_k, Y = y_j) \\ &= \sum_k u(x_k)p_k \sum_j v(y_j)q_j = \mathbf{E} u(X) \mathbf{E} v(Y).\end{aligned}$$

Si X e Y tienen distribución absolutamente continua y $r(x, y)$ designa la densidad del vector (X, Y) , aplicando la proposición ??, obtenemos que $r(x, y) = p(x)q(y)$, donde $p(x)$ es la densidad de la variable aleatoria X , $q(y)$ la densidad de la variable aleatoria Y . Por ésto, tenemos

$$\begin{aligned}\mathbf{E}(u(X)v(Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x)v(y)r(x, y) dx dy \\ &= \int_{-\infty}^{\infty} u(x)p(x) dx \int_{-\infty}^{\infty} v(y)q(y) dy = \mathbf{E} u(X) \mathbf{E} v(Y),\end{aligned}$$

concluyendo la demostración. \square

Observación. El lema recién formulado es válido también en el caso en que las funciones de $u(x)$ y $v(x)$ de argumento real, tomen valores complejos, es decir, si tenemos

$$u(x) = u_1(x) + iu_2(x), \quad v(x) = v_1(x) + iv_2(x),$$

donde $u_k(x)$ y $v_k(x)$ son funciones de argumento real, que toman valores reales ($k = 1, 2$). Esto es sencillo de demostrar, aplicando el lema anterior a las partes real e imaginaria del producto $u(X)v(Y)$.

Propiedad 6. Consideremos dos variables aleatorias independientes X e Y , con funciones características $f(t)$ y $g(t)$ respectivamente. Sea $h(t)$ la función característica de la suma $X + Y$. Entonces, se verifica $h(t) = f(t)g(t)$.

Demostración. Tenemos

$$h(t) = \mathbf{E} e^{it(X+Y)} = \mathbf{E} \left(e^{itX} e^{itY} \right) = \mathbf{E} e^{itX} \mathbf{E} e^{itY} = f(t)g(t),$$

en vista de la observación posterior al lema 4. \square

Es válida la siguiente generalización de la propiedad recién demostrada: si X_1, X_2, \dots, X_n es un conjunto de variables aleatorias mutuamente independientes, con funciones características $f_1(t), f_2(t), \dots, f_n(t)$ respectivamente, entonces, la función característica $h(t)$ de la suma $X_1 + X_2 + \dots + X_n$ es igual al producto de las funciones características de los sumandos:

$$h(t) = f_1(t)f_2(t) \cdots f_n(t).$$

Propiedad 7. Para todo t real, se verifica $f(-t) = \overline{f(t)}$.

La propiedad anterior se obtiene de la igualdad

$$f(-t) = \mathbf{E} e^{-itX} = \mathbf{E} \overline{e^{itX}} = \overline{\mathbf{E} e^{itX}} = \overline{f(t)}.$$

Definición 26. Una variable aleatoria X y su distribución $F(x)$ se dicen simétricas cuando las funciones de distribución de las variables aleatorias X y $-X$ son idénticas.

Propiedad 8. Si la variable aleatoria X es simétrica, su función característica $f(t)$ es una función real.

A la conclusión de la propiedad anterior conducen las igualdades

$$f(t) = \mathbf{E} e^{itX} = \mathbf{E} e^{it(-X)} = f(-t) = \overline{f(t)}.$$

En la sección 2 demostraremos el recíproco de la propiedad anterior: si la función característica de una variable aleatoria X es real, la variable aleatoria X es simétrica.

3.2. Fórmula de inversión. Teorema de unicidad

Teorema 17. Consideremos una variable aleatoria X con función de distribución $F(x)$, y función característica $f(t)$. Sean x_1, x_2 dos puntos de continuidad de $F(x)$. Entonces, tiene lugar la igualdad

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t) dt. \quad (3.11)$$

La igualdad (3.11) se denomina fórmula de inversión.

Demostración. Comenzamos introduciendo la función auxiliar

$$R(h, T) = \int_0^T \frac{\text{sen } ht}{t} dt = \int_0^{hT} \frac{\text{sen } u}{u} du.$$

Del cálculo integral, son conocidas las siguientes afirmaciones:

$$\int_0^\infty \frac{\text{sen } u}{u} du = \frac{\pi}{2}, \quad \left| \int_0^x \frac{\text{sen } u}{u} du \right| \leq C \quad (x \geq 0),$$

de donde obtenemos, que

$$\lim_{T \rightarrow \infty} R(h, T) = \begin{cases} -\pi/2, & \text{si } h < 0, \\ \pi/2, & \text{si } h > 0. \end{cases} \quad (3.12)$$

Es importante observar, que esta convergencia es uniforme en los intervalos de la forma $(-\infty, \delta]$, y $[\delta, \infty)$, para todo $\delta > 0$.

La segunda etapa de la demostración, consiste en representar a la integral en (3.11), que designamos I , en términos de la función $R(h, T)$. Aplicando la definición (3.4), e intercambiando el orden de integración (dado que el integrando es una función continua y acotada), tenemos

$$\begin{aligned} I &= \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t) dt = \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{-it} \left(\int_{-\infty}^{\infty} e^{ity} dF(y) \right) dt \\ &= \int_{-\infty}^{\infty} \left(\int_{-T}^T \frac{e^{it(y-x_2)} - e^{it(y-x_1)}}{-it} dt \right) dF(y) \\ &= \int_{-\infty}^{\infty} \left(\int_{-T}^T \frac{\text{sen}(t(y-x_2)) - \text{sen}(t(y-x_1))}{t} dt \right) dF(y) \\ &= 2 \int_{-\infty}^{\infty} (R(y-x_2, T) - R(y-x_1, T)) dF(y). \end{aligned}$$

donde utilizamos que $\int_{-T}^T (\cos(\alpha t)/t) dt = 0$ para todo α real, para obtener la última igualdad. Respecto del comportamiento asintótico del último integrando, tomando por ejemplo $x_1 < x_2$, en vista de (3.12), tenemos

$$\lim_{T \rightarrow \infty} (R(y-x_2, T) - R(y-x_1, T)) = \pi \mathbf{1}_{\{x_1 < y < x_2\}}, \quad (3.13)$$

Como

$$|R(y-x_2, T) - R(y-x_1, T)| \leq 2C,$$

aplicando convergencia dominada, tenemos

$$I \rightarrow 2 \int_{-\infty}^{\infty} \pi \mathbf{1}_{\{x_1 < y < x_2\}} dF(y) = 2\pi (F(x_2) - F(x_1)),$$

donde utilizamos que x_1, x_2 son de continuidad de F , concluyendo la demostración. \square

Teorema 18 (Unicidad). *Sean $f(t), g(t)$ las funciones características correspondientes a dos funciones de distribución $F(x), G(x)$. Supongamos que $f(t) = g(t)$ para todo t real. Entonces, se verifica $F(x) = G(x)$ para todo x real.*

Demostración. Sea \mathbf{C} el conjunto de los puntos en el que ambas funciones $F(x)$ y $G(x)$ son continuas. Como $F(x)$ y $G(x)$ son funciones de distribución, el complemento del conjunto \mathbf{C} es, a lo sumo, numerable. Sean entonces x, y_1, y_2, \dots puntos de \mathbf{C} , tales que $y_n \rightarrow -\infty$ ($n \rightarrow \infty$). Aplicando el teorema 17, obtenemos que $F(x) - F(y_n) = G(x) - G(y_n)$, y tomando límite si $n \rightarrow \infty$ en la igualdad anterior, resulta

$$F(x) = G(x) \text{ para todo } x \text{ en } \mathbf{C}. \quad (3.14)$$

Sea ahora z un real arbitrario. Consideremos $x_1 > x_2 > \dots$ puntos de \mathbf{C} , tales que $x_n \rightarrow z$ ($n \rightarrow \infty$). En vista de (3.14), tenemos $F(x_n) = G(x_n)$. Como ambas funciones de distribución son continuas por la derecha, al tomar límite si $n \rightarrow \infty$ en la igualdad anterior, obtenemos la igualdad $F(z) = G(z)$. Como z es arbitrario, esto concluye la demostración. \square

De acuerdo al teorema 18, denominado *teorema de unicidad*, la función característica de una variable aleatoria define unívocamente (es decir “caracteriza”) su función de distribución. Veamos algunas aplicaciones de este teorema.

Ejemplo 20. Consideremos dos variables aleatorias independientes X_1 y X_2 , cada una de las cuales tiene distribución normal con parámetros (a_1, σ_1) y (a_2, σ_2) respectivamente. Con la ayuda del teorema de unicidad es sencillo demostrar que la suma $X_1 + X_2$ tiene distribución normal (como vimos en la sección 3.4).

La función característica de X_k es $f_k(t) = e^{ia_k t - \sigma_k^2 t^2 / 2}$ ($k = 1, 2$). Como las variables aleatorias son independientes, la función característica de la suma $X_1 + X_2$ es

$$\begin{aligned} g(t) &= \mathbf{E} e^{it(X_1 + X_2)} = \mathbf{E} e^{itX_1} \mathbf{E} e^{itX_2} \\ &= f_1(t) f_2(t) = e^{i(a_1 + a_2)t - (\sigma_1^2 + \sigma_2^2)t^2 / 2}. \end{aligned}$$

Es claro que la función $g(t)$ coincide con la función característica de una variable aleatoria con distribución normal, con parámetros $(a_1 + a_2, (\sigma_1^2 + \sigma_2^2)^{1/2})$. Aplicando el teorema 18 se deduce que la suma $X_1 + X_2$ es una variable aleatoria con distribución normal con éstos parámetros.

Ejemplo 21. Consideremos dos variables aleatorias independientes X e Y , cada una de las cuales tiene distribución exponencial con parámetros $\alpha > 0$ y $\beta > 0$ respectivamente. Veamos que la variable aleatoria $Z = X - Y$ tiene densidad, dada por

$$p(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha x} & \text{si } x > 0, \\ \frac{\alpha\beta}{\alpha+\beta} e^{\beta x} & \text{si } x \leq 0. \end{cases}$$

Por una parte, utilizando el ejemplo 18, tenemos

$$\mathbf{E} e^{itZ} = \mathbf{E} e^{itX} \mathbf{E} e^{-itY} = \frac{\alpha}{\alpha - it} \frac{\beta}{\beta + it}. \quad (3.15)$$

Por otra parte,

$$\begin{aligned} \int_{-\infty}^{\infty} e^{itx} p(x) dx &= \frac{\alpha\beta}{\alpha + \beta} \left(\int_0^{\infty} e^{(it-\alpha)x} dx + \int_{-\infty}^0 e^{(it+\beta)x} dx \right) \\ &= \frac{\alpha\beta}{\alpha + \beta} \left(\frac{1}{\alpha - it} + \frac{1}{\beta + it} \right) = \frac{\alpha}{\alpha - it} \frac{\beta}{\beta + it}. \end{aligned} \quad (3.16)$$

Como los resultados en (3.15) y (3.16) coinciden, del teorema de unicidad se deduce que $p(x)$ es la densidad de Z .

Ejercicio 5. Demostrar que si la función característica $f(t)$ de una variable aleatoria dada es real, entonces la variable aleatoria es simétrica.

3.3. Teoremas de Helly

Los teoremas de Helly juegan un importante rol en la demostración de los teoremas de convergencia de funciones características, que estudiamos en la sección 4. Dado que no se incluyen en los cursos habituales de cálculo, los presentamos aquí con sus correspondientes demostraciones.

Definición 27. Consideremos funciones $F(x), F_1(x), F_2(x), \dots$, acotadas y no decrecientes.

(a) Decimos que la sucesión $\{F_n(x)\}$ converge débilmente a $F(x)$ y escribimos $F_n \rightarrow F$, si $F_n(x) \rightarrow F(x)$ ($n \rightarrow \infty$) para todo punto x de continuidad de la función $F(x)$.

(b) Decimos que la sucesión $\{F_n(x)\}$ converge completamente a $F(x)$ y escribimos $F_n \Rightarrow F$, si $\{F_n(x)\}$ converge débilmente a $F(x)$ (es decir, si $F_n \rightarrow F$) y además¹ $F_n(-\infty) \rightarrow F(-\infty)$, $F_n(\infty) \rightarrow F(\infty)$ si $n \rightarrow \infty$.

¹Designamos $G(-\infty) = \lim_{x \rightarrow -\infty} G(x)$, cuando existe este límite para una cierta función $G(x)$. Análogamente, designamos $G(\infty) = \lim_{x \rightarrow \infty} G(x)$.

Observemos que si $F(x), F_1(x), F_2(x), \dots$ son funciones de distribución, la convergencia débil definida en (a) coincide con la convergencia débil de variables aleatorias definida en la sección 5.1. Además, en este caso, las definiciones (a) y (b) coinciden. Sin embargo, en el caso general esta equivalencia no es cierta, como se ve en el siguiente ejemplo.

Consideremos, para cada $n = 1, 2, \dots$, la función

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

Es claro que $F_n(x) \rightarrow 1/2$ ($n \rightarrow \infty$) para todo x real, y en consecuencia la sucesión $\{F_n(x)\}$ considerada converge débilmente a la función $F(x) = 1/2$. Sin embargo, la convergencia completa no tiene lugar, dado que $F_n(-\infty) = 0$, $F_n(+\infty) = 1$ para todo n , y tenemos $F(-\infty) = F(+\infty) = 1/2$.

Teorema 19 (Helly). *Consideremos una sucesión $\{F_n(x)\}$ de funciones no decrecientes. Supongamos que existen constantes A y B tales que se verifica $A \leq F_n(x) \leq B$ para todo x real y para todo n . Entonces, la sucesión dada contiene una subsucesión $\{F_{n_k}(x)\}$ que converge débilmente a una cierta función $F(x)$, no decreciente y acotada.*

En la demostración de este teorema utilizamos el siguiente resultado.

Lema 5. *Si $F_n(x) \rightarrow F(x)$ para todo $x \in \mathbf{D}$, donde \mathbf{D} es un conjunto denso en la recta real, entonces $F_n \rightarrow F$.*

Demostración.[Demostración del lema] Sea x un punto de continuidad de $F(x)$. Como \mathbf{D} es denso, existen dos sucesiones $\{x'_k\}$ y $\{x''_k\}$ que verifican $x'_k < x < x''_k$ para todo k y $\lim_{k \rightarrow \infty} x'_k = \lim_{k \rightarrow \infty} x''_k = x$. Para cada k y cada n , tenemos

$$F_n(x'_k) \leq F_n(x) \leq F_n(x''_k).$$

Como $\lim_{n \rightarrow \infty} F_n(x'_k) = F(x'_k)$ y $\lim_{n \rightarrow \infty} F_n(x''_k) = F(x''_k)$, de las desigualdades anteriores, si $n \rightarrow \infty$, obtenemos

$$F(x'_k) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x''_k).$$

Hagamos ahora tender k a infinito. Dado que x es un punto de continuidad de $F(x)$, se verifica $\lim_{k \rightarrow \infty} F(x'_k) = \lim_{k \rightarrow \infty} F(x''_k) = F(x)$, por lo que

$$F(x) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x).$$

Entonces, el $\lim_{n \rightarrow \infty} F_n(x)$ existe, y vale $F(x)$. Como el punto de continuidad es arbitrario, esto concluye la demostración. \square

Demostración.[Demostración del teorema 19 de Helly] Sea \mathbf{D} un conjunto denso numerable de números reales x'_1, x'_2, \dots . La sucesión numérica $\{F_n(x'_1)\}$ es acotada, por lo que contiene una subsucesión $\{F_{1n}(x'_1)\}$ que converge a un cierto límite, que designamos $F(x'_1)$.

La sucesión $\{F_{1n}(x'_2)\}$ también contiene una subsucesión $\{F_{2n}(x'_2)\}$, convergente a un cierto límite, que designamos $F(x'_2)$. Además, se verifica $\lim_{n \rightarrow \infty} F_{2n}(x'_1) = F(x'_1)$.

Continuando este proceso, obtenemos, que para cualquier natural k , existen k sucesiones $\{F_{kn}(x'_i)\}$ ($i = 1, \dots, k$) para las cuales se verifica $\lim_{n \rightarrow \infty} F_{kn}(x'_i) = F(x'_i)$ ($i = 1, \dots, k$).

Consideremos ahora la sucesión diagonal compuesta por las funciones $\{F_{nn}(x)\}$. Sea $x'_k \in \mathbf{D}$. Es claro que $\lim_{n \rightarrow \infty} F_{nn}(x'_k) = F(x'_k)$, dado que $\{F_{nn}(x'_k)\}$ es una subsucesión de la sucesión numérica $\{F_{kn}(x'_k)\}$, si $n \geq k$. Hemos así definido una función $F(x)$ en el conjunto \mathbf{D} . Si $x < y$ son dos puntos de \mathbf{D} , entonces $F(x) = \lim_{n \rightarrow \infty} F_{nn}(x) \leq \lim_{n \rightarrow \infty} F_{nn}(y) = F(y)$, y la función $F(x)$ es no decreciente en \mathbf{D} . Es claro también que $A \leq F(x) \leq B$. Estas propiedades permiten extender la función $F(x)$ a toda la recta real, conservando las propiedades mencionadas. Estamos entonces en condiciones de aplicar el lema 5, para concluir la demostración del teorema. \square

Observación. Se puede ver que la función límite $F(x)$ puede elegirse continua por la derecha, si definimos $F(x) = \lim_{n \rightarrow \infty} F(x_n)$, donde $\{x_n\} \in \mathbf{D}$, $x_n \rightarrow x$ ($n \rightarrow \infty$), y $x_n \geq x$ para todo n .

Teorema 20 (Helly). *Consideremos funciones no decrecientes y acotadas $F(x), F_1(x), F_2(x), \dots$ tales que $F_n \Rightarrow F$, y una función $g(x)$ continua y acotada. Entonces*

$$\int_{-\infty}^{\infty} g(x) dF_n(x) \rightarrow \int_{-\infty}^{\infty} g(x) dF(x) \quad (n \rightarrow \infty).$$

Demostración. Sea $\varepsilon > 0$ arbitrario. Designemos

$$G = \sup_{x \in \mathbf{R}} |g(x)|, \quad C = F(\infty) - F(-\infty).$$

Como $\lim_{x \rightarrow -\infty} F(x) = F(-\infty)$, $\lim_{x \rightarrow \infty} F(x) = F(\infty)$, existen $a < b$, puntos de continuidad de $F(x)$, tales que se verifica

$$F(\infty) - F(b) < \varepsilon/(3G), \quad F(a) - F(-\infty) < \varepsilon/(3G). \quad (3.17)$$

Como $g(x)$ es una función continua, existe una partición del intervalo $[a, b]$, designada $a = x_0 < x_1 < \dots < x_N = b$, formada por puntos de continuidad de $F(x)$, y tales que se verifica $|g(x) - g(x_k)| < \delta$ si $x \in (x_{k-1}, x_k)$ ($k = 1, \dots, N$).

Consideremos la función auxiliar

$$g_0(x) = \begin{cases} g(x_k), & \text{si } x \in (x_{k-1}, x_k] \text{ } (k = 1, \dots, N), \\ 0, & \text{en otro caso,} \end{cases}$$

que podemos también definir como $g_0(x) = \sum_{k=1}^N g(x_k) \mathbf{1}_{(x_{k-1}, x_k]}(x)$. Sean

$$I = \int_{-\infty}^{\infty} g(x) dF(x), \quad I_n = \int_{-\infty}^{\infty} g(x) dF_n(x).$$

Sumando y restando, obtenemos

$$I_n - I = \int_{-\infty}^{\infty} (g(x) - g_0(x)) dF_n(x) \quad (3.18)$$

$$+ \int_{-\infty}^{\infty} g_0(x) dF_n(x) - \int_{-\infty}^{\infty} g_0(x) dF(x) \quad (3.19)$$

$$+ \int_{-\infty}^{\infty} (g(x) - g_0(x)) dF(x) \quad (3.20)$$

$$= S_1 + S_2 + S_3.$$

Acotemos cada uno de los tres sumandos anteriores. Para S_3 en (3.20), tenemos

$$\begin{aligned} |S_3| &\leq \int_{-\infty}^a |g(x)| dF(x) + \int_a^b |g(x) - g_0(x)| dF(x) + \int_b^{\infty} |g(x)| dF(x) \\ &\leq G(F(a) - F(-\infty)) + \delta(F(b) - F(a)) + G(F(\infty) - F(b)) \quad (3.21) \\ &\leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon, \end{aligned}$$

en vista de (3.17) y la desigualdad $F(b) - F(a) \leq C$.

Para S_1 en (3.18), cambiando F_n por F , obtenemos

$$\begin{aligned} |S_1| &\leq G(F_n(a) - F_n(-\infty)) \\ &\quad + \delta(F_n(b) - F_n(a)) + G(F_n(\infty) - F_n(b)) < \varepsilon, \quad (3.22) \end{aligned}$$

si n es suficientemente grande, dado que, por la convergencia completa $F_n \Rightarrow F$, la cota obtenida en (3.22) converge a la cota en (3.21).

Finalmente, también utilizando la convergencia completa $F_n \Rightarrow F$, obtenemos $S_2 \rightarrow 0$ ($n \rightarrow \infty$) en (3.19), porque

$$\begin{aligned} \int_{-\infty}^{\infty} g_0(x) dF_n(x) &= \sum_{k=1}^N g(x_k) (F_n(x_k) - F_n(x_{k-1})) \\ &\rightarrow \sum_{k=1}^N g(x_k) (F(x_k) - F(x_{k-1})) = \int_{-\infty}^{\infty} g_0(x) dF(x), \end{aligned}$$

si $n \rightarrow \infty$. En conclusión, para n suficientemente grande, tenemos

$$\left| \int_{-\infty}^{\infty} g(x) dF_n(x) - \int_{-\infty}^{\infty} g(x) dF(x) \right| < \varepsilon + \varepsilon + \varepsilon = 3\varepsilon,$$

lo que concluye la demostración. \square

3.4. Relación entre la convergencia de distribuciones y de funciones características

El objetivo de esta sección es la demostración del siguiente resultado.

Teorema 21. *Consideremos las funciones de distribución*

$$F(x), F_1(x), F_2(x), \dots$$

con sus correspondientes funciones características

$$f(t), f_1(t), f_2(t), \dots$$

Entonces, la sucesión $\{F_n(x)\}$ converge débilmente a $F(x)$ (es decir $F_n \rightarrow F$), si y solo si se verifica

$$f_n(t) \rightarrow f(t) \quad (n \rightarrow \infty) \text{ para todo } t \text{ real.} \quad (3.23)$$

Demostración. Supongamos primero que $F_n \rightarrow F$. Tenemos

$$f_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x), \quad f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x).$$

Como $e^{itx} = \cos tx + i \sin tx$, donde las funciones $\sin tx, \cos tx$ son continuas y acotadas; y tenemos $F_n \Rightarrow F(x)$ (porque se trata de funciones de distribución), obtenemos (3.23) aplicando el teorema 20 de Helly.

Supongamos ahora que se verifica la condición (3.23). En virtud del teorema 19 de Helly, la sucesión $\{F_n(x)\}$ contiene una cierta subsucesión $\{F_{n_k}(x)\}$, que converge débilmente a una cierta función $F(x)$ no decreciente, que verifica $0 \leq F(x) \leq 1$, y que elegimos continua por la derecha. Demostremos que $F(x)$ es una función de distribución. Para ésto hay que demostrar que $F(-\infty) = 0$ y $F(+\infty) = 1$, lo que equivale a demostrar que $\delta = F(+\infty) - F(-\infty) = 1$.

Supongamos que $\delta < 1$ y sea ε tal que $0 < \varepsilon < 1 - \delta$. Como $f(0) = 1$ y $f(t)$ es continua, para γ suficientemente pequeño es válida la desigualdad

$$\frac{1}{2\gamma} \int_{-\gamma}^{\gamma} f(t) dt > 1 - \frac{\varepsilon}{2} > \delta + \frac{\varepsilon}{2}. \quad (3.24)$$

Como $F_{n_k} \rightarrow F$, podemos elegir un real $A > \gamma/(4\varepsilon)$ que verifique: $F(x)$ es continua en A y en $-A$; para todo k suficientemente grande $\delta_k(A) = F_{n_k}(A) - F_{n_k}(-A) < \delta + \varepsilon/4$.

Introducimos ahora la función

$$B(x) = \int_{-\gamma}^{\gamma} e^{itx} dt = \frac{2}{x} \operatorname{sen}(\gamma x)$$

que, como $|e^{itx}| \leq 1$, verifica $|B(x)| \leq 2\gamma$. Además, como $|\operatorname{sen}(\gamma x)| \leq 1$, tenemos $|B(x)| \leq 2/A$, si $|x| \geq A$.

Cambiando el orden de integración y utilizando la función $B(x)$ recién introducida, tenemos

$$\int_{-\gamma}^{\gamma} f_{n_k}(t) dt = \int_{-\gamma}^{\gamma} \left(\int_{-\infty}^{\infty} e^{itx} dF_{n_k}(x) \right) dt = \int_{-\infty}^{\infty} B(x) dF_{n_k}(x).$$

Partiendo el intervalo de integración, obtenemos la siguiente acotación:

$$\begin{aligned} \left| \int_{-\gamma}^{\gamma} f_{n_k}(t) dt \right| &\leq \left| \int_{\{|x| \leq A\}} B(x) dF_{n_k}(x) \right| \\ &\quad + \left| \int_{\{|x| > A\}} B(x) dF_{n_k}(x) \right| \leq 2\gamma\delta_k + \frac{2}{A}. \end{aligned}$$

En vista de la elección de A , dividiendo por 2γ , obtenemos

$$\frac{1}{2\gamma} \left| \int_{-\gamma}^{\gamma} f_{n_k}(t) dt \right| \leq \delta + \frac{\varepsilon}{2}.$$

Como $f_{n_k}(t) \rightarrow f(t)$ ($n \rightarrow \infty$) con $|f_{n_k}(t)| \leq 1$, tomando límite en la desigualdad anterior, obtenemos

$$\frac{1}{2\gamma} \left| \int_{-\gamma}^{\gamma} f(t) dt \right| \leq \delta + \frac{\varepsilon}{2},$$

lo que contradice la desigualdad (3.24). Entonces $\delta = 1$ y $F(x)$ es una función de distribución.

Por último observemos, que como $F_{n_k} \rightarrow F$, aplicando la primera parte del teorema, obtenemos que $f(t)$ es la función característica que corresponde a esta distribución $F(x)$.

Para terminar, resta demostrar que toda la sucesión $\{F_n(x)\}$ converge débilmente a $F(x)$. Supongamos que esto no es cierto. Existe entonces una subsucesión $\{F_{n'}(x)\}$, que converge débilmente a una cierta función $G(x)$, que es distinta de $F(x)$, en por lo menos un punto de continuidad. Una argumentación similar a la aplicada a $\{F_{n_k}(x)\}$ nos permite obtener, que $G(x)$ es una función de distribución, y que $f(t)$ es su función característica. Aplicando el teorema 18 de unicidad de funciones características obtenemos que $F(x) = G(x)$ para todo x real, contradiciendo nuestro supuesto. Entonces $F_n \rightarrow F$, lo que concluye la demostración. \square

En el capítulo 7 estudiaremos distintas variantes del *teorema central del límite*, cuyas demostraciones se basan el teorema recién demostrado.

3.5. Ejercicios

1. Hallar la función característica de una variable aleatoria: (a) con distribución uniforme en el intervalo $(-\ell, \ell)$; (b) con densidad dada por $p(x) = (1 - \cos x)/(\pi x^2)$.
2. Dada una variable aleatoria X , la variable aleatoria simetrizada, designada X^s , se define mediante la igualdad $X^s = X - Y$, donde Y es una variable aleatoria independiente de X , y con su misma distribución. Demostrar que si X tiene función característica $f(t)$, entonces, X^s tiene función característica $|f(t)|^2$.
3. Consideremos una función característica $f(t)$. Demostrar la desigualdad

$$1 - |f(2t)|^2 \leq 4(1 - |f(t)|^2),$$

válida para todo t real.

4. Consideremos funciones características $f_1(t), \dots, f_n(t)$, y constantes positivas b_1, \dots, b_n , que verifican $b_1 + \dots + b_n = 1$. Demostrar que $b_1 f_1(t) + \dots + b_n f_n(t)$ es una función característica.
5. Determinar si las siguientes son funciones características: (a) $\sin t$; (b) $\cos t$; (c) $\cos^2 t$; (d) $\sin t + \cos t$; (e) $(e^{it} + e^{2it})^3/8$; (f) $\Re f(t)$, donde $f(t)$ es una función característica; (g) $\Im f(t)$ donde $f(t)$ es una función característica.

6. Calcular la varianza $\text{var } X$, de una variable aleatoria X , con función característica $f(t) = (1 + e^{3it})^2/4$.
7. Sea X una variable aleatoria con distribución discreta. Esta distribución se denomina *látice*, si existen dos reales $h > 0$ y a , tales que se verifica $\sum_{k=-\infty}^{\infty} \mathbf{P}(X = a + hk) = 1$. (a) Encontrar una distribución discreta que no sea *látice*. (b) Demostrar que una distribución con función característica $f(t)$ es *látice* si y solo si existe $t_0 \neq 0$ tal que $|f(t_0)| = 1$.
8. Sea X una variable aleatoria con distribución *látice*, que toma los valores $a + hk$ ($k = 0, \pm 1, \pm 2, \dots$), con probabilidades $p_k = \mathbf{P}(X = a + kh)$. Demostrar que se verifica

$$p_k = \frac{h}{2\pi} \int_{\frac{\{t\}}{h} \ominus} e^{-it(a+kh)} f(t) dt$$

- para todo k entero, donde $f(t)$ es la función característica de X .
9. Utilizando el teorema 18 de unicidad, demostrar: si X e Y son variables aleatorias independientes, con distribución de Poisson con parámetros λ_1 y λ_2 respectivamente, entonces $X + Y$ tiene distribución de Poisson con parámetro $\lambda_1 + \lambda_2$.
10. Consideremos una función característica $f(t)$ y dos constantes b, c , que verifican $0 < c < 1$, $b > 0$. Demostrar que si $|f(t)| \leq c$, cuando $|t| \geq b$, entonces $|f(t)| \leq 1 - (1 - c^2)t^2/(8b^2)$, si $|t| < b$.
11. Demostrar que si una función característica verifica la condición

$$\limsup_{|t| \rightarrow \infty} |f(t)| < 1$$

- (denominada *condición (C) de Cramér*), entonces, para todo $\varepsilon > 0$, existe un real positivo $c < 1$, tal que $|f(t)| \leq c$, cuando $|t| \geq \varepsilon$. Sugerencia: Utilizar el ejercicio 3.5.
12. Consideremos una variable aleatoria con función característica $f(t)$. Demostrar que si la variable aleatoria es absolutamente continua, entonces $f(t) \rightarrow 0$, si $|t| \rightarrow \infty$. (Sugerencia: utilizar el teorema de Riemann-Lebesgue).
13. Sea $F(x)$ una función de distribución con función característica $f(t)$. Demostrar la igualdad

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^2} dF(x) = \int_0^{\infty} e^{-t} (1 - \Re f(t)) dt.$$

14. Una función característica se denomina *infinitamente divisible* si para cada natural $n \geq 1$, existe una función característica $f_n(t)$, tal que $f(t) = (f_n(t))^n$. Demostrar las siguientes proposiciones: (a) Si $f(t)$ y $g(t)$ son funciones características infinitamente divisibles, entonces, $f(t)g(t)$ es una función característica infinitamente divisible. (b) Si $f(t)$ es función característica infinitamente divisible, entonces, $f(t) \neq 0$ para todo t real. (Sugerencia: utilizar la desigualdad del ejercicio 3.5.)
15. Si una función característica verifica $f(t) = 1 + o(t^2)$ ($t \rightarrow 0$), entonces $f(t) = 1$ para todo t real.
16. Sea $f(t)$ la función característica de una variable aleatoria con distribución no degenerada. Demostrar que existen reales positivos δ y ε , tales que $|f(t)| \leq 1 - \varepsilon t^2$ para $|t| \leq \delta$.
17. Sea X una variable aleatoria con función característica $f(t)$, con distribución lártice, que toma los valores $a + hk$ ($k = 0, \pm 1, \pm 2, \dots$), donde $h > 0$ y a son números reales fijos. El número h se denomina el *paso* de la distribución. El paso h se denomina *maximal*, si no existe un par h_1, a_1 , con $h_1 > h$, tal que $\sum_{k=-\infty}^{\infty} \mathbf{P}(X = a_1 + h_1 k) = 1$. Demostrar que el paso h es maximal si y solo si se verifican las condiciones $|f(2\pi/h)| = 1$, y $|f(t)| < 1$ para $0 < |t| < 2\pi/h$.
18. Una función de distribución, o su función característica $f(t)$, se denomina *estable*, cuando para todo par a_1 y a_2 de números reales positivos, existen $a > 0$ y b reales, tales que $f(a_1 t)f(a_2 t) = e^{ibt} f(at)$. (a) Demostrar que esta definición es equivalente a la siguiente: la función de distribución $F(x)$ es estable, si para todo $a_1 > 0$, $a_2 > 0$, b_1 y b_2 , existen $a > 0$ y b reales tales que $F(a_1 x + b_1) * F(a_2 x + b_2) = F(ax + b)$, donde $*$ designa la convolución. (b) Determinar si son estables las siguientes distribuciones: (i) degenerada, (ii) normal, (iii) uniforme, (iv) binomial, (v) de Poisson.
19. (a) Hallar la función característica de una variable aleatoria con distribución Gama, con parámetros (α, β) . (b) Demostrar que si T_1, \dots, T_n son variables aleatorias independientes con distribución común exponencial de parámetro α , entonces, su suma $T_1 + \dots + T_n$, tiene densidad dada por $p(x) = \alpha^n x^{n-1} e^{-\alpha x} / (n-1)!$.

Capítulo 4

Teorema central del límite

Se denomina *teorema central del límite* a cualquier proposición que establece, bajo determinadas condiciones, que la función de distribución de la suma de una cantidad creciente a infinito de variables aleatorias, converge a la función de distribución normal. Aplicando el teorema central del límite podemos aproximar la distribución de la suma de un gran número de variables aleatorias, mediante la distribución normal. En este capítulo, dedicado a estudiar diversas variantes del teorema central del límite, comenzamos por el teorema de Lindeberg–Lévy que considera sumas de variables independientes e idénticamente distribuidas. En la sección 2 se demuestra un resultado más general: el teorema de Lindeberg, en el cual no se supone que las variables aleatorias consideradas tienen la misma distribución.

4.1. Teorema de Lindeberg–Lévy

Decimos que X_1, X_2, \dots es una *sucesión de variables aleatorias independientes*, cuando las variables aleatorias X_1, \dots, X_n son mutuamente independientes para cada $n = 1, 2, \dots$. Recordemos, que X_1, X_2, \dots es una *sucesión de variables aleatorias idénticamente distribuidas* cuando todas las variables aleatorias consideradas tienen la misma distribución.

Teorema 22 (Lindeberg–Lévy).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes e idénticamente distribuidas, con esperanza matemática $\mathbf{E} X_1 = a$ y varianza $\text{var } X_1 = \sigma^2 > 0$. Designemos

$$F_n(x) = \mathbf{P} \left(\frac{X_1 + \dots + X_n - na}{\sigma\sqrt{n}} \leq x \right).$$

Entonces,

$$F_n(x) \rightarrow \Phi(x) \quad (n \rightarrow \infty) \text{ para todo } x \text{ real,} \quad (4.1)$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ es la distribución normal estándar.

Demostración. La primer etapa consiste en demostrar el teorema en el caso particular en el que $\mathbf{E} X_1 = a = 0$ y $\text{var } X_1 = \sigma^2 = 1$.

Consideremos entonces para cada $n = 1, 2, \dots$ la variable aleatoria

$$Z_n = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n).$$

Como $a = 0$ y $\sigma^2 = 1$, tenemos $\mathbf{P}(Z_n \leq x) = F_n(x)$. La demostración se basa en la aplicación del teorema 21. Calculemos $f_n(t)$, la función característica de Z_n , en términos de $v(t)$, la función característica de X_1 :

$$\begin{aligned} f_n(t) &= \mathbf{E} e^{itZ_n} = \mathbf{E} e^{i(t/\sqrt{n}) \sum_{k=1}^n X_k} = \mathbf{E} \prod_{k=1}^n e^{i(t/\sqrt{n}) X_k} \\ &= \prod_{k=1}^n \mathbf{E} e^{i(t/\sqrt{n}) X_k} = \left[v\left(\frac{t}{\sqrt{n}}\right) \right]^n, \end{aligned} \quad (4.2)$$

donde utilizamos que las variables aleatorias son idénticamente distribuidas en la última igualdad, y que son independientes en la ante última. Como $\alpha_2 = \text{var } X_1 = 1 < \infty$, aplicando el desarrollo de Taylor (3.10) de orden $k = 2$ para la función característica de X_1 , tenemos

$$v(u) = 1 - \frac{u^2}{2} + o(u^2) \quad (u \rightarrow 0), \quad (4.3)$$

dado que $\alpha_1 = \mathbf{E} X_1 = 0$. Consideremos un real t arbitrario y fijo. Queremos calcular $\lim_{n \rightarrow \infty} f_n(t)$. Si en (4.3) ponemos $u = t/\sqrt{n}$, tenemos

$$v\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \quad (n \rightarrow \infty),$$

dado que $1/n \rightarrow 0$ si y solo si $u \rightarrow 0$.

Verifiquemos ahora la validez de la identidad

$$\ln(1+z) = z + r(z), \quad (4.4)$$

donde $|r(z)| \leq 2|z|^2$, si z es un número complejo que verifica $|z| < 1/2$. En efecto, considerando el desarrollo de Taylor de la función logaritmo, tenemos

$$r(z) = \ln(1+z) - z = \sum_{m=2}^{\infty} (-1)^{m-1} z^m / m.$$

Acotando y sumando la serie geométrica que se obtiene, tenemos

$$|r(z)| \leq \sum_{m=2}^{\infty} |z|^m = |z|^2 \sum_{m=0}^{\infty} |z|^m = \frac{|z|^2}{1-|z|} \leq 2|z|^2,$$

porque $(1-|z|)^{-1} \leq 2$, dado que $|z| \leq 1/2$. Esto prueba (4.4).

Si $z = v(t/\sqrt{n}) - 1 = -t^2/(2n) + o(1/n)$, para n suficientemente grande podemos aplicar la fórmula (4.4), para obtener

$$\ln v\left(\frac{t}{\sqrt{n}}\right) = -\frac{t^2}{2n} + o\left(\frac{1}{n}\right),$$

porque, en este caso $|r(z)| \leq 2|z|^2 = t^4/(2n^2) + o(1/n^2)$.

Estamos en condiciones de tomar logaritmo en la fórmula (4.2):

$$\ln f_n(t) = n \ln v\left(\frac{t}{\sqrt{n}}\right) = n \left[-\frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right] = -\frac{t^2}{2} + o(1).$$

En otras palabras, $f_n(t) \rightarrow e^{-t^2/2}$ si $n \rightarrow \infty$. Como $f(t) = e^{-t^2/2}$ es la función característica de la distribución normal $\Phi(x)$, aplicando el teorema 21 obtenemos que $F_n(x) \rightarrow \Phi(x)$ para todo x real (dado que $\Phi(x)$ es continua en \mathbf{R}), concluyendo la primer etapa de la demostración ($a = 0, \sigma^2 = 1$).

Supongamos ahora que $\mathbf{E}X_1 = a, \text{var} X_1 = \sigma^2$, con $a, \sigma > 0$ arbitrarios. Consideremos las variables aleatorias auxiliares $Y_n = (X_n - a)/\sigma$ ($n = 1, 2, \dots$). Es fácil de ver que $\{Y_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, con $\mathbf{E}Y_1 = 0$, y $\text{var} Y_1 = 1$. Entonces

$$\begin{aligned} F_n(x) &= \mathbf{P}\left(\frac{X_1 + \dots + X_n - na}{\sigma\sqrt{n}} \leq x\right) \\ &= \mathbf{P}\left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \leq x\right) \rightarrow \Phi(x), \end{aligned}$$

para todo x real si $n \rightarrow \infty$, dado que $\{Y_n\}$ verifica las condiciones de la primer etapa de la demostración. Esto es la tesis del teorema. \square

Observación. Es posible demostrar que la convergencia en (4.1) es uniforme en el conjunto de los x reales. No es difícil verificar esta afirmación directamente; es consecuencia del *teorema de Pólya*: si una sucesión de funciones de distribución $\{G_n(x)\}$ converge a una función de distribución continua $G(x)$ para todo x , entonces, esta convergencia es uniforme en la recta real (ver ejercicio ??, capítulo 5).

Veamos que el teorema límite integral de De Moivre–Laplace de la sección 2.3, es un corolario del teorema de Lindeberg–Lévy recién demostrado.

Consideremos entonces una serie de n experimentos independientes, con dos resultados posibles cada uno (éxito y fracaso), y probabilidad de éxito igual a p en cada experimento ($0 < p < 1$). Sea μ la cantidad de éxitos en n experimentos. Veamos como se formula el teorema límite integral de De Moivre–Laplace en términos de variables aleatorias. Para ésto consideremos una sucesión de variables aleatorias X_1, X_2, \dots , cada una de las cuales toma el valor 1 con probabilidad p (si ocurre un éxito) y el valor 0 con probabilidad $q = 1 - p$ (si ocurre un fracaso). Tenemos $\mathbf{E} X_k = p$, $\mathbf{var} X_k = pq > 0$ para cada $i = 1, 2, \dots$. Además, $\mu = \sum_{k=1}^n X_k$, porque la suma contiene tantos sumandos iguales a uno como éxitos ocurren en los primeros n experimentos, siendo nulos los sumandos restantes. La sucesión $\{X_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, por lo que es aplicable el teorema de Lindeberg–Lévy. De (4.1) obtenemos, que

$$\mathbf{P} \left(\frac{\mu - np}{\sqrt{npq}} \leq x \right) - \Phi(x) \rightarrow 0 \quad (n \rightarrow \infty), \quad (4.5)$$

uniformemente en el conjunto de los x reales, en vista de la última observación. Poniendo entonces en (4.5) primero $x = b$, luego $x = a$, y restando, obtenemos

$$\mathbf{P} \left(a < \frac{\mu - np}{\sqrt{npq}} \leq b \right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \rightarrow 0 \quad (n \rightarrow \infty),$$

uniformemente, en el conjunto de los reales $a < b$, que es el contenido del teorema límite integral de De Moivre–Laplace.

4.2. Teorema de Lindeberg

Teorema 23 (Lindeberg).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes, con esperanzas a_1, a_2, \dots y varianzas $\sigma_1^2, \sigma_2^2, \dots$, no todas nulas. Designemos

$$V_n(x) = \mathbf{P}(X_n \leq x), \quad B_n = \sum_{k=1}^n \sigma_k^2,$$

$$F_n(x) = \mathbf{P} \left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n (X_k - a_k) \leq x \right).$$

Supongamos que se verifica la condición de Lindeberg: Para todo $\varepsilon > 0$

$$\Lambda_n(\varepsilon) = \frac{1}{B_n} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} (x - a_k)^2 dV_k(x) \rightarrow 0 \quad (n \rightarrow \infty). \quad (4.6)$$

Entonces

$$F_n(x) \rightarrow \Phi(x) \quad (n \rightarrow \infty) \text{ para todo } x \text{ real.}$$

La demostración del teorema de Lindeberg utiliza el siguiente resultado de cálculo, que incluimos por conveniencia del lector.

Lema 6. Vale la desigualdad

$$\left| e^{ix} - \sum_{\nu=0}^{k-1} \frac{(ix)^\nu}{\nu!} \right| \leq \frac{1}{k!} |x|^k, \quad (4.7)$$

para todo real x , y todo natural $k \geq 1$.

Demostración. [Demostración del lema 6] Como

$$\int_0^x e^{it} dt = \frac{1}{i} (e^{ix} - 1)$$

obtenemos que $|e^{ix} - 1| \leq |x|$, demostrando la fórmula (4.7) para $k = 1$.

Para demostrar la validez de (4.7) para $k + 1$, a partir de su validez para k , escribimos

$$I = \int_0^x \left(e^{it} - \sum_{\nu=0}^{k-1} \frac{(it)^\nu}{\nu!} \right) dt = \frac{1}{i} \left(e^{ix} - \sum_{\nu=0}^k \frac{(ix)^\nu}{\nu!} \right).$$

Entonces

$$\left| e^{ix} - \sum_{\nu=0}^k \frac{(ix)^\nu}{\nu!} \right| = |I| \leq \int_0^{|x|} \frac{t^k}{k!} dt = \frac{|x|^{k+1}}{(k+1)!},$$

concluyendo la demostración. \square

Demostración. [Demostración del teorema 23 de Lindeberg] Observemos en primer lugar que podemos suponer $a_n = 0$ ($n = 1, 2, \dots$). El caso general en el que las variables aleatorias tienen esperanzas arbitrarias, se reduce a este caso particular mediante la consideración de las variables aleatorias $Y_n = X_n - a_n$ ($n = 1, 2, \dots$) que verifican $\mathbf{E} Y_n = 0$, $\text{var } Y_n = \sigma_n^2$.

Supongamos entonces que $\mathbf{E} X_n = 0$ ($n = 1, 2, \dots$). En primer lugar, demostremos que la condición (4.6) de Lindeberg, que en nuestro caso es: Para todo $\varepsilon > 0$

$$\Lambda_n(\varepsilon) = \frac{1}{B_n} \sum_{k=1}^n \int_{\{|x| \geq \varepsilon \sqrt{B_n}\}} x^2 dV_k(x) \rightarrow 0 \quad (n \rightarrow \infty),$$

implica la condición:

$$\frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty). \quad (4.8)$$

En efecto, para $\varepsilon > 0$ arbitrario y para cada $k = 1, 2, \dots, n$, tenemos

$$\begin{aligned} \sigma_k^2 &= \int_{\{|x| < \varepsilon \sqrt{B_n}\}} x^2 dV_k(x) + \int_{\{|x| \geq \varepsilon \sqrt{B_n}\}} x^2 dV_k(x) \\ &\leq \varepsilon^2 B_n + \sum_{k=1}^n \int_{\{|x| \geq \varepsilon \sqrt{B_n}\}} x^2 dV_k(x). \end{aligned}$$

Entonces, como la cota obtenida no depende de k , dividiendo por B_n tenemos

$$\frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \leq \varepsilon^2 + \Lambda_n(\varepsilon) < \varepsilon^2 + \varepsilon,$$

para n suficientemente grande. Como $\varepsilon > 0$ es arbitrario, hemos demostrado (4.8).

Consideremos ahora $Z_n = \sum_{k=1}^n X_k / \sqrt{B_n}$ y calculemos su función característica $f_n(t)$ en términos de $v_k(t)$, la función característica de X_k ($k = 1, \dots, n$). Tenemos

$$\begin{aligned} f_n(t) &= \mathbf{E} e^{itZ_n} = \mathbf{E} e^{i(t/\sqrt{B_n}) \sum_{k=1}^n X_k} = \mathbf{E} \prod_{k=1}^n e^{i(t/\sqrt{B_n}) X_k} \\ &= \prod_{k=1}^n \mathbf{E} e^{i(t/\sqrt{B_n}) X_k} = \prod_{k=1}^n v_k\left(\frac{t}{\sqrt{B_n}}\right), \end{aligned} \quad (4.9)$$

donde utilizamos que las variables aleatorias son independientes.

Para la demostración del teorema es suficiente verificar que $f_n(t) \rightarrow e^{-t^2/2}$ ($n \rightarrow \infty$) y aplicar el teorema 21. Para demostrar entonces la convergencia de las funciones características, tomamos logaritmo en (4.9) y utilizamos el siguiente resultado.

Lema 7. Para cada t real, tiene lugar la igualdad

$$\ln f_n(t) = \sum_{k=1}^n \ln v_k\left(\frac{t}{\sqrt{B_n}}\right) = \sum_{k=1}^n \left[v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right] + R_n(t),$$

donde $R_n(t) \rightarrow 0$ ($n \rightarrow \infty$).

Demostración. [Demostración del lema 7] Consideremos

$$r_k(t) = \ln v_k\left(\frac{t}{\sqrt{B_n}}\right) - \left[v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right] \text{ para } k = 1, \dots, n,$$

$$R_n(t) = \sum_{k=1}^n r_k(t).$$

Como $\mathbf{E} X_k = \int_{-\infty}^{\infty} x dV_k(x) = 0$, aplicando el lema 6 con $k = 2$, tenemos

$$\left| v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right| = \left| \int_{-\infty}^{+\infty} \left(e^{\frac{itx}{\sqrt{B_n}}} - 1 - \frac{itx}{\sqrt{B_n}} \right) dV_k(x) \right|$$

$$\leq \frac{t^2}{2B_n} \int_{-\infty}^{+\infty} x^2 dV_k(x) = \frac{t^2 \sigma_k^2}{2B_n} \quad (4.10)$$

$$\leq \frac{t^2}{2} \times \frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty), \quad (4.11)$$

según vimos en la fórmula (4.8). Luego, si n es suficientemente grande, designando $z_k = v_k(t/\sqrt{B_n}) - 1$, se verifica $|z_k| < 1/2$ para todo $k = 1, \dots, n$, y podemos utilizar el desarrollo del logaritmo (4.4), para obtener

$$|r_k(t)| \leq 2 \left| v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right|^2 \leq \frac{t^4}{2B_n^2} \sigma_k^4 \leq \frac{t^4 \sigma_k^2}{2B_n} \times \frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2,$$

donde utilizamos (4.10). Por ésto,

$$|R_n(t)| = \sum_{k=1}^n |r_k(t)| \leq \frac{t^4}{2} \sum_{k=1}^n \frac{\sigma_k^2}{B_n} \times \frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty),$$

en vista de (4.8), concluyendo la demostración del lema. \square

Resta la etapa final, que consiste en demostrar que

$$\ln f_n(t) + t^2/2 \rightarrow 0 \quad (n \rightarrow \infty). \quad (4.12)$$

Con este fin, introducimos

$$I_k = v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 + \frac{t^2\sigma_k^2}{2B_n} = \int_{-\infty}^{\infty} \left(e^{itx/\sqrt{B_n}} - 1 - \frac{itx}{\sqrt{B_n}} - \frac{(itx)^2}{2B_n} \right) dV_k(x)$$

para cada $k = 1, 2, \dots$. Acotamos I_k partiendo el dominio de integración en en las regiones $\{|x| < \varepsilon\sqrt{B_n}\}$ y $\{|x| \geq \varepsilon\sqrt{B_n}\}$, y utilizando el lema 6 con $k = 3$ en la primer región, y con $k = 2$ en la segunda. Las acotaciones que obtenemos de dicho lema, son

$$\begin{aligned} \left| e^{iy} - 1 - iy - \frac{(iy)^2}{2} \right| &\leq \frac{1}{6}|y|^3, \\ \left| e^{iy} - 1 - iy - \frac{(iy)^2}{2} \right| &\leq |e^{iy} - 1 - iy| + \frac{1}{2}|y|^2 \leq |y|^2. \end{aligned}$$

Por eso, si $y = ix/\sqrt{B_n}$, tenemos

$$\begin{aligned} |I_k| &\leq \int_{\{|x| < \varepsilon\sqrt{B_n}\}} \frac{|tx|^3}{6B_n^{3/2}} dV_k(x) + \int_{\{|x| \geq \varepsilon\sqrt{B_n}\}} \frac{|tx|^2}{B_n} dV_k(x) \\ &\leq \frac{\varepsilon\sigma_k^2|t|^3}{6B_n} + \frac{t^2}{B_n} \int_{\{|x| \geq \varepsilon\sqrt{B_n}\}} |x|^2 dV_k(x). \end{aligned} \quad (4.13)$$

Estamos en condiciones de concluir la demostración. Utilizando el lema 7, tenemos

$$\begin{aligned} \ln f_n(t) + \frac{t^2}{2} &= \sum_{k=1} \left[\ln v_k\left(\frac{t}{\sqrt{B_n}}\right) + \frac{t^2\sigma_k^2}{2B_n} \right] \\ &= \sum_{k=1} \left[v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 - \frac{(it)^2\sigma_k^2}{2B_n} \right] + R_n(t) \\ &= \sum_{k=1}^n I_k + R_n(t). \end{aligned}$$

Aplicando ahora la acotación (4.13), tenemos

$$\left| \ln f_n(t) + \frac{t^2}{2} \right| \leq \sum_{k=1}^n |I_k| + R_n(t) \leq \frac{\varepsilon|t|^3}{6} + t^2\Lambda_n(\varepsilon) + R_n(t).$$

Como el primer sumando es arbitrariamente pequeño, el segundo converge a cero (aplicando la condición de Lindeberg), y el tercero también tiende a cero (según demostramos en el lema 7), obtuvimos (4.12). Con la aplicación del teorema 21 concluimos la demostración. \square

El teorema 22 de Lindeberg–Lévy, resulta ser un corolario del teorema de Lindeberg, recién demostrado. En efecto, en el caso de variables aleatorias con distribución común $V(x)$, esperanza matemática a y variancia $\sigma^2 > 0$, obtenemos que $B_n = n\sigma^2$, y, para $\varepsilon > 0$ arbitrario, se verifica

$$\Lambda_n(\varepsilon) = \frac{1}{\sigma^2} \int_{\{|x-a| \geq \varepsilon\sigma\sqrt{n}\}} (x-a)^2 dV(x) \rightarrow 0 \quad (n \rightarrow \infty)$$

porque $\sigma^2 = \int_{-\infty}^{\infty} (x-a)^2 V(x) < \infty$. En conclusión, si se verifican las hipótesis del teorema de Lindeberg–Lévy, también se verifican las del teorema de Lindeberg; mientras que las tesis de estos dos teoremas, en el caso particular considerado, coinciden.

Volvamos ahora al caso general, en el que las distribuciones no necesariamente son idénticas. Sea $\{X_n\}$ una sucesión de variables aleatorias que verifica las condiciones del teorema 23 de Lindeberg. Consideremos

$$X_{nk} = \frac{X_k - a_k}{\sqrt{B_n}} \quad (k = 1, \dots, n).$$

Poniendo $Z_n = \sum_{k=1}^n (X_k - a_k)/\sqrt{B_n}$, tenemos $Z_n = \sum_{k=1}^n X_{nk}$. Demostremos que la condición de Lindeberg implica la condición: Para todo $\varepsilon > 0$

$$\mathbf{P} \left(\max_{1 \leq k \leq n} |X_{nk}| \geq \varepsilon \right) \rightarrow 0 \quad (n \rightarrow \infty). \quad (4.14)$$

La fórmula (4.14) significa que las variables aleatorias son uniformemente “pequeñas”. Veamos su demostración. Dado $\varepsilon > 0$, tenemos

$$\begin{aligned} \mathbf{P}(|X_{nk}| \geq \varepsilon) &= \mathbf{P}(|X_k - a_k| \geq \varepsilon\sqrt{B_n}) = \int_{\{|x-a_k| \geq \varepsilon\sqrt{B_n}\}} dV_k(x) \\ &\leq \frac{1}{\varepsilon^2 B_n} \int_{\{|x-a_k| \geq \varepsilon\sqrt{B_n}\}} (x - a_k)^2 dV_k(x). \end{aligned}$$

Por ésto,

$$\begin{aligned} \mathbf{P} \left(\max_{1 \leq k \leq n} |X_{nk}| \geq \varepsilon \right) &\leq \mathbf{P} \left(\bigcup_{k=1}^n \{|X_{nk}| \geq \varepsilon\} \right) \leq \sum_{k=1}^n \mathbf{P}(|X_{nk}| \geq \varepsilon) \\ &\leq \frac{1}{\varepsilon^2 B_n} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon\sqrt{B_n}\}} (x - a_k)^2 dV_k(x) = \frac{1}{\varepsilon^2} \Lambda_n(\varepsilon), \end{aligned}$$

obteniendo la condición (4.14).

4.3. Teorema de Lyapunov

Teorema 24 (Lyapunov).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes, con esperanzas a_1, a_2, \dots y varianzas $\sigma_1^2, \sigma_2^2, \dots$, no todas nulas. Designemos

$$B_n = \sum_{k=1}^n \sigma_k^2, \quad F_n(x) = \mathbf{P} \left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n (X_k - a_k) \leq x \right).$$

Supongamos que se verifica la condición de Lyapunov: Existe $\delta > 0$ tal que

$$L_n(\delta) = \frac{1}{B_n^{1+\delta/2}} \sum_{k=1}^n \mathbf{E} |X_k - a_k|^{2+\delta} \rightarrow 0 \quad (n \rightarrow \infty). \quad (4.15)$$

Entonces

$$F_n(x) \rightarrow \Phi(x) \quad (n \rightarrow \infty) \text{ para todo } x \text{ real.}$$

Observemos que la condición de Lyapunov implica que existen los momentos de orden $2 + \delta$ de las variables aleatorias X_k ($k = 1, 2, \dots$).

Demostración. Como la tesis del teorema 23 y la del teorema 24 coinciden, es suficiente demostrar que la condición (4.15) de Lyapunov implica la condición (4.6) de Lindeberg. En efecto, supongamos que se verifica la condición (4.15) para un cierto $\delta > 0$. Dado $\varepsilon > 0$ arbitrario, tenemos

$$\begin{aligned} \Lambda_n(\varepsilon) &= \frac{1}{B_n} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} (x - a_k)^2 dV_k(x) \\ &\leq \frac{1}{\varepsilon^\delta B_n^{1+\delta/2}} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} |x - a_k|^{2+\delta} dV_k(x) \\ &\leq \frac{1}{\varepsilon^\delta B_n^{1+\delta/2}} \sum_{k=1}^n \int_{-\infty}^{\infty} |x - a_k|^{2+\delta} dV_k(x) = \frac{1}{\varepsilon^\delta} L_n(\delta) \rightarrow 0 \quad (n \rightarrow \infty). \end{aligned}$$

De aquí se obtiene la demostración, resultando el teorema de Lyapunov un caso particular del teorema de Lindeberg. \square

Como conclusión de este capítulo haremos algunas consideraciones relativas a la *velocidad de convergencia* en el teorema central del límite.

Lyapunov demostró que si se verifican las condiciones del teorema 24 con $0 < \delta < 1$, entonces, existe una constante C tal que, para n suficientemente grande, se verifica

$$|F_n(x) - \Phi(x)| \leq CL_n(\delta),$$

uniformemente en el conjunto de los x reales.

En el caso $\delta = 1$ Esseen obtuvo la siguiente desigualdad, válida para *todo* natural $n = 1, 2, \dots$. Sean X_1, \dots, X_n variables aleatorias independientes, con esperanzas a_1, \dots, a_n y varianzas $\sigma_1^2, \dots, \sigma_n^2$, no todas nulas. Supongamos que $\mathbf{E}|X_k - a_k|^3 < \infty$ ($k = 1, \dots, n$). Designemos

$$B_n = \sum_{k=1}^n \sigma_k^2, \quad F_n(x) = \mathbf{P} \left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n (X_k - a_k) \leq x \right),$$

$$L_n = L_n(1) = \frac{1}{B_n^{3/2}} \sum_{k=1}^n \mathbf{E}|X_k - a_k|^3.$$

Entonces

$$|F_n(x) - \Phi(x)| \leq AL_n \quad \text{para todo } x \text{ real y todo } n = 1, 2, \dots, \quad (4.16)$$

donde $A > 0$ es una constante absoluta. (Está demostrado que la acotación (4.16) es válida con $A = 0,8$). De (4.16) se obtiene que si $L_n \rightarrow 0$ entonces $F_n(x) \rightarrow \Phi(x)$ para todo x real.

La desigualdad (4.16) es válida también si $\mathbf{E}|X_n - a_n|^{2+\delta} < \infty$ ($k = 1, \dots, n$) para algún $0 < \delta \leq 1$, definiendo $L_n(\delta)$ como en (4.15).

En el caso particular en el que X_1, \dots, X_n son variables aleatorias idénticamente distribuidas, con $a = \mathbf{E}X_1$, $\sigma^2 = \text{var} X_1$, y $\delta = 1$, los resultados anteriores son los siguientes. Si designamos

$$L_n = \frac{\rho}{\sqrt{n}}, \quad \text{con } \rho = \frac{\mathbf{E}|X_1 - a|^3}{\sigma^3},$$

aplicando la desigualdad (4.16) de Esseen, obtenemos

$$|F_n(x) - \Phi(x)| \leq \frac{A\rho}{\sqrt{n}} \quad (4.17)$$

para todo x real y todo $n = 1, 2, \dots$. Es posible demostrar (ver por ejemplo §5.2 en [?]), que sin la introducción de condiciones adicionales, esta acotación es óptima en el siguiente sentido: el término \sqrt{n} en el denominador, a la derecha en (4.17), no se puede sustituir por una función $g(n)$, que verifique $g(n)/\sqrt{n} \rightarrow \infty$ ($n \rightarrow \infty$).

4.4. Ejercicios

1. Al disparar a un blanco, se obtienen 10 puntos con probabilidad 0,3; 9 puntos con probabilidad 0,3; 8 con probabilidad 0,2; 7 con probabilidad 0,1 y 6

- con probabilidad 0,1. Utilizando el teorema central del límite, estimar la probabilidad de que, al realizar 100 disparos, se obtengan más de 870 puntos.
2. Se tira un dado 500 veces. Hallar un valor aproximado para la probabilidad de que la suma de puntos obtenidos sea mayor que 1800.
 3. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, e idénticamente distribuidas, con $\mathbf{E} X_1 = a$, y $\mathbf{var} X_1 = \sigma^2 > 0$. Verificar que $\{Y_n\}$, con $Y_n = (X_n - a)/\sigma$, es una sucesión de variables aleatorias independientes, idénticamente distribuidas, con esperanza nula, y varianza igual a uno.
 4. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, que verifican $\mathbf{P}(X_n = n^a) = \mathbf{P}(X_n = -n^a) = 1/2$ para cada $n = 1, 2, \dots$, donde $a > -1/2$. ¿Es aplicable el teorema central del límite a esta sucesión?
 5. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, con esperanza nula. Supongamos que $|X_n| \leq C$ para todo n , donde C es una cierta constante. Sea $B_n = \sum_{k=1}^n \mathbf{var} X_k^2$. Demostrar que si $B_n \rightarrow \infty$ ($n \rightarrow \infty$), entonces, es aplicable el teorema central del límite a esta sucesión, es decir $\mathbf{P}\left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n X_k \leq x\right) \rightarrow \Phi(x)$.
 6. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, que verifican $\mathbf{P}(X_n = -1/\sqrt{n}) = \mathbf{P}(X_n = 1/\sqrt{n}) = p$, $\mathbf{P}(X_n = 0) = 1 - 2p$ para todo n , y algún p en el intervalo $0 < p \leq 1/2$. ¿Es aplicable el teorema central del límite a esta sucesión?
 7. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tales que, cada variable aleatoria X_n , tiene distribución uniforme, en el intervalo $(-\sqrt{n}, \sqrt{n})$. Demostrar que para esta sucesión, es aplicable el teorema central del límite.
 8. Sea μ la cantidad de éxitos, en una serie n experimentos independientes, con dos resultados posibles cada uno (éxito y fracaso). Sea p_k la probabilidad de que ocurra un éxito en el k -ésimo experimento, y $q_k = 1 - p_k$, la probabilidad de que ocurra un fracaso ($k = 1, 2, \dots$). Demostrar que si $\sum_{k=1}^{\infty} p_k q_k = \infty$, entonces, la función de distribución de la variable aleatoria $(\mu - \sum_{k=1}^n p_k)(\sum_{k=1}^n p_k q_k)^{-1/2}$ (cantidad de éxitos, normalizados), converge a la distribución normal estándar, si $n \rightarrow \infty$.
 9. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, que verifica la condición de Lindeberg. Demostrar que $B_n \rightarrow \infty$. (Sugerencia: Utilizar, que la condición de Lindeberg implica la condición (4.8).)

10. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, con esperanza matemática nula. Supongamos que se verifica

$$\sum_{k=1}^n \mathbf{E} |X_k|^3 \leq Bn, \quad \sum_{k=1}^n \mathbf{E} |X_k|^2 \geq An$$

para todo n , donde A y B son constantes positivas. ¿Es aplicable el teorema central del límite a esta sucesión?

11. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, con esperanza matemática nula. Supongamos que se verifica

$$\sum_{k=1}^n \mathbf{E} |X_k|^2 \left| \ln |X_k| \right|^{1+\delta} \leq Bn, \quad \sum_{k=1}^n \mathbf{E} |X_k|^2 \geq An,$$

para todo n y algún $\delta > 0$, donde A y B son constantes positivas. Demostrar que para esta sucesión es aplicable el teorema central del límite. (Sugerencia: verificar la condición de Lindeberg).

12. Consideremos una sucesión $\{X_n\}$ de variables aleatorias independientes, idénticamente distribuidas, con distribución de Poisson con parámetro 1. Hallar $\lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{1}{\sqrt{n}} \left(\sum_{k=1}^n X_k - n \right) \leq x \right)$.

13. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tal que para cada $n = 1, 2, \dots$, la variable aleatoria X_n tiene distribución normal, con $\mathbf{E} X_n = 0$, y $\mathbf{var} X_n = 2^{2n}$. (a) ¿Es aplicable el teorema central del límite a esta sucesión? (b) ¿Se cumple la condición de Lindeberg para esta sucesión de variables aleatorias?

14. Sea $\Phi(x)$ la distribución normal estándar. Demostrar la desigualdad

$$\frac{1 - \Phi(x)}{x\sqrt{2\pi}} \sim e^{-x^2/2},$$

para todo $x > 0$.

15. Demostrar la fórmula

$$1 - \Phi(x) = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \left(1 + O\left(\frac{1}{x^2}\right) \right) \quad (x \rightarrow \infty). \quad (4.18)$$

(Sugerencia: Utilizar la identidad $\int_x^\infty e^{-t^2/2} dt = \int_x^\infty \frac{1}{t} d(e^{-t^2/2})$ e integrar por partes.)

- 16.** Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, idénticamente distribuidas, con esperanza nula y varianza $\sigma^2 > 0$. Designemos $F_n(x) = \mathbf{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n X_k \leq x\right)$. Demostrar que si se verifica

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sqrt{n}}$$

para todo x real, y todo n natural, donde C es una constante positiva, entonces, para $0 < \varepsilon < 1$, se verifica

$$\frac{1 - F_n(x)}{1 - \Phi(x)} \rightarrow 1 \quad (n \rightarrow \infty),$$

uniformemente, en el conjunto $0 \leq x \leq (1 - \varepsilon)\sqrt{\ln n}$. (Sugerencia: utilizar la fórmula (4.18).)

Capítulo 5

Martingalas

Continuamos con el estudio de las sucesiones de variables aleatorias dependientes a través de las *martingalas*. Esta noción proviene de los juegos de azar, mas precisamente de situaciones en las que el conocimiento de resultados anteriores no permite aumentar la fortuna esperada de un apostador. Las martingalas son además una potente herramienta para el estudio del comportamiento asintótico y el cálculo de funcionales de procesos estocásticos.

Índice general

5.1. Esperanza condicional

En esta sección introducimos la *esperanza condicional*, herramienta esencial en la definición y el estudio de las martingalas. Consideremos un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$. Recordemos que la probabilidad condicional de un suceso \mathbf{B} con respecto a otro \mathbf{A} con $\mathbf{P}(\mathbf{A}) > 0$ se define mediante

$$\mathbf{P}(\mathbf{B} \mid \mathbf{A}) = \frac{\mathbf{P}(\mathbf{B}\mathbf{A})}{\mathbf{P}(\mathbf{A})}$$

y se interpreta como la probabilidad de que ocurra \mathbf{B} cuando sabemos que \mathbf{A} es cierto.

La esperanza condicional de una variable aleatoria X con respecto de una segunda variable aleatoria Y , o más en general respecto de una σ -álgebra $\mathcal{D} \subset \mathcal{F}$ representará el valor esperado de la variable X *modificado* según el valor que tomó la variable Y , o según el suceso de \mathcal{D} que aconteció.

Supongamos en principio que Y es una variable aleatoria con distribución discreta, que toma valores en un conjunto $\mathbf{I} = \{i_0, i_1, \dots\}$. Notemos $\mathbf{D}_k = \{\omega : Y(\omega) = i_k\}$, y consideremos la σ -álgebra $\sigma(Y)$ definida por

$$\sigma(Y) = \{\mathbf{D} = \cup_{k \in \mathbf{K}} \mathbf{D}_k : \mathbf{K} \text{ subconjunto de } \mathbf{I}\}. \quad (5.1)$$

$\sigma(Y)$ está formada por uniones arbitrarias de conjuntos \mathbf{D}_k , siendo la mínima σ -álgebra tal que Y es una variable aleatoria en Ω .

En el caso considerado, dada X variable aleatoria con $\mathbf{E}|X| < \infty$, definimos la esperanza condicional de X con respecto de $\sigma(Y)$ mediante

$$\mathbf{E}(X \mid \sigma(Y)) = \sum_{k=0}^{\infty} \frac{\mathbf{E}(X \mathbf{1}_{\mathbf{D}_k})}{\mathbf{P}(\mathbf{D}_k)} \mathbf{1}_{\mathbf{D}_k}(\omega), \quad (5.2)$$

que también notaremos $\mathbf{E}(X \mid Y)$.

Observemos en primer lugar que $\mathbf{E}(X | Y)$ es una variable aleatoria cuyos valores dependen únicamente de los valores que toma Y . En otras palabras, $\mathbf{E}(X | Y)$ es constante en los conjuntos donde Y es constante. Esto es equivalente a decir que

(A) $\mathbf{E}(X | Y)$ es una variable aleatoria $\sigma(Y)$ medible.

Sea ahora $\mathbf{D} = \cup_{j \in \mathbf{J}} \mathbf{D}_j$ con $\mathbf{J} \subseteq \mathbf{N}$. Como los conjuntos $\mathbf{D}_1, \mathbf{D}_2, \dots$ son disjuntos dos a dos

$$\int_{\mathbf{D}} \mathbf{E}(X | Y) d\mathbf{P} = \sum_{j \in \mathbf{J}} \int_{\mathbf{D}_j} \frac{\mathbf{E}(X \mathbf{1}_{\mathbf{D}_j})}{\mathbf{P}(\mathbf{D}_j)} \mathbf{1}_{\mathbf{D}_j} d\mathbf{P} = \sum_{j \in \mathbf{J}} \mathbf{E}(X \mathbf{1}_{\mathbf{D}_j}) = \int_{\mathbf{D}} X d\mathbf{P}.$$

Es decir, las integrales de X y su esperanza condicional respecto de $\sigma(Y)$ coinciden en los conjuntos de $\sigma(Y)$. Distinguiamos esta propiedad.

(B) $\int_{\mathbf{D}} \mathbf{E}(X | Y) d\mathbf{P} = \int_{\mathbf{D}} X d\mathbf{P}$ para todo $\mathbf{D} \in \sigma(Y)$.

Ejemplo 22. Esperanza condicional bajo independencia o medibilidad Sea Y variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$, que toma valores en $\mathbf{I} = \{i_0, i_1, \dots\}$, y $\sigma(Y)$ como en (5.1).

(a) Supongamos que X e Y son variables aleatorias independientes. Entonces

$$\mathbf{E}(X | Y) = \sum_{k=0}^{\infty} \frac{\mathbf{E}(X \mathbf{1}_{\mathbf{D}_k})}{\mathbf{P}(\mathbf{D}_k)} \mathbf{1}_{\mathbf{D}_k} = \sum_{k=0}^{\infty} \frac{\mathbf{E} X \mathbf{P}(\mathbf{D}_k)}{\mathbf{P}(\mathbf{D}_k)} \mathbf{1}_{\mathbf{D}_k} = \mathbf{E} X.$$

Es decir, la esperanza condicional coincide con la esperanza.

(b) Supongamos ahora que $X = f(Y)$ donde $f: \mathbf{R} \rightarrow \mathbf{R}$ es una función medible¹. En este caso

$$\mathbf{E}(X | Y) = \sum_{k=0}^{\infty} \frac{\mathbf{E}(X \mathbf{1}_{\{Y=i_k\}})}{\mathbf{P}(Y = i_k)} \mathbf{1}_{\mathbf{D}_k} = \sum_{k=0}^{\infty} f(i_k) \mathbf{1}_{\mathbf{D}_k} = f(Y) = X.$$

Es decir, la esperanza condicional *coincide* con la variable aleatoria.

Consideremos ahora una σ -álgebra arbitraria $\mathcal{D} \subset \mathcal{F}$. El hecho relevante es que las propiedades (A) y (B) son suficientes para caracterizar la esperanza condicional de una variable aleatoria respecto de \mathcal{D} , generalizando la fórmula (5.2).

¹Esto equivale a decir que X es $\sigma(Y)$ -medible.

Definición 28. Sea X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$ con $\mathbf{E}|X| < \infty$, y \mathcal{D} una σ -álgebra contenida en \mathcal{F} . La esperanza condicional de X con respecto a \mathcal{D} es una variable aleatoria notada $\mathbf{E}(X | \mathcal{D})$ que cumple

- (A) $\mathbf{E}(X | \mathcal{D})$ es \mathcal{D} medible.
- (B) $\int_{\mathbf{D}} \mathbf{E}(X | \mathcal{D}) d\mathbf{P} = \int_{\mathbf{D}} X d\mathbf{P}$ para todo $\mathbf{D} \in \mathcal{D}$.

El siguiente resultado demuestra que la esperanza condicional así definida existe y es única, a menos de conjuntos de probabilidad nula.

Teorema 25. En las condiciones de la definición 28 existe un variable aleatoria que cumple (A) y (B). Además, esta variable aleatoria es única a menos de conjuntos de probabilidad nula.

Demostración. Sea $\mu: \mathcal{D} \rightarrow \mathbf{R}$ una medida signada finita definida mediante $\mu(\mathbf{D}) = \int_{\mathbf{D}} X d\mathbf{P}$, si $\mathbf{D} \in \mathcal{D}$. Si notamos $\bar{\mathbf{P}}$ la restricción de \mathbf{P} a (Ω, \mathcal{D}) tenemos que μ es absolutamente continua respecto de $\bar{\mathbf{P}}$ en (Ω, \mathcal{D}) , porque $\bar{\mathbf{P}}(\mathbf{D}) = 0$ implica $\mu(\mathbf{D}) = 0$. Existe entonces, por el teorema de Radon-Nikodym una variable aleatoria Z , \mathcal{D} medible, tal que $\mu(\mathbf{D}) = \int_{\mathbf{D}} Z d\bar{\mathbf{P}}$ para todo \mathbf{D} en \mathcal{D} . Entonces

$$\int_{\mathbf{D}} Z d\mathbf{P} = \int_{\mathbf{D}} Z d\bar{\mathbf{P}} = \mu(\mathbf{D}) = \int_{\mathbf{D}} X d\mathbf{P}.$$

Luego Z cumple (A) y (B) en la definición 28. La unicidad es consecuencia del resultado que sigue. \square

Lema 8. Sean X, Y, Z variables aleatorias definidas en $(\Omega, \mathcal{F}, \mathbf{P})$, integrables. Entonces

- (a) $\int_{\mathbf{A}} X d\mathbf{P} \geq 0$ para todo $\mathbf{A} \in \mathcal{F} \iff X \geq 0$ P -c.s.
- (b) $\int_{\mathbf{A}} Y d\mathbf{P} \geq \int_{\mathbf{A}} Z d\mathbf{P}$ para todo $\mathbf{A} \in \mathcal{F} \iff Y \geq Z$ P -c.s.
- (c) $\int_{\mathbf{A}} Y d\mathbf{P} = \int_{\mathbf{A}} Z d\mathbf{P}$ para todo $\mathbf{A} \in \mathcal{F} \iff Y = Z$ P -c.s.

Demostración. Las implicaciones \Leftarrow son inmediatas. Veamos \Rightarrow en (a). Si existe $\varepsilon > 0$ tal que $\mathbf{P}(X \leq -\varepsilon) > 0$, con $\mathbf{A}_\varepsilon = \{\omega: X(\omega) \leq -\varepsilon\}$ tenemos $0 \leq \int_{\mathbf{A}_\varepsilon} X d\mathbf{P} \leq -\varepsilon \mathbf{P}(\mathbf{A}_\varepsilon) < 0$, lo que es absurdo, luego $X \geq 0$ \mathbf{P} -c.s. \Rightarrow en (b) se sigue de (a) si $X = Y - Z$. \Rightarrow en (c) sigue de (b). \square

Observación. La unicidad de la esperanza condicional recién demostrada permite concluir, que si \mathcal{D} en la definición 28 está generada como en (5.1) la fórmula (5.2) da la esperanza condicional ya que cumple (A) y (B).

Una forma de determinar la esperanza condicional es verificar (A) y (B) y apelar a la unicidad del teorema 25, como se ve a continuación

Ejemplo 23. Sea X variable aleatoria definida en $(\Omega, \mathcal{F}, \mathbf{P})$.

(a) Supongamos que $\mathbf{P}(X = C) = 1$, con $C \in \mathbf{R}$. Entonces $\mathbf{E}(X | \mathcal{D}) = C$ para toda $\mathcal{D} \subset \mathcal{F}$, dado que verifica (A) y (B).

(b) Si $\mathcal{D} = \{\emptyset, \Omega\}$ entonces $\mathbf{E}(X | \mathcal{D}) = \mathbf{E}X$ verifica (A) y (B).

(c) Si X es \mathcal{D} medible, entonces $\mathbf{E}(X | \mathcal{D}) = X$ porque verifica (A) y (B). Obsérvese que, en el caso general, si X no es \mathcal{D} medible, se cumple (B) pero no (A).

Un caso particular muy importante es cuando la σ -álgebra que condiciona es generada por una o varias variables aleatorias. Sea Y una variable aleatoria definida en $(\Omega, \mathcal{F}, \mathbf{P})$. $\sigma(Y)$ denota la mínima σ -álgebra que hace que Y sea variable aleatoria, es decir, la generada por los conjuntos de la forma $\{\omega: Y(\omega) \leq x\}$ para todo $x \in \mathbf{R}$. En este caso, también notamos

$$\mathbf{E}(X | \sigma(Y)) = \mathbf{E}(X | Y).$$

Más en general, si notamos $\sigma(Y_1, \dots, Y_n)$ la σ -álgebra generada por los conjuntos $\{\omega: Y_1(\omega) \leq x_1, \dots, Y_n(\omega) \leq x_n\}$ para todo $(x_1, \dots, x_n) \in \mathbf{R}^n$ notaremos

$$\mathbf{E}(X | \sigma(Y_1, \dots, Y_n)) = \mathbf{E}(X | Y_1, \dots, Y_n).$$

Ejemplo 24. Sea (X, Y) un vector aleatorio bidimensional definido en $(\Omega, \mathcal{F}, \mathbf{P})$ con densidad $p = p(x, y)$. Es decir, para cada boreliano de \mathbf{R}^2 , $\mathbf{P}((X, Y) \in \mathbf{B}) = \int_{\mathbf{B}} p(x, y) dx dy$. Veamos que

$$\mathbf{E}(X | Y) = \frac{\int_{\mathbf{R}} xp(x, Y) dx}{\int_{\mathbf{R}} p(x, Y) dx}, \quad (5.3)$$

en los puntos $\omega \in \Omega$ en los que $\int_{\mathbf{R}} p(x, Y(\omega)) dx > 0$, tomando $\mathbf{E}(X | Y) = 0$ en caso contrario. Como (5.3) es una función de Y , es $\sigma(Y)$ medible. Resta verificar (B), y es suficiente hacerlo en los conjuntos de la forma $\{\omega: Y(\omega) \leq b\}$, con $b \in \mathbf{R}$. En efecto,

$$\begin{aligned} \int_{\{Y \leq b\}} \left(\frac{\int_{\mathbf{R}} xp(x, Y) dx}{\int_{\mathbf{R}} p(x, Y) dx} \right) d\mathbf{P} &= \int_{-\infty}^b \frac{\int_{\mathbf{R}} xp(x, y) dx}{\int_{\mathbf{R}} p(x, y) dx} \left(\int_{\mathbf{R}} p(x, y) dx \right) dy \\ &= \int_{(-\infty, b]} dy \int_{\mathbf{R}} xp(x, y) dx = \int_{\{Y \leq b\}} X d\mathbf{P} \end{aligned}$$

donde en la primera y la última igualdad hemos realizado cambios de variable.

Ejemplo 25. Sea ahora (X, Y) un vector gaussiano, con densidad

$$p(x, y) = \frac{1}{2\pi(1-\rho^2)} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\left(\frac{x-a_1}{\sigma_1} \right)^2 + \left(\frac{y-a_2}{\sigma_2} \right)^2 - 2\rho \frac{x-a_1}{\sigma_1} \frac{y-a_2}{\sigma_2} \right] \right\}.$$

Calculemos $\mathbf{E}(X | Y)$, que se interpreta como un estimador de X cuando observamos Y . Veamos que

$$\mathbf{E}(X | Y) = a_1 + \rho \frac{\sigma_1}{\sigma_2} (Y - a_2).$$

Notando $z = x - a_1 - \rho \frac{\sigma_1}{\sigma_2} (Y - a_2)$, utilizando (5.3)

$$\mathbf{E}(X | Y) = \frac{\int_{\mathbf{R}} xp(x, Y)dx}{\int_{\mathbf{R}} p(x, Y)dx} = a_1 + \rho \frac{\sigma_1}{\sigma_2} (Y - a_2) + \frac{\int_{\mathbf{R}} zp(x, Y)dx}{\int_{\mathbf{R}} p(x, Y)dx}.$$

Veamos que $\int_{\mathbf{R}} zp(x, Y)dx = 0$. En efecto, cambiando de variable en la integral

$$\int_{\mathbf{R}} zp(x, Y)dx = \frac{1}{2\pi(1-\rho^2)} \exp \left\{ -\frac{1}{2} \left(\frac{Y-a_2}{\sigma_2} \right)^2 \right\} \int_{\mathbf{R}} z \exp \left\{ \frac{-1}{2(1-\rho^2)\sigma_1^2} z^2 \right\} dz = 0,$$

porque el integrando último es una función impar de z .

5.2. Propiedades de la esperanza condicional

Comenzamos estudiando aquellas propiedades que generalizan las correspondientes de la esperanza de una variable aleatoria.

Teorema 26. Sean X, Y variables aleatorias en $(\Omega, \mathcal{F}, \mathbf{P})$, con $\mathbf{E}|X| < \infty$, $\mathbf{E}|Y| < \infty$. Sea \mathcal{D} una σ -álgebra contenida en \mathcal{F} . Valen las siguientes propiedades:

- (i) *Monotonía.* Si $X \leq Y$ \mathbf{P} -c.s., entonces $\mathbf{E}(X | \mathcal{D}) \leq \mathbf{E}(Y | \mathcal{D})$ \mathbf{P} -c.s.
- (ii) *Linealidad.* Para a, b reales $\mathbf{E}(aX + bY | \mathcal{D}) = a\mathbf{E}(X | \mathcal{D}) + b\mathbf{E}(Y | \mathcal{D})$
- (iii) $|\mathbf{E}(X | \mathcal{D})| \leq \mathbf{E}(|X| | \mathcal{D})$.

Demostración. (i) Sabemos que para $\mathbf{D} \in \mathcal{D}$ tenemos $\int_{\mathbf{D}} X d\mathbf{P} \leq \int_{\mathbf{D}} Y d\mathbf{P}$, de donde por (B),

$$\int_{\mathbf{D}} \mathbf{E}(X | \mathcal{D}) d\mathbf{P} \leq \int_{\mathbf{D}} \mathbf{E}(Y | \mathcal{D}) d\mathbf{P}. \quad (5.4)$$

Por el lema 8 parte (b), y por ser \mathcal{D} medibles los integrandos en (5.4), se obtiene (i).

(ii) Para $\mathbf{D} \in \mathcal{D}$,

$$\begin{aligned} \int_{\mathbf{D}} \mathbf{E}(aX + bY | \mathcal{D}) d\mathbf{P} &= \int_{\mathbf{D}} (aX + bY) d\mathbf{P} = a \int_{\mathbf{D}} X d\mathbf{P} + b \int_{\mathbf{D}} Y d\mathbf{P} \\ &= a \int_{\mathbf{D}} \mathbf{E}(X | \mathcal{D}) d\mathbf{P} + b \int_{\mathbf{D}} \mathbf{E}(Y | \mathcal{D}) d\mathbf{P} = \int_{\mathbf{D}} [a\mathbf{E}(X | \mathcal{D}) + b\mathbf{E}(Y | \mathcal{D})] d\mathbf{P}, \end{aligned}$$

y (ii) sigue del Lema 8 parte (c).

(iii) Sabemos $-|X| \leq X \leq |X|$. Luego, por (i) y (ii)

$$-\mathbf{E}(|X| | \mathcal{D}) \leq \mathbf{E}(X | \mathcal{D}) \leq \mathbf{E}(|X| | \mathcal{D}),$$

que equivale a (iii). \square

Las siguientes propiedades son específicas de la esperanza condicional.

Teorema 27. *Propiedades telescópicas*

Sea X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$, con $\mathbf{E}|X| < \infty$, y \mathcal{C}, \mathcal{D} σ -álgebras contenidas en \mathcal{F} . Entonces

- (i) $\mathbf{E}\mathbf{E}(X | \mathcal{D}) = \mathbf{E}X$.
- (ii) $\mathbf{E}|\mathbf{E}(X | \mathcal{D})| \leq \mathbf{E}|X|$.
- (iii) Si $\mathcal{C} \subseteq \mathcal{D}$ entonces $\mathbf{E}(\mathbf{E}(X | \mathcal{C}) | \mathcal{D}) = \mathbf{E}(X | \mathcal{C})$.
- (iv) Si $\mathcal{C} \subseteq \mathcal{D}$ entonces $\mathbf{E}(\mathbf{E}(X | \mathcal{D}) | \mathcal{C}) = \mathbf{E}(X | \mathcal{C})$.

Demostración. (i) Como $\Omega \in \mathcal{D}$ por (B)

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

(ii) Por el teorema 26 parte (iii) y la propiedad anterior

$$\mathbf{E}|\mathbf{E}(X | \mathcal{D})| \leq \mathbf{E}\mathbf{E}(|X| | \mathcal{D}) = \mathbf{E}|X|.$$

(iii) Notemos $Z = \mathbf{E}(X | \mathcal{C})$. Como $\mathcal{C} \subset \mathcal{D}$, Z es \mathcal{D} medible, luego según

(c) en el ejemplo anterior $\mathbf{E}(Z | \mathcal{D}) = Z$ que es (iii).

(iv) Sea $\mathbf{C} \in \mathcal{C}$. Como además $\mathbf{C} \in \mathcal{D}$ tenemos

$$\int_{\mathbf{C}} \mathbf{E}(X | \mathcal{C}) d\mathbf{P} = \int_{\mathbf{C}} X d\mathbf{P} = \int_{\mathbf{C}} \mathbf{E}(X | \mathcal{D}) d\mathbf{P} = \int_{\mathbf{C}} \mathbf{E}(\mathbf{E}(X | \mathcal{D}) | \mathcal{C}) d\mathbf{P},$$

y la igualdad se sigue de (c) en el Lema 8 dado que el primer y último integrando son \mathcal{C} medibles. \square

Estudiamos ahora casos particulares en los cuales poseemos información sobre la relación entre la variable aleatoria a condicionar y la σ -álgebra condicionante.

Teorema 28. *Medibilidad e independencia*

Sean X, Y variables aleatorias en $(\Omega, \mathcal{F}, \mathbf{P})$ y \mathcal{D} una σ -álgebra contenida en \mathcal{F} .

(i) Si $\mathbf{E}|X| < \infty$ y X e $\mathbf{1}_{\mathbf{D}}$ son independientes para todo $\mathbf{D} \in \mathcal{D}$, entonces

$$\mathbf{E}(X | \mathcal{D}) = \mathbf{E} X.$$

(ii) Si X es \mathcal{D} medible, $\mathbf{E}|Y| < \infty$ y $\mathbf{E}|XY| < \infty$, entonces

$$\mathbf{E}(XY | \mathcal{D}) = X \mathbf{E}(Y | \mathcal{D}).$$

Demostración. (i) $\mathbf{E} X$ es \mathcal{D} medible por ser constante. Si $\mathbf{D} \in \mathcal{D}$,

$$\int_{\mathbf{D}} X d\mathbf{P} = \int_{\Omega} \mathbf{1}_{\mathbf{D}} X d\mathbf{P} = \mathbf{E}(X \mathbf{1}_{\mathbf{D}}) = \mathbf{P}(\mathbf{D}) \mathbf{E} X = \int_{\mathbf{D}} \mathbf{E} X d\mathbf{P},$$

luego se verifican (A) y (B).

(ii) Supongamos primero que X e Y son no negativas. Como $X \mathbf{E}(Y | \mathcal{D})$ es \mathcal{D} medible debemos verificar (B), que seguirá de verificar

$$\int_{\mathbf{D}} X \mathbf{E}(Y | \mathcal{D}) d\mathbf{P} = \int_{\mathbf{D}} XY d\mathbf{P} \quad \text{para todo } \mathbf{D} \in \mathcal{D}. \quad (5.5)$$

Si $X = \mathbf{1}_{\mathbf{D}_1}$ con $\mathbf{D}_1 \in \mathcal{D}$, (5.5) es (B) en la definición de esperanza condicional para la variable Y y el conjunto $\mathbf{D} \cap \mathbf{D}_1$. Por linealidad, (5.5) vale para una variables aleatoria de la forma $X = \sum_{i=1}^K c_k \mathbf{1}_{\mathbf{D}_k}$ con $\mathbf{D}_k \in \mathcal{D}$ para $k = 1, \dots, K$. Sea ahora $\{X_n\}$ una sucesión de variables aleatorias de la forma anterior, creciente y con $0 \leq X_n \rightarrow X$ si $n \rightarrow \infty$ \mathbf{P} -c.s. Por convergencia monótona

$$\int_{\mathbf{D}} X \mathbf{E}(Y | \mathcal{D}) d\mathbf{P} = \lim_n \int_{\mathbf{D}} X_n \mathbf{E}(Y | \mathcal{D}) d\mathbf{P} = \lim_n \int_{\mathbf{D}} X_n Y d\mathbf{P} = \int_{\mathbf{D}} XY d\mathbf{P}.$$

En particular deducimos que $\int_{\Omega} X \mathbf{E}(Y | \mathcal{D}) d\mathbf{P} < \infty$. Sean ahora X e Y arbitrarias. Consideremos $X = X^+ - X^-$, $Y = Y^+ - Y^-$.

$$\begin{aligned} \mathbf{E}(XY | \mathcal{D}) &= X^+ \mathbf{E}(Y^+ | \mathcal{D}) + X^- \mathbf{E}(Y^- | \mathcal{D}) - X^+ \mathbf{E}(Y^- | \mathcal{D}) - X^- \mathbf{E}(Y^+ | \mathcal{D}) \\ &= (X^+ - X^-) \mathbf{E}(Y^+ - Y^- | \mathcal{D}) = X \mathbf{E}(Y | \mathcal{D}). \end{aligned}$$

□

Ejemplo 26 (Esperanza condicional en el paseo al azar simple). Sean X_1, X_2, \dots v.a.i.i.d. de Bernoulli con $\mathbf{P}(X_1 = 1) = p$, $\mathbf{P}(X_1 = -1) = 1 - p$, definidas en $(\Omega, \mathcal{F}, \mathbf{P})$. Notamos $S_0 = 0$ y $S_n = X_1 + \dots + X_n$. Dado n ,

notamos por \mathcal{F}_n la σ -álgebra que hace medibles las variables X_1, \dots, X_n . Observese que en este caso \mathcal{F}_n está formado por 2^n conjuntos de la forma

$$\{\omega: X_1(\omega) = e_1, \dots, X_n(\omega) = e_n\}$$

donde e_1, \dots, e_n es una sucesión de 1 y -1 . Queremos calcular

$$\mathbf{E}(S_{n+1} | \mathcal{F}_n) = \mathbf{E}(S_n + X_{n+1} | \mathcal{F}_n) = \mathbf{E} X_{n+1} + S_n = p - q + S_n.$$

$$\mathbf{E} \left(\left(\frac{q}{p} \right)^{S_{n+1}} | \mathcal{F}_n \right) = \left(\frac{q}{p} \right)^{S_n} \mathbf{E} \left(\frac{q}{p} \right)^{X_{n+1}} = \left(\frac{q}{p} \right)^{S_n}.$$

Consideramos finalmente la propiedad relativa a intercambiar el orden de la esperanza condicional y la composición con una función convexa.

Teorema 29 (Desigualdad de Jensen). *Sean X una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$, \mathcal{D} una σ -álgebra contenida en \mathcal{F} y $\phi: \mathbf{R} \rightarrow \mathbf{R}$ una función convexa². Supongamos que $\mathbf{E}|X| < \infty$ y $\mathbf{E}|\phi(X)| < \infty$. Entonces*

$$\phi(\mathbf{E}(Y | \mathcal{D})) \leq \mathbf{E}(\phi(Y) | \mathcal{D}). \quad (5.6)$$

La fórmula (5.6) se denomina desigualdad de Jensen (para la esperanza condicional).

Demostración. Consideremos un conjunto $\{y_1, y_2, \dots\}$ denso y numerable de puntos de la recta real. Para cada $n = 1, 2, \dots$, definimos

$$a_n = \phi'(y_n) = \lim_{y \rightarrow y_n^+} \frac{\phi(y) - \phi(y_n)}{y - y_n}, \quad b_n = \phi(y_n) - a_n y_n.$$

La constante a_n es la derivada por la derecha de la función $\phi(y)$ en el punto y_n , que siempre existe, dado que $\phi(y)$ es convexa. Por eso mismo, obtenemos

$$a_n y + b_n \leq \phi(y) \quad \text{para } y, n \text{ arbitrarios,} \quad (5.7)$$

y como es continua, no es difícil verificar que

$$\phi(y) = \sup_n (a_n y + b_n). \quad (5.8)$$

Si sustituimos $y = Y$ en la desigualdad (5.7), aplicando la monotonía y la linealidad de la esperanza condicional, obtenemos

$$a_n \mathbf{E}(Y | \mathcal{D}) + b_n \leq \mathbf{E}(\phi(Y) | \mathcal{D}).$$

² $\phi: \mathbf{R} \rightarrow \mathbf{R}$ se dice convexa cuando $\phi(\lambda x + (1 - \lambda)y) \leq \lambda\phi(x) + (1 - \lambda)\phi(y)$ para todo x, y reales y $\lambda \in (0, 1)$.

Para concluir la demostración tomamos supremo al variar $n = 1, 2, \dots$ en la desigualdad anterior, y nos referimos a (5.8). \square

Con el mismo esquema de demostración (tomando esperanza en vez de esperanza condicional en la fórmula (5.7)), se obtiene la *desigualdad de Jensen*: dadas una variable aleatoria Y y una función convexa $\phi(y)$, tales que existen las esperanzas $\mathbf{E} Y, \mathbf{E} \phi(Y)$, se verifica

$$\phi(\mathbf{E} Y) \leq \mathbf{E} \phi(Y). \quad (5.9)$$

Alternativamente, la desigualdad (5.9) se puede obtener como un caso particular de la desigualdad (5.6) en el que $X(\omega) = 1$ ($\omega \in \Omega$).

5.3. Martingalas

Definición 29. *Filtración.* Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad. Una sucesión creciente de σ -álgebras contenidas en \mathcal{F} , $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}$ se dirá una *filtración*, notándose $\mathbf{F} = (\mathcal{F}_0, \mathcal{F}_1, \dots)$. Notaremos $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ a un espacio de probabilidad con filtración.

Una filtración en un espacio de probabilidad describe la información en cada momento del tiempo. La propiedad $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ para cada n significa que la información aumenta con el paso del tiempo.

Definición 30. *Martingala, sub y supermartingala*

Dado un espacio de probabilidad con filtración $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$, y una sucesión de variables aleatorias $X = (X_0, X_1, X_2, \dots)$, diremos que X es una *martingala* si

- (a) $\mathbf{E} |X_n| < \infty$ para $n = 0, 1, \dots$
- (b) X_n es \mathcal{F}_n medible para $n = 0, 1, \dots$
- (c) $\mathbf{E}(X_{n+1} | \mathcal{F}_n) = X_n$ \mathbf{P} -c.s. para $n = 0, 1, \dots$

Diremos que X es una *submartingala* o una *supermartingala* cuando valgan (a), (b), y respectivamente (d) o (e):

- (d) $\mathbf{E}(X_{n+1} | \mathcal{F}_n) \geq X_n$ \mathbf{P} -c.s.
- (e) $\mathbf{E}(X_{n+1} | \mathcal{F}_n) \leq X_n$ \mathbf{P} -c.s.

Observaciones.

1. Es claro que si (X_1, X_2, \dots) es una submartingala, entonces la sucesión $(-X_1, -X_2, \dots)$ es una supermartingala. Esto permite considerar para el estudio únicamente las submartingalas, por ejemplo.
2. En caso de cumplirse (b), se dice también que X es adaptado a \mathcal{F} .
3. Cuando sea necesario destacar la filtración, diremos que X es una \mathcal{F} martingala o una martingala con respecto a \mathcal{F} (idem sub o supermartingala).

Ejemplo 27. Filtración natural. Es frecuente construir una filtración a partir de una sucesión de variables aleatorias X_0, X_1, \dots en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ de la siguiente forma. Definimos $\mathcal{F}_0 = \sigma(X_0)$, $\mathcal{F}_1 = \sigma(X_0, X_1)$, \dots , $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$, donde $\sigma(X_0, X_1, \dots, X_n)$ es la σ -álgebra generada por las variables X_0, X_1, \dots, X_n (ver pg. 91). En ese caso, la filtración $\mathcal{F} = (\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \dots)$ se llamará filtración natural de la sucesión X . Obsérvese que, en caso de considerar la filtración natural para una sucesión X , la condición (b) se cumple automáticamente.

Ejemplo 28. Suma de variables independientes Sea $X_0, X_1, \dots, X_n, \dots$ una sucesión de v.a.i. en $(\Omega, \mathcal{F}, \mathbf{P})$ con $\mathbf{E}|X_n| < \infty$ para $n = 0, 1, \dots$. Notemos $S_0 = X_0$, $S_n = X_0 + \dots + X_n$, y $S = (S_0, S_1, \dots)$. Sabemos que $\mathbf{E}|S_n| \leq \sum_{k=0}^n \mathbf{E}|X_k|$, y de las propiedades de la esperanza condicional resulta

$$\mathbf{E}(S_{n+1} | S_n, \dots, S_0) = \mathbf{E} X_{n+1} + S_n.$$

Luego, respecto de la filtración natural de S , tenemos

- Si $\mathbf{E} X_n = 0$ para todo n , S es una martingala.
- Si $\mathbf{E} X_n \geq 0$ para todo n , S es una submartingala.
- Si $\mathbf{E} X_n \leq 0$ para todo n , S es una supermartingala.

Ejemplo 29. Sea Z una variable aleatoria definida en un espacio de probabilidad con filtración $(\Omega, \mathcal{F}, \mathcal{F}, \mathbf{P})$. Consideremos para cada n , $X_n = \mathbf{E}(Z | \mathcal{F}_n)$. Por la propiedad telescópica (iv) del teorema 27, como $\mathcal{F}_n \subset \mathcal{F}_{n+1}$

$$\mathbf{E}(X_{n+1} | \mathcal{F}_n) = \mathbf{E}(\mathbf{E}(Z | \mathcal{F}_{n+1}) | \mathcal{F}_n) = \mathbf{E}(Z | \mathcal{F}_n) = X_n,$$

luego la sucesión X_1, X_2, \dots es una \mathcal{F} martingala.

Ejemplo 30. Martingalas y funciones convexas.

(a) Sea (X_0, X_1, \dots) una martingala en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ y $\phi: \mathbf{R} \rightarrow \mathbf{R}$ una función convexa. Supongamos que $\mathbf{E}|\phi(X_n)| < \infty$ para $n = 0, 1, \dots$. Por la desigualdad de Jensen

$$\mathbf{E}(\phi(X_{n+1}) | \mathcal{F}_n) \geq \phi(\mathbf{E}(X_{n+1} | \mathcal{F}_n)) = \phi(X_n),$$

luego $(\phi(X_0), \phi(X_1), \dots)$ es una \mathbf{F} submartingala.

(b) Sea (X_0, X_1, \dots) una submartingala en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ y $\phi: \mathbf{R} \rightarrow \mathbf{R}$ una función convexa y creciente, con $\mathbf{E}|\phi(X_n)| < \infty$ para $n = 0, 1, \dots$. Entonces,

$$\mathbf{E}(\phi(X_{n+1}) | \mathcal{F}_n) \geq \phi(\mathbf{E}(X_{n+1} | \mathcal{F}_n)) \geq \phi(X_n),$$

y $(\phi(X_0), \phi(X_1), \dots)$ es una \mathbf{F} submartingala.

Ejemplo 31. Martingalas y apuestas

El siguiente ejemplo relaciona las martingalas con los juegos de azar. Sean X_1, X_2, \dots variables aleatorias independientes con distribución común de Bernoulli simétrica, es decir $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$. Un jugador que apuesta b al resultado de X_k recibe $2b$ si $X_k = 1$ y 0 si $X_k = -1$. es decir, su capital aumenta en b si gana y disminuye en b si pierde. Si apuesta teniendo en cuenta los resultados obtenidos, es decir, su apuesta para $n + 1$ es $b_n = b_n(X_1, \dots, X_n)$. Su capital $\{Y_n\}$ cumplirá la relación

$$Y_{n+1} = Y_n + X_{n+1}b_n(X_1, \dots, X_n). \quad (5.10)$$

Si comienza el juego con un capital Y_0 , en tiempo n tendrá

$$Y_n = Y_0 + \sum_{k=0}^{n-1} (Y_{k+1} - Y_k) = Y_0 + \sum_{k=0}^{n-1} X_{k+1}b_n(X_1, \dots, X_k),$$

que es una función de X_1, \dots, X_n , y permite calcular, de (5.10)

$$\mathbf{E}(Y_{n+1} | X_1 \dots X_n) = Y_n + b_n(X_1, \dots, X_n) \mathbf{E} X_{n+1} = Y_n.$$

es decir el capital $\{Y_n\}$ es una martingala respecto de su filtración natural para *cualquier* estrategia b_n . Es decir, no hay estrategia que permita aumentar el capital esperado para el tiempo $n + 1$. Estamos en presencia de un *juego justo*.

El siguiente resultado estudia equivalencias de la propiedad (c) en la definición de martingala.

Teorema 30. Sea X_1, X_2, \dots una sucesión de variables aleatorias definidas en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$, que cumplen (a) y (b) en la definición 30 de martingala. Son equivalentes:

$$(c) \quad \mathbf{E}(X_{n+1} | \mathcal{F}_n) = X_n \quad \mathbf{P}\text{-c.s. para } n = 0, 1, \dots$$

$$(c_1) \quad \mathbf{E}(X_n | \mathcal{F}_m) = X_m, \quad \mathbf{P}\text{-c.s. para todo } m < n.$$

$$(c_2) \quad \int_{\mathbf{A}} X_n d\mathbf{P} = \int_{\mathbf{A}} X_m d\mathbf{P} \quad \text{para todo } \mathbf{A} \in \mathcal{F}_m \text{ con } m < n.$$

Demostración. Claramente $(c_1) \Rightarrow (c)$. Veamos $(c) \Rightarrow (c_1)$. Sea $m < n$.

$$\begin{aligned} \mathbf{E}(X_n | \mathcal{F}_m) &= \mathbf{E}(\mathbf{E}(X_n | \mathcal{F}_{n-1}) | \mathcal{F}_m) = \mathbf{E}(X_{n-1} | \mathcal{F}_m) \\ &= \dots = \mathbf{E}(X_{m+1} | \mathcal{F}_m) = X_m. \end{aligned}$$

Veamos que $(c_1) \Leftrightarrow (c_2)$. En efecto, (c_2) vale si y solo si

$$\int_{\mathbf{A}} \mathbf{E}(X_n | \mathcal{F}_m) d\mathbf{P} = \int_{\mathbf{A}} X_m d\mathbf{P} \quad \text{para todo } \mathbf{A} \in \mathcal{F}_m, \quad m < n. \quad (5.11)$$

y concluimos por (c) en el Lema 8, que (5.11) es equivalente a (c_1) . \square

Observación. Un resultado análogo vale para las propiedades (d) y (e) en la definición de sub y supermartingalas.

Definición 31 (Tiempos de parada). Sea $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ un espacio de probabilidad con filtración.

(a) Una variable aleatoria $\tau: \Omega \rightarrow \{0, 1, \dots, \infty\}$ es un tiempo de parada si el conjunto $\{\tau \leq n\}$ pertenece a \mathcal{F}_n para cada $n = 0, 1, \dots$

(b) Dado un tiempo de parada τ notamos por

$$\mathcal{F}_\tau = \{\mathbf{A} \in \mathcal{F}: \mathbf{A} \cap \{\tau \leq n\} \in \mathcal{F}_n \text{ para cada } n = 0, 1, \dots\},$$

que representa la σ -álgebra de los sucesos que ocurren hasta el momento τ .

Observaciones.

1. Como $\{\tau \leq n\} = \cup_{k=0}^n \{\tau = k\}$, para verificar que τ es tiempo de parada es equivalente ver que $\{\tau = n\} \in \mathcal{F}_n$ para $n = 0, 1, \dots$, y para verificar que $\mathbf{A} \in \mathcal{F}_\tau$ ver que $\mathbf{A} \cap \{\tau = n\} \in \mathcal{F}_n$ para $n = 0, 1, \dots$
2. En la proposición siguiente se verifica que \mathcal{F}_τ es una σ -álgebra.
3. Si $\tau = N$ con $N \in \mathbf{N}$ entonces $\{\tau \leq n\} = \emptyset$ si $n < N$ y $\{\tau \leq n\} = \Omega$ si $n \geq N$, verificándose la definición (a). Los tiempos de parada *generalizan* los tiempos constantes.

Ejemplo 32. El ejemplo fundamental de tiempo de parada es el siguiente. Sea $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ y X_1, X_2, \dots una sucesión \mathcal{F} adaptada. Sea \mathbf{B} un boreliano de \mathbf{R} .

$$\tau_{\mathbf{B}} = \inf\{n \geq 0: X_n \in \mathbf{B}\}$$

es un tiempo de parada, porque $\{\tau_{\mathbf{B}} = n\} = \{X_0 \notin \mathbf{B}, \dots, X_{n-1} \notin \mathbf{B}, X_n \in \mathbf{B}\}$ es \mathcal{F}_n medible.

Dadas $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ y $X = (X_1, X_2, \dots)$ una sucesión \mathcal{F} adaptada, introducimos la variable aleatoria

$$X_\tau = \sum_{n=0}^{\infty} X_n \mathbf{1}_{\{\tau=n\}},$$

con la convención $X_\infty = 0$, que representa el valor de la sucesión X en el instante τ .

Proposición 8. Sea X_1, X_2, \dots definido en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ y adaptada. Sean σ y τ tiempos de parada.

- (a) \mathcal{F}_τ es una σ -álgebra.
- (b) Si $\sigma \leq \tau$ entonces $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.
- (c) $\sigma \wedge \tau$ es tiempo de parada.
- (d) La variable X_τ es \mathcal{F}_τ medible.

Demostración. (a) Sea $\mathbf{A} \in \mathcal{F}_\tau$. El suceso $\mathbf{A}^c \cap \{\tau \leq n\} = (\mathbf{A} \cap \{\tau \leq n\})^c \cap \{\tau \leq n\} \in \mathcal{F}_n$, por pertenecer a ésta σ -álgebra los conjuntos $\mathbf{A} \cap \{\tau \leq n\}$ y $\{\tau \leq n\} \in \mathcal{F}_n$. Si $\mathbf{A}_1, \mathbf{A}_2, \dots$ es una sucesión finita o infinita en \mathcal{F}_τ , $(\cup_k \mathbf{A}_k) \cap \{\tau \leq n\} = \cup_k (\mathbf{A}_k \cap \{\tau \leq n\}) \in \mathcal{F}_n$. Luego \mathcal{F}_τ es σ -álgebra.

(b) Sea $\mathbf{A} \in \mathcal{F}_\sigma$. Como $\sigma \leq \tau$, tenemos $\{\tau \leq n\} \subset \{\sigma \leq n\}$. Entonces

$$\mathbf{A} \cap \{\tau \leq n\} = \mathbf{A} \cap \{\sigma \leq n\} \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \text{para } n = 0, 1, \dots,$$

por pertenecer a \mathcal{F}_n los conjuntos $\mathbf{A} \cap \{\sigma \leq n\}$ y $\{\tau \leq n\}$.

(c) $\{\tau \wedge \sigma \leq n\} = \{\tau \leq n\} \cup \{\sigma \leq n\} \in \mathcal{F}_n$ para cada n .

(d) Verifiquemos $\{X_\tau \leq x\} \in \mathcal{F}_\tau$. Eso equivale a $\{X_\tau \leq x\} \cap \{\tau = n\} = \{X_n \leq x\} \cap \{\tau = n\} \in \mathcal{F}_n$, por ser X una sucesión \mathbf{F} adaptada. \square

5.4. Teorema del muestreo opcional

Una idea clave en la teoría de las martingalas, es la sustitución de los tiempos determinísticos por tiempos de parada. Así definimos \mathcal{F}_τ y X_τ . Los siguientes teoremas (del muestreo opcional) dan condiciones para realizar esta sustitución en la definición de martingala.

Teorema 31. *Sea X_1, X_2, \dots una martingala en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$. Sean σ y τ tiempos de parada tales que, para $N \in \mathbf{N}$ se cumple $0 \leq \sigma \leq \tau \leq N$, \mathbf{P} -c.s. Entonces*

$$(a) \quad \mathbf{E}(X_\tau | \mathcal{F}_\sigma) = X_\sigma, \mathbf{P}\text{-c.s.}$$

$$(b) \quad \mathbf{E} X_0 = \mathbf{E} X_\sigma = \mathbf{E} X_\tau = \mathbf{E} X_N.$$

Observación. De la demostración se puede ver que el teorema anterior vale cuando X es una submartingala, con las \geq en (a) y \leq en (b).

Demostración. Observemos primero que, como $X_\tau = \sum_{k=0}^N X_k \mathbf{1}_{\{\tau=k\}}$ obtenemos que X_τ es integrable, y análogamente, que X_σ también lo es. (b) se obtiene de tomar esperanza en (a) en las parejas de tiempos de parada 0 y σ , σ y τ , τ y N . Veamos entonces (a). Para eso verificamos primero

$$(a_0) \quad \mathbf{E}(X_N | \mathcal{F}_\sigma) = X_\sigma, \mathbf{P}\text{-c.s.}$$

Como X_σ es \mathcal{F}_σ medible, es suficiente verificar

$$\int_{\mathbf{A}} X_N d\mathbf{P} = \int_{\mathbf{A}} X_\sigma d\mathbf{P} \quad \text{para todo } \mathbf{A} \in \mathcal{F}_\sigma.$$

Tenemos, si $\mathbf{A} \in \mathcal{F}_\sigma$, $\mathbf{A} \cap \{\sigma = k\} \in \mathcal{F}_k$ para $k = 0, \dots, N$, y

$$\int_{\mathbf{A}} X_N d\mathbf{P} = \sum_{k=0}^N \int_{\mathbf{A} \cap \{\sigma=k\}} X_N d\mathbf{P} = \sum_{k=0}^N \int_{\mathbf{A} \cap \{\sigma=k\}} X_k d\mathbf{P} = \int_{\mathbf{A}} X_\sigma d\mathbf{P}, \quad (5.12)$$

por ser X martingala, probando (a₀). Ahora, por la propiedad telescópica, como $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$, y (a₀) dan

$$\mathbf{E}(X_\tau | \mathcal{F}_\sigma) = \mathbf{E}(\mathbf{E}(X_N | \mathcal{F}_\tau) | \mathcal{F}_\sigma) = \mathbf{E}(X_N | \mathcal{F}_\sigma) = X_\sigma.$$

□

Observación. El resultado anterior, así como el teorema siguiente, vale para submartingalas, poniendo \geq en vez de $=$ en (a) y (b), así como en la demostración.

La extensión de este resultado para tiempos de parada no acotados exige hipótesis adicionales. Por ejemplo, si $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, con X_1, X_2, \dots v.a.i.i.d. con $\mathbf{P}(X_n = 1) = \mathbf{P}(X_n = -1) = \frac{1}{2}$, entonces (S_0, S_1, \dots) es una martingala respecto de su filtración natural. Si

$$\tau = \inf\{n \geq 0: S_n = 1\},$$

se tiene $\mathbf{P}(\tau < \infty) = 1$ (ver Ejemplo 8.6 en [3]). Sin embargo, con $\sigma = 0$,

$$0 = \mathbf{E} S_\sigma \neq \mathbf{E} S_\tau = 1,$$

y (b) en el Teorema 31 no se cumple. En su forma general, el teorema del muestreo opcional es el siguiente.

Teorema 32. *Sea $X = (X_1, X_2, \dots)$ una martingala en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$. Sean σ y τ tiempos de parada tales que se cumple $0 \leq \sigma \leq \tau < \infty$, \mathbf{P} -c.s. Supongamos que*

- (i) $\mathbf{E} |X_\sigma| < \infty$, $\mathbf{E} |X_\tau| < \infty$.
- (ii) $\liminf_n \mathbf{E} |X_n| \mathbf{1}_{\{\tau > n\}} = 0$.

Entonces

- (a) $\mathbf{E}(X_\tau | \mathcal{F}_\sigma) = X_\sigma$, \mathbf{P} -c.s.
- (b) $\mathbf{E} X_0 = \mathbf{E} X_\sigma = \mathbf{E} X_\tau$.

Demostración. Como en el teorema anterior, (b) se obtiene a partir de (a). Comencemos verificando

$$(a_1) \mathbf{E}(X_N | \mathcal{F}_\sigma) = X_{N \wedge \sigma}, \mathbf{P}\text{-c.s.}$$

En efecto, como $X_{N \wedge \sigma}$ es $\mathcal{F}_{N \wedge \sigma}$ medible, y $\mathcal{F}_{N \wedge \sigma} \subset \mathcal{F}_\sigma$, $X_{N \wedge \sigma}$ es \mathcal{F}_σ medible. Resta verificar que, si $\mathbf{A} \in \mathcal{F}_\sigma$

$$\int_{\mathbf{A}} X_N d\mathbf{P} = \int_{\mathbf{A}} X_{N \wedge \sigma} d\mathbf{P},$$

y para esto

$$\int_{\mathbf{A}} X_{N \wedge \sigma} d\mathbf{P} = \sum_{k=0}^N \int_{\mathbf{A} \cap \{\sigma=k\}} X_k d\mathbf{P} + \sum_{k=N+1}^{\infty} \int_{\mathbf{A} \cap \{\sigma=k\}} X_N d\mathbf{P} = \int_{\mathbf{A}} X_N d\mathbf{P},$$

porque $\int_{\mathbf{A} \cap \{\sigma=k\}} X_k d\mathbf{P} = \int_{\mathbf{A} \cap \{\sigma=k\}} X_N d\mathbf{P}$, ya que $\mathbf{A} \cap \{\sigma = k\} \in \mathcal{F}_k$, y X es martingala. Aplicando (a₁), como $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$

$$X_{N \wedge \sigma} = \mathbf{E}(X_N | \mathcal{F}_\sigma) = \mathbf{E}(\mathbf{E}(X_N | \mathcal{F}_\tau | \mathcal{F}_\sigma)) = \mathbf{E}(X_{N \wedge \tau} | \mathcal{F}_\sigma). \quad (5.13)$$

Tomamos límite si $N \rightarrow \infty$.

$$\mathbf{P}(|X_{N \wedge \sigma} - X_\sigma| \neq 0) \leq \mathbf{P}(\sigma > N) \rightarrow 0 \quad \text{si } N \rightarrow \infty.$$

Por otra parte

$$\begin{aligned} \mathbf{E} |\mathbf{E}(X_{N \wedge \tau} | \mathcal{F}_\sigma) - \mathbf{E}(X_\tau | \mathcal{F}_\sigma)| &\leq \mathbf{E} |X_{N \wedge \tau} - X_\tau| \\ &\leq \mathbf{E} |X_\tau| \mathbf{1}_{\{\tau > N\}} + \mathbf{E} |X_N| \mathbf{1}_{\{\tau > N\}}. \end{aligned}$$

Como $\mathbf{E} |X_\tau| \mathbf{1}_{\{\tau > N\}} \rightarrow 0$ si $N \rightarrow \infty$ por (i), y por (ii), existe una subsucesión $\{N_j\}$ tal que $\mathbf{E} |X_{N_j}| \mathbf{1}_{\{\tau > N_j\}} \rightarrow 0$, si $j \rightarrow \infty$, obtenemos

$$\mathbf{E}(X_{N_j \wedge \tau} | \mathcal{F}_\sigma) \rightarrow \mathbf{E}(X_\tau | \mathcal{F}_\sigma)$$

en media, y en probabilidad. La unicidad del límite en probabilidad en (5.13) da (a). \square

Ejemplo 33. Problema de barrera para el paseo al azar simple Sea $S_0 = 0$, $S_n = X_1 + \dots + X_n$ con X_1, X_2, \dots v.a.i.i.d. de Bernoulli, $\mathbf{P}(X_n = 1) = p$, $\mathbf{P}(X_n = -1) = q$, $p + q = 1$, $0 < p < 1$. Sea A natural y positivo, y

$$\tau_A = \inf\{n \geq 0: S_n = A\}.$$

Queremos calcular $\mathbf{P}(\tau_A < \infty)$, es decir, la probabilidad de que el paseo al azar llegue a la barrera A .

Si $p > q$, tenemos, con $\delta = p - q > 0$ y N tal que $\delta - \frac{A}{N} > \frac{\delta}{2}$,

$$\mathbf{P}(\tau_A = \infty) = \mathbf{P}(S_n < A \forall n) \leq \mathbf{P}\left(\left|\frac{S_N}{N} - \delta\right| > \frac{\delta}{2}\right),$$

luego

$$\mathbf{P}(\tau_A = \infty) \leq \lim_N \mathbf{P}\left(\left|\frac{S_N}{N} - \delta\right| > \frac{\delta}{2}\right) = 0,$$

y $\mathbf{P}(\tau_A < \infty) = 1$.

Si $p = q$ el paseo al azar es recurrente, y $\mathbf{P}(\tau_A = \infty) = \mathbf{P}(X_n \neq A \forall n) = 0$

Resta el caso $p < q$. Del ejemplo 26 sabemos que $\left\{\left(\frac{q}{p}\right)^{S_n}\right\}_{n \geq 0}$ es martingala, luego por el teorema 31

$$1 = \mathbf{E} S_0 = \mathbf{E} \left(\frac{q}{p}\right)^{S_{\tau_A \wedge N}}. \quad (5.14)$$

Como $\lim S_n = -\infty$, **P**-c.s., si $N \rightarrow \infty$, tenemos

$$\left(\frac{q}{p}\right)^{S_{\tau_A \wedge N}} = \left(\frac{q}{p}\right)^{S_{\tau_A \wedge N}} \mathbf{1}_{\{\tau_A < \infty\}} + \left(\frac{q}{p}\right)^{S_{\tau_A \wedge N}} \mathbf{1}_{\{\tau_A = \infty\}} \rightarrow \left(\frac{q}{p}\right)^A \mathbf{1}_{\{\tau_A < \infty\}}.$$

Tomando límite en (5.14), obtenemos

$$\mathbf{P}(\tau_A < \infty) = \left(\frac{p}{q}\right)^A.$$

Ejemplo 34. Problema de dos barreras para el paseo al azar simple.

Sea como en el ejemplo anterior $S_0 = 0$, $S_n = X_1 + \dots + X_n$ con X_1, X_2, \dots v.a.i.i.d. de Bernoulli, $\mathbf{P}(X_n = 1) = p$, $\mathbf{P}(X_n = -1) = q$, $p + q = 1$, $0 < p < 1$. Sean A, B naturales positivos, y

$$\tau_A = \inf\{n \geq 0: S_n = A\}, \quad \tau_B = \inf\{n \geq 0: S_n = -B\}.$$

Nos interesa calcular $\mathbf{P}(\tau_A < \tau_B)$, es decir, la probabilidad de que el paseo al azar llegue a A antes de llegar a $-B$. Como vimos en el ejemplo anterior, si $p \leq q$ tenemos $\mathbf{P}(\tau_B < \infty) = 1$, y si $p \geq q$ entonces $\mathbf{P}(\tau_A < \infty) = 1$. Luego, si $\tau_0 = \inf(\tau_A, \tau_B)$, tenemos $\mathbf{P}(\tau_0 < \infty) = 1$, cualquiera sean p y q .

Interpretamos este modelo como un juego de apuestas, sucesivas entre dos jugadores I con un capital A y II con un capital B . El que pierde paga una unidad al otro. $B + S_n$ representa el capital del jugador II. Como $\mathbf{P}(\tau_0 < \infty) = 1$, el juego se termina con la ruina de alguno de los dos jugadores. $\mathbf{P}(\tau_A < \tau_B)$ es la probabilidad de que gane II. Como $\left\{\left(\frac{q}{p}\right)^{S_n}\right\}_{n \geq 0}$ es martingala, y los tiempos de parada $\sigma = 0$ y τ_0 cumplen (i) y (ii) en el teorema 32

$$1 = \mathbf{E} S_0 = \mathbf{E} \left(\frac{q}{p}\right)^{S_{\tau_A \wedge \tau_B}} = \left(\frac{q}{p}\right)^A \mathbf{P}(\tau_A < \tau_B) - \left(\frac{p}{q}\right)^B \mathbf{P}(\tau_B < \tau_A). \quad (5.15)$$

De aquí, como $\mathbf{P}(\tau_A < \tau_B) + \mathbf{P}(\tau_B < \tau_A) = 1$,

$$\mathbf{P}(\tau_A < \tau_B) = \frac{1 - (q/p)^{-B}}{(q/p)^A - (q/p)^{-B}}.$$

Una importante aplicación de los teoremas del muestreo opcional son las desigualdades que dan información sobre el máximo de una (sub) martingala, conocidas como desigualdades maximales, de las cuales estudiaremos las mas elementales.

Teorema 33. *Sea X_1, X_2, \dots una submartingala en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$. Entonces, para $\lambda > 0$ vale*

- (a) $\lambda \mathbf{P}(\max_{0 \leq k \leq n} X_k \geq \lambda) \leq \mathbf{E} X_n^+$.
- (b) $\lambda \mathbf{P}(\min_{0 \leq k \leq n} X_k \leq -\lambda) \leq \mathbf{E} X_n^+ - \mathbf{E} X_0$.
- (c) $\lambda \mathbf{P}(\max_{0 \leq k \leq n} |X_k| \geq \lambda) \leq 3 \max_{0 \leq k \leq n} \mathbf{E} |X_k|$.

Demostración. (a) Sea $\tau = \inf\{k \geq 0: X_k \geq \lambda\} \wedge n$. Como $\tau \leq n$, por el teorema 31

$$\begin{aligned} \mathbf{E} X_n &\geq \mathbf{E} X_\tau = \mathbf{E} X_\tau \mathbf{1}_{\{\max_{0 \leq k \leq n} X_k \geq \lambda\}} + \mathbf{E} X_\tau \mathbf{1}_{\{\max_{0 \leq k \leq n} X_k < \lambda\}} \\ &\geq \lambda \mathbf{P}(\max_{0 \leq k \leq n} X_k \geq \lambda) + \mathbf{E} X_n \mathbf{1}_{\{\max_{0 \leq k \leq n} X_k < \lambda\}}. \end{aligned}$$

Entonces

$$\begin{aligned} \lambda \mathbf{P}(\max_{0 \leq k \leq n} X_k \geq \lambda) &\leq \mathbf{E} X_n \mathbf{1}_{\{\max_{0 \leq k \leq n} X_k \geq \lambda\}} \\ &\leq \mathbf{E} X_n^+ \mathbf{1}_{\{\max_{0 \leq k \leq n} X_k < \lambda\}} \leq \mathbf{E} X_n^+. \end{aligned}$$

- (b) Sea $\tau = \inf\{k \geq 0: X_k < -\lambda\} \wedge n$.

$$\begin{aligned} \mathbf{E} X_0 &\leq \mathbf{E} X_\tau \leq \mathbf{E} X_\tau \mathbf{1}_{\{\inf_{0 \leq k \leq n} X_k \leq -\lambda\}} + \mathbf{E} X_\tau \mathbf{1}_{\{\inf_{0 \leq k \leq n} X_k > -\lambda\}} \\ &\leq -\lambda \mathbf{P}(\min_{0 \leq k \leq n} X_k \leq -\lambda) + \mathbf{E} X_n^+. \end{aligned}$$

y de aquí resulta (b).

- (c) Por (a) y (b)

$$\begin{aligned} \lambda \mathbf{P}(\max_{0 \leq k \leq n} |X_k| \geq \lambda) &\leq \lambda \mathbf{P}(\max_{0 \leq k \leq n} X_k \geq \lambda) + \lambda \mathbf{P}(\min_{0 \leq k \leq n} X_k \leq -\lambda) \\ &\leq 2 \mathbf{E} X_n^+ - \mathbf{E} X_0 \leq 3 \max_{0 \leq k \leq n} \mathbf{E} |X_k|. \end{aligned}$$

□

Como caso particular de (a) obtenemos el resultado clásico de Kolmogorov.

Corolario 3. *Desigualdad de Kolmogorov*

Sean X_1, \dots, X_n v.a.i. con $\mathbf{E} X_k = 0$ y $\mathbf{E} X_k^2 < \infty$ para $k = 1, \dots, n$. Sea $S_k = X_1 + \dots + X_k$, $k = 1, \dots, n$. Entonces

$$\mathbf{P}(\max_{1 \leq k \leq n} |S_k| \geq \lambda) \leq \frac{1}{\lambda^2} \mathbf{E} S_n^2 = \frac{1}{\lambda^2} \sum_{k=1}^n \mathbf{E} X_k^2.$$

Demostración. Como (S_1, S_2, \dots) es una martingala, como en el ejemplo 30 obtenemos que (S_1^2, S_2^2, \dots) es submartingala. Por (a) en el teorema anterior

$$\mathbf{P}(\max_{1 \leq k \leq n} |S_k| \geq \lambda) = \mathbf{P}(\max_{1 \leq k \leq n} |S_k|^2 \geq \lambda^2) \leq \frac{1}{\lambda^2} \mathbf{E} S_n^2.$$

Por último, $\mathbf{E} S_n^2 = \mathbf{E}(X_1 + \dots + X_n)^2 = \sum_{k=1}^n \mathbf{E} X_k^2$, porque $\mathbf{E} X_i X_j = 0$ si $i \neq j$. \square

5.5. Convergencia casi segura de Martingalas

En esta sección y la siguiente estudiamos el comportamiento asintótico de una submartingala. El resultado central es el siguiente.

Teorema 34 (de convergencia de Doob). *Sea X_1, X_2, \dots una submartingala en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$. Supongamos que*

$$\sup_n \mathbf{E} |X_n| < \infty.$$

Entonces, existe $X_\infty = \lim_n X_n$, \mathbf{P} -c.s. Además $\mathbf{E} |X_\infty| \leq \sup_n \mathbf{E} |X_n|$.

La demostración se basa en la siguiente desigualdad, que controla la oscilación de una submartingala. Fijemos $a < b$, reales, y definimos para una sucesión X_1, X_2, \dots , $\tau_0 = \sigma_0 = 0$, y en forma inductiva

$$\begin{aligned} \sigma_1 &= \inf\{k \geq 0: X_k \leq a\} \\ \tau_1 &= \inf\{k \geq \sigma_1: X_k \geq b\} \\ &\vdots \\ \sigma_n &= \inf\{k \geq \tau_{n-1}: X_k \leq a\} \\ \tau_n &= \inf\{k \geq \sigma_n: X_k \geq b\} \end{aligned}$$

La cantidad de cruces de la sucesión X de valores menores que a a mayores que b , antes de N está dada por

$$U_n(a, b) = \max\{k \geq 0: \tau_k \leq n\}. \quad (5.16)$$

Como $U_n(a, b)$ es creciente con n , podemos definir además $U(a, b) = \lim_n U_n(a, b)$.

Teorema 35 (Desigualdad de cruces de Doob). *Sea X_1, X_2, \dots una submartingala definida en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$. Entonces,*

$$\mathbf{E} U_n(a, b) \leq \frac{\mathbf{E}(X_n - a)^+}{b - a},$$

con $a < b$ reales y $U_n(a, b)$ definido en (5.16).

Demostración. Observemos primero, que con $Y_0 = 0$ y $Y_n = (X_n - a)^+$, para $n = 1, 2, \dots$, el proceso Y_1, Y_2, \dots es una submartingala.

Para contar los cruces, asociamos a cada tiempo n un uno si la sucesión este pasando de a hacia b , o un en caso contrario. Es decir, para $k = 0, 1, 2, \dots$, introducimos las variables auxiliares,

$$\varphi_j = \begin{cases} 1, & \text{si } j \in [\sigma_k, \tau_k) \text{ para algún } k \leq n, \\ 0, & \text{en otro caso.} \end{cases}$$

La variable aleatoria φ_{k-1} depende de las variables X_0, X_1, \dots, X_{k-1} por lo que es \mathcal{F}_{k-1} medible. Además

$$(b - a)U_n(a, b) \leq \sum_{k=1}^n \varphi_{k-1}(Y_k - Y_{k-1}).$$

Como Y es submartingala, y φ_{k-1} es \mathcal{F}_{k-1} medible,

$$\mathbf{E} \varphi_{k-1}(Y_k - Y_{k-1}) = \mathbf{E} \varphi_{k-1} \mathbf{E}(Y_k - Y_{k-1} | \mathcal{F}_{k-1}) \leq \mathbf{E}(Y_k - Y_{k-1}).$$

Entonces

$$(b - a) \mathbf{E} U_n(a, b) \leq \sum_{k=1}^n \mathbf{E}(Y_k - Y_{k-1}) = \mathbf{E} Y_n = \mathbf{E}(X_n - a)^+,$$

concluyendo la prueba del Lema de cruces. \square

Demostración. Veamos que

$$\mathbf{P}(\not\exists \lim_n X_n) = 0. \tag{5.17}$$

Para eso

$$\begin{aligned} \mathbf{P}(\not\exists \lim_n X_n) &= \mathbf{P}(\liminf_n X_n < \limsup_n X_n) = \\ &= \mathbf{P}(\cup \{ \liminf_n X_n \leq a < b \leq \limsup_n X_n : a < b, a, b \in \mathbf{Q} \}). \end{aligned}$$

Ahora, por convergencia monótona

$$\mathbf{E}U(a, b) = \lim_n \mathbf{E}U_n(a, b) \leq \lim_n \frac{\mathbf{E}(X_n - a)^+}{b - a} \leq \sup_n \frac{\mathbf{E}|X_n| + |a|}{b - a}.$$

Entonces $\mathbf{E}U(a, b) < \infty$, lo que da

$$\mathbf{P}(\cup_n \{\liminf_n X_n \leq a < b \leq \limsup_n X_n : a < b, a, b \in \mathbf{Q}\}) = \mathbf{P}(U(a, b) = \infty) = 0,$$

concluyendo (5.17). Sea $X_\infty(\omega) = \lim_n X_n(\omega)$ Respecto de (b) $|X_\infty| = \lim_n |X_n|$, y por el lema de Fatou

$$\mathbf{E}|X_\infty| = \mathbf{E} \lim_n |X_n| \leq \liminf_n \mathbf{E} \lim_n |X_n| \leq \sup_n \mathbf{E}|X_n|.$$

Ejemplo 35 (Lema de Borel Cantelli). Sea $(\Omega, \mathcal{F}, \mathbf{P})$, y $\mathbf{A}_1, \mathbf{A}_2, \dots$ una sucesión de conjuntos en \mathcal{F} . El lema de Borel Cantelli afirma que si

$$\sum_{n=1}^{\infty} \mathbf{P}(\mathbf{A}_n) < \infty,$$

entonces

$$\mathbf{P}(\mathbf{A}_n \text{ infinitas veces}) = \mathbf{P}(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} \mathbf{A}_k) = 0.$$

Sea $S_n = \sum_{k=1}^n \mathbf{1}_{\mathbf{A}_k}$. Tenemos

$$\{\omega : \mathbf{A}_n \text{ infinitas veces}\} = \{\omega : \lim_n S_n = \infty\}.$$

$\{S_n\}$ es una submartingala positiva, que cumple

$$\sup_n \mathbf{E}|S_n| = \sup_n \mathbf{E} \sum_{k=1}^n \mathbf{1}_{\mathbf{A}_k} = \sum_{n=1}^{\infty} \mathbf{P}(\mathbf{A}_n) < \infty,$$

de donde existe $S_\infty(\omega) = \lim_n S_n$, con $\mathbf{E}S_\infty < \infty$. De aquí $\sum_{n=1}^{\infty} \mathbf{1}_{\mathbf{A}_n} < \infty$ \mathbf{P} -c.s. y $\mathbf{P}(\mathbf{A}_n \text{ infinitas veces}) = 0$.

Ejemplo 36. Sean X_1, X_2, \dots v.a.i.i.d. con $\mathbf{P}(X_1 = 0) = \mathbf{P}(X_1 = 2) = 1/2$. Definimos $P_0 = 1$, $P_n = \prod_{k=1}^n X_k$. La sucesión $\{P_n\}$ es una martingala respecto de su filtración natural. Además $\sup_n \mathbf{E}|P_n| = 1$. Luego existe $P_\infty = \lim_n P_n$. En este caso, podemos conocer la ley de P_∞ .

$$\mathbf{P}(P_\infty \neq 0) = \mathbf{P}(\lim_n \prod_{k=1}^n X_k \neq 0) = \mathbf{P}(\cap_{k=1}^{\infty} \{X_k \neq 0\}) =$$

$$\lim_n \mathbf{P}(\cap_{k=1}^n \{X_k \neq 0\}) = \lim_n \frac{1}{2^n} = 0,$$

luego $\mathbf{P}(P_\infty = 0) = 1$. Observemos que

$$0 = \mathbf{E}P_\infty \neq \lim_n \mathbf{E}P_n = 1.$$

5.6. Integrabilidad uniforme y convergencia en media de martingalas

El ejemplo 36 muestra que la convergencia casi segura de una submartingala no conlleva necesariamente la convergencia de los valores esperados. Estudiaremos entonces condiciones que aseguren además de la convergencia puntual o casi segura, la convergencia en media.

Definición 32. *Integrabilidad uniforme*

La sucesión de variables aleatorias $X = (X_1, X_2, \dots)$ es uniformemente integrable si

$$\lim_{H \rightarrow \infty} \sup_n \int_{\{|X_n| > H\}} |X_n| d\mathbf{P} = 0. \quad (5.18)$$

Ejemplo 37. Sea Z v.a. en $(\Omega, \mathcal{F}, \mathbf{P})$ y $X_n = Z$ para todo n . Entonces

$$\lim_{H \rightarrow \infty} \sup_n \int_{\{|X_n| > H\}} |X_n| d\mathbf{P} = \lim_{H \rightarrow \infty} \int_{\{|Z| > H\}} |Z| d\mathbf{P} = 0$$

si y solo si $\mathbf{E}|Z| < \infty$. En este sentido 5.18 generaliza la integrabilidad de variables aleatorias.

Ejemplo 38. Sea $X = (X_1, X_2, \dots)$ una sucesión de variables aleatorias. Supongamos que existe Y v.a. no negativa, con $\mathbf{E}Y < \infty$ tal que $|X_n| \leq Y$ para todo n . Entonces,

$$\sup_n \int_{\{|X_n| > H\}} |X_n| d\mathbf{P} \leq \int_{\{Y > H\}} Y d\mathbf{P} = 0$$

y X es uniformemente integrable. El recíproco no es cierto, como se ve en el ejercicio 9.9 de [3].

Ejemplo 39. Sea Z una variable aleatoria en $(\Omega, \mathcal{F}, \mathbf{P})$ con $\mathbf{E}|Z| < \infty$. Definamos $X_n = \mathbf{E}(Z | \mathcal{F}_n)$, para $n = 0, 1, \dots$. Entonces (X_1, X_2, \dots) es uniformemente integrable. En efecto, dado $\varepsilon > 0$ sea α real tal que $\mathbf{E}|Z| \mathbf{1}_{\{|Z| > \alpha\}} < \varepsilon$.

$$\begin{aligned} \mathbf{E}|X_n| \mathbf{1}_{\{|X_n| > H\}} &= \mathbf{E}|\mathbf{E}(Z | \mathcal{F}_n)| \mathbf{1}_{\{|X_n| > H\}} \leq \mathbf{E}|Z| \mathbf{1}_{\{|X_n| > H\}} \\ &= \mathbf{E}|Z| \mathbf{1}_{\{|X_n| > H, |Z| > \alpha\}} + \mathbf{E}|Z| \mathbf{1}_{\{|X_n| > H, |Z| \leq \alpha\}} \\ &\leq \mathbf{E}|Z| \mathbf{1}_{\{|Z| > \alpha\}} + \alpha \mathbf{P}(|X_n| > H) \\ &\leq \varepsilon + \frac{\alpha}{H} \mathbf{E}|X_n| \\ &\leq \varepsilon + \frac{\alpha}{H} \mathbf{E}|Z| \rightarrow \varepsilon \end{aligned}$$

si $H \rightarrow \infty$.

Es inmediato ver que si (X_1, X_2, \dots) es uniformemente integrable, entonces $\sup_n \mathbf{E} |X_n| < \infty$. En efecto, si $\varepsilon = 1$ existe H_0 tal que $\int_{\{|X_n| > H_0\}} |X_n| d\mathbf{P} \leq 1$, y

$$\sup_n \mathbf{E} |X_n| = \sup_n \mathbf{E} |X_n| \mathbf{1}_{\{|X_n| \leq H_0\}} + 1 < \infty$$

El recíproco de este resultado no es cierto como se ve en el ejercicio 9.9 de [3]. Sin embargo, vale el siguiente resultado.

Teorema 36 (Criterio de La Vallée-Poussin). *Sea X_1, X_2, \dots una sucesión de variables aleatorias. Supongamos que existe una función $\Phi: [0, \infty) \rightarrow [0, \infty)$ creciente, con $\lim_{x \rightarrow \infty} \frac{\Phi(x)}{x} = \infty$ y tal que $\sup_n \mathbf{E} \Phi(|X_n|) < \infty$. Entonces X es uniformemente integrable.*

Demostración. Notemos $M = \sup_n \mathbf{E} \Phi(|X_n|)$. Sea $\varepsilon > 0$ Existe c tal que $\frac{\Phi(x)}{x} > \frac{M}{\varepsilon}$, si $x > c$. En el conjunto $\{\omega: |X_n| > c\}$ tenemos $|X_n| \leq \frac{\varepsilon}{M} \Phi(|X_n|)$, y, para cada n ,

$$\int_{\{|X_n| > c\}} |X_n| d\mathbf{P} \leq \frac{\varepsilon}{M} \int_{\{|X_n| > c\}} \Phi(|X_n|) d\mathbf{P} < \varepsilon,$$

luego X es uniformemente integrable. \square

Observación. Es interesante destacar que si X es uniformemente integrable, se puede construir una función Φ como la del teorema ³.

El siguiente resultado de convergencia muestra la utilidad de la noción introducida, y generaliza el teorema de convergencia dominada.

Teorema 37. *Sea X_1, X_2, \dots una sucesión uniformemente integrable de variables aleatorias en $(\Omega, \mathcal{F}, \mathbf{P})$, tal que existe $X_\infty = \lim_n X_n$, \mathbf{P} -c.s. Entonces*

- (a) $\mathbf{E} X_n \rightarrow \mathbf{E} X_\infty$ si $n \rightarrow \infty$.
- (b) $\mathbf{E} |X_n - X_\infty| \rightarrow 0$ si $n \rightarrow \infty$.

Demostración. Como $|\mathbf{E} X_n - \mathbf{E} X_\infty| \leq \mathbf{E} |X_n - X_\infty|$ (b) resulta de (a). Veamos (b). Sea $\alpha > 0$. Definamos $f_\alpha: \mathbf{R} \rightarrow \mathbf{R}$ mediante

$$f_\alpha(x) = \begin{cases} x, & \text{si } |x| \leq \alpha, \\ \alpha, & \text{si } x \geq \alpha, \\ -\alpha, & \text{si } x \leq -\alpha. \end{cases}$$

³ver Dellacherie et Meyer, *Probabilités et Potentiel*.

f_α es continua, y además

$$|x - f_\alpha(x)| \leq |x| \mathbf{1}_{|x| \geq \alpha}. \quad (5.19)$$

Como $X_\infty = \lim_n X_n$ y X es uniformemente integrable,

$$\mathbf{E}|X_\infty| \leq \sup \mathbf{E}|X_n| < \infty.$$

Dado $\varepsilon > 0$ elegimos α_0 real tal que

$$\sup_n \mathbf{E} \mathbf{1}_{\{|X_n| > \alpha_0\}} |X_n| < \frac{\varepsilon}{3}, \quad \mathbf{E} \mathbf{1}_{\{|X_\infty| > \alpha_0\}} |X_\infty| < \frac{\varepsilon}{3}.$$

Ahora, aplicando (5.19)

$$\begin{aligned} \mathbf{E}|X_n - X_\infty| &\leq \\ &\mathbf{E}|X_n - f_{\alpha_0}(X_n)| + \mathbf{E}|f_{\alpha_0}(X_n) - f_{\alpha_0}(X_\infty)| + \mathbf{E}|X_\infty - f_{\alpha_0}(X_\infty)| \\ &\leq \mathbf{E} \mathbf{1}_{\{|X_n| > \alpha_0\}} |X_n| + \mathbf{E} \mathbf{1}_{\{|X_\infty| > \alpha_0\}} |X_\infty| \\ &\quad + \mathbf{E}|f_{\alpha_0}(X_n) - f_{\alpha_0}(X_\infty)| < \varepsilon \end{aligned}$$

si $n \geq n_0$, donde n_0 es tal que $\mathbf{E}|f_{\alpha_0}(X_n) - f_{\alpha_0}(X_\infty)| \leq \frac{\varepsilon}{3}$ para $n \geq n_0$ dado que $f_{\alpha_0}(X_n) - f_{\alpha_0}(X_\infty) \rightarrow 0$, \mathbf{P} -c.s. y son variables uniformemente acotadas. \square

El resultado anterior nos permite completar el estudio sobre la convergencia de martingalas

Teorema 38. *Convergencia en media y casi segura de martingalas*

Sea X_1, X_2, \dots una submartingala uniformemente integrable. Entonces

- (a) *Existe $X_\infty = \lim_n X_n$, \mathbf{P} -c.s.*
- (b) $\mathbf{E}|X_n - X_\infty| \rightarrow 0$ si $n \rightarrow \infty$.
- (c) $\mathbf{E}(X_\infty | \mathcal{F}_n) \geq X_n$ para todo n .

Demostración. Como X es uniformemente integrable, $\sup_n \mathbf{E}|X_n| < \infty$, y el teorema 34 da (a). (b) se obtiene de (a) y el teorema 37. Veamos (c). Como X es submartingala, dado $\mathbf{A} \in \mathcal{F}_n$ $\{\int_{\mathbf{A}} X_k d\mathbf{P}\}_{k \geq n}$ es creciente. Por (b) $\lim_k \int_{\mathbf{A}} X_k d\mathbf{P} = \int_{\mathbf{A}} X_\infty d\mathbf{P}$. De allí,

$$\int_{\mathbf{A}} X_n d\mathbf{P} \leq \int_{\mathbf{A}} X_\infty d\mathbf{P} = \int_{\mathbf{A}} \mathbf{E}(X_\infty | \mathcal{F}_n) d\mathbf{P}$$

de donde sigue (c) \square

La propiedad (c) recién demostrada permite considerar martingalas y submartingalas con conjunto extendido de índices $\{0, 1, \dots, \infty\}$ donde X_∞ es una variable más y $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

Corolario 4 (Teorema de Lévy). *Sea Z variable aleatoria con $\mathbf{E}|Z| < \infty$, definida en $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$. Si $X_n = \mathbf{E}(Z | \mathcal{F}_n)$ para $n = 0, 1, \dots$ con $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$, tenemos*

- (a) $\mathbf{E}(Z | \mathcal{F}_\infty) = \lim_n X_n$, \mathbf{P} -c.s.
- (b) $\mathbf{E}|X_n - \mathbf{E}(Z | \mathcal{F}_\infty)| \rightarrow 0$ si $n \rightarrow \infty$.

Demostración. Como X_1, X_2, \dots es una martingala uniformemente integrable existe una variable aleatoria X_∞ con $X_\infty = \lim_n X_n$. Resta ver $X_\infty = \mathbf{E}(Z | \mathcal{F}_\infty)$. Como X_∞ es \mathcal{F}_∞ medible, y esta σ -álgebra se genera por $\cup_n \mathcal{F}_n$, es suficiente ver que ⁴, si $n \in \mathbf{N}$, y $\mathbf{A} \in \mathcal{F}_n$,

$$\int_{\mathbf{A}} X_\infty d\mathbf{P} = \int_{\mathbf{A}} Z d\mathbf{P}.$$

Si $m \geq n$ por (b), con $m \rightarrow \infty$

$$\int_{\mathbf{A}} X_m d\mathbf{P} \rightarrow \int_{\mathbf{A}} X_\infty d\mathbf{P}.$$

y

$$\int_{\mathbf{A}} X_m d\mathbf{P} = \int_{\mathbf{A}} \mathbf{E}(Z | \mathcal{F}_m) d\mathbf{P} = \int_{\mathbf{A}} Z d\mathbf{P},$$

porque $\mathbf{A} \in \mathcal{F}_n$ concluyendo la demostración. \square

⁴Nos basamos en el teorema de extensión de medidas de Charatédory.

Capítulo 6

El Proceso de Wiener

Estudiaremos *procesos estocásticos con tiempo continuo*, es decir, familias de variables aleatorias $X = (X_t)_{t \geq 0}$, definidas en un espacio de probabilidad común $(\Omega, \mathcal{F}, \mathbf{P})$, cuyo índice t toma valores en la semirrecta real no negativa $[0, \infty)$. Una manera alternativa de ver un proceso aleatorio de tiempo continuo es considerar fijo cada suceso elemental ω del espacio de estados Ω , obteniéndose una función $X(t, \omega)$ ¹. Una tal función se denomina *trayectoria* del proceso, y aquí supondremos que todas las trayectorias de los procesos considerados son continuas por la derecha con límites a la izquierda.

Diremos que un proceso estocástico X tiene *incrementos independientes* cuando para cualquier elección de índices $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n$, las variables aleatorias

$$X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}, \dots, X_{t_n} - X_{s_n} \text{ son mutuamente independientes. (6.1)}$$

Un proceso estocástico tiene *incrementos estacionarios* (también decimos *incrementos homogéneos en el tiempo*) cuando cualesquiera sean $t \geq 0$, y $h \geq 0$

$$\text{la distribución de } X_{t+h} - X_t \text{ es idéntica a la de } X_h. \quad (6.2)$$

6.1. Vectores gaussianos

Comenzamos recordando la definición de un vector normal (o gaussiano) multidimensional.

¹las variables aleatorias de un proceso estocástico X serán notadas indistintamente mediante X_t , $X_t(\omega)$ o $X(t, \omega)$.

Definición 33. (a) Una variable aleatoria X tiene distribución normal con media m y varianza σ^2 cuando tiene densidad dada por

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

(b) Dados un vector $m \in \mathbf{R}^n$ y una matriz $n \times n$ semidefinida positiva Γ , decimos que el vector aleatorio $X = (X_1, \dots, X_n)$ tiene distribución normal multidimensional con parámetros (m, Γ) , cuando para todo $\alpha \in \mathbf{R}^n$ se tiene²

$$\mathbf{E} e^{i\alpha^t X} = \exp\left(i\alpha^t m - \frac{1}{2}\alpha^t \Gamma \alpha\right) \quad (6.3)$$

(c) Decimos que el vector aleatorio $X = (X_1, \dots, X_n)$ tiene distribución normal multidimensional estándar cuando $m = 0$ y $\Gamma = I_n$ (matriz identidad $n \times n$).

Proposición 9. Son equivalentes:

- (a) el vector aleatorio $X = (X_1, \dots, X_n)$ tiene distribución normal multidimensional con parámetros $m = \mathbf{E} X$ y $\Gamma = \mathbf{cov} X = (\mathbf{cov}(X_i, X_j))_{1 \leq i, j \leq n}$,
- (b) para todo $\alpha \in \mathbf{R}^n$ el vector $\alpha^t X$ tiene distribución normal en \mathbf{R} .

Demostración. Vemos que (a) \Rightarrow (b) se obtiene de la fórmula (6.3) y el teorema de unicidad de las funciones características. Para ver (b) \Rightarrow (a), elegimos $\alpha \in \mathbf{R}^n$ fijo y arbitrario. Tenemos $\mathbf{E} \alpha^t X = \alpha^t \mathbf{E} X$, y $\mathbf{var}(\alpha^t X) = \sum_{i,j=1}^n \alpha_i \alpha_j \mathbf{cov}(X_i, X_j)$. Si definimos $m = \mathbf{E} X$ y $\Gamma = (\mathbf{cov}(X_i, X_j))$ tenemos que se verifica (6.3). \square

Vista la identificación de las distribuciones con sus funciones características, de la definición surge que los parámetros m y Γ caracterizan la distribución gaussiana. Mas aún, de la proposición anterior obtenemos que

$$m = \mathbf{E} X, \quad \Gamma = \mathbf{cov} X = \mathbf{E}(X - m)(X - m)^t,$$

es decir, m es el vector de las esperanzas de las coordenadas, y Γ la matriz que reúne a las covarianzas de las coordenadas.

Proposición 10. Dado el vector gaussiano X en \mathbf{R}^n con parámetros (m, Γ) , si $\det \Gamma \neq 0$ existe una matriz A de tamaño $n \times n$ tal que

$$X = m + AZ, \quad (6.4)$$

²Identificamos los vectores x de \mathbf{R}^n con las matrices columna, correspondientemente x^t denota la matriz traspuesta de x .

donde la igualdad es en distribución, y Z es un vector normal estándar en \mathbf{R}^n .

Demostración. Como Γ es definida positiva es diagonalizable en una base ortonormal. Es decir, existe una matriz de pasaje ortogonal P y una matriz diagonal D con diagonal estrictamente positiva tal que $\Gamma = P^t D P$. Tomando \sqrt{D} la matriz diagonal con entradas diagonales correspondientes a las raíces cuadradas de las entradas de D . Podemos escribir entonces, con $A = \sqrt{D} R$, que $\Gamma = A^t A$. Si Z es un vector con coordenadas independientes normales es un vector normal estándar, y se verifica (6.4). \square

6.2. Definición del Proceso de Wiener

En esta sección estudiaremos el proceso de Wiener, o Movimiento Browniano. Se trata de un modelo probabilístico para la evolución temporal de un sistema sujeto a cambios instantáneos. Su relevancia radica en dos hechos: por una parte este modelo juega un rol central en toda la teoría de los procesos estocásticos (cálculo de Itô, difusiones, martingalas, procesos de incrementos independientes, procesos autosimilares) y por otro, encuentra aplicaciones en las mas diversas ramas del conocimiento (física, biología, economía, etc.).

La denominación de este proceso se debe a las investigaciones del botánico inglés Robert Brown, que en 1828 observó y describió el movimiento caótico de una partícula de polen suspendida en agua, destacando la naturaleza física (y no biológica) del movimiento observado. Norbert Wiener, en 1923, construyó el primer modelo matemático para la descripción de dicha dinámica, similar a la que presentamos aquí. Comenzamos con la definición.

Definición 34. Diremos que un proceso estocástico $W = (W(t))_{t \geq 0}$ es un proceso de Wiener o un movimiento Browniano si cumple las siguientes propiedades

- (a) $\mathbf{P}(W(0) = 0) = 1$.
- (b) W tiene trayectorias continuas.
- (c) W tiene incrementos independientes y estacionarios.
- (d) Para cada $t > 0$ la variable aleatoria $W(t)$ tiene distribución normal con parámetros $(0, t)$.

Comencemos observando que las propiedades (c) y (d) son equivalentes a la propiedad

(e) Para todos $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ el vector aleatorio

$$V = (W(t_1), W(t_2), \dots, W(t_n))$$

tiene distribución multidimensional normal centrada, con matriz de covarianzas $\Gamma_V = ((t_i \wedge t_j))_{(i,j=1,\dots,n)}$.

que a su vez, es equivalente a la condición

(f) Para todos $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ el vector aleatorio

$$U = (W(t_1), W(t_2) - W(t_1), \dots, W(t_n) - W(t_{n-1}))$$

es gaussiano centrado, con coordenadas independientes y varianzas $\text{var}(W(t_k) - W(t_{k-1})) = t_k - t_{k-1}$ ($k = 1, \dots, n$), es decir, con matriz de covarianza $\Gamma_U = \text{diag}(t_1, t_2 - t_1, \dots, t_n - t_{n-1})$.

La equivalencia de (e) y (f) resulta de la relación $U = AV$, donde

$$A = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

Veamos por ejemplo (f) \Rightarrow (e). Para $\alpha \in \mathbf{R}^n$, tenemos $m_U = \mathbf{E}U = \mathbf{E}AV = A\mathbf{E}V = 0$. Verificamos multiplicando que $\Gamma_U = A\Gamma_V A^t$, entonces

$$\Gamma_U = \Gamma_{AV} = \mathbf{E}(AV(AV)^t) = A\Gamma_V A^t.$$

Para (e) \Rightarrow (f) utilizamos que la matriz A es invertible.

Veamos ahora que (f) \Rightarrow (c)+(d). Efectivamente (f) implica que los incrementos $W(t_{i+1}) - W(t_i)$ son centrados e independientes, con varianza igual a $t_{i+1} - t_i$, es decir, homogéneos, concluyendo (c). Como además son gaussianos, obtenemos (d). Si (c) y (d) son ciertas, un vector con coordenadas independientes y gaussianas es un vector gaussiano con coordenadas independientes (basta verificar la definición 33, que resulta centrado y con la matriz de covarianza correspondiente).

6.3. Construcción del proceso de Wiener

Una forma de demostrar la existencia del proceso de Wiener es mediante la construcción de un proceso aleatorio que verifique la definición 34.

Para esto es necesario considerar una base ortonormal de funciones de cuadrado integrable (con respecto a la medida de Lebesgue) en el intervalo $[0, 1]$.

Consideramos entonces las funciones de Haar, definidas de la siguiente forma. Notemos $a(0) = 0$ y $a(n) = 2^{n-1} - 1$ si $n \geq 1$. Para $n = 0$ ponemos $H_{00}(t) = 1$ ($0 \leq t \leq 1$). Para $n \geq 1$ y $0 \leq k \leq a(n)$, ponemos

$$H_{nk} = \begin{cases} 2^{\frac{n-1}{2}} & \text{si } \frac{k}{2^{n-1}} < t \leq \frac{k+1/2}{2^{n-1}} \\ -2^{\frac{n-1}{2}} & \text{si } \frac{k+1/2}{2^{n-1}} < t \leq \frac{k+1}{2^{n-1}} \\ 0 & \text{en otro caso.} \end{cases}$$

La *fórmula de Parseval* establece, que dadas dos funciones de cuadrado integrable $f(v)$ y $g(w)$, vale la igualdad

$$\sum_{k,m} \int_0^1 f(v) H_{mk}(v) dv \int_0^1 g(w) H_{mk}(w) dw = \int_0^1 f(v) g(v) dv. \quad (6.5)$$

Teorema 39. *Existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ en el cual existe un proceso aleatorio $W = \{W_t\}_{t \geq 0}$ que verifica las propiedades definición 34.*

Demostración. Comenzamos observando que existe un espacio de probabilidad que contiene una sucesión de variables aleatorias gaussianas estándar independientes (como por ejemplo el espacio $\mathbf{R}^{\mathbf{N}}$ con la medida producto de sus coordenadas, si estas son gaussianas estándar en \mathbf{R}). La demostración consta de dos etapas. Comenzamos construyendo un proceso $Y = \{Y_t\}_{0 \leq t \leq 1}$ que verifica la definición 6.6 para los tiempos en el intervalo $[0, 1]$.

El proceso Y se construirá como el límite uniforme en el intervalo $[0, 1]$ en un conjunto de probabilidad uno de la siguiente sucesión de funciones aleatorias. Sea

$$Y_t^n = \sum_{m=0}^n f_m(t) \quad (n \geq 0) \quad (6.6)$$

donde

$$f_m(t) = \sum_{k=0}^{a(m)} X_{mk} \int_0^t H_{mk}(u) du \quad (m \geq 0).$$

Para estudiar la convergencia de la sucesión $\{Y^n\}$, observemos que para $m \geq 1$ las funciones $\int_0^t H_{mk}(u) du$ ($0 \leq t \leq 1$) presentan su valor máximo

en el punto medio de su soporte, tomando este máximo el valor $2^{-\frac{m+1}{2}}$ independientemente del valor de k . Por ser además estos soportes disjuntos para distintos valores de k , dado un cierto m fijo, para $\varepsilon > 0$ tenemos

$$\begin{aligned} \mathbf{P}\left(\max_{0 \leq t \leq 1} |f_m(t)| \geq \varepsilon\right) &= \mathbf{P}\left(\max_{0 \leq k \leq a(m)} |X_{mk}| \geq \varepsilon 2^{\frac{m+1}{2}}\right) \\ &\leq \sum_{k=0}^{a(m)} \mathbf{P}(|X_{mk}| \geq \varepsilon 2^{\frac{m+1}{2}}) = 2^m [1 - \Phi(\varepsilon 2^{\frac{m+1}{2}})] \\ &\leq 2^m \frac{1}{\varepsilon 2^{\frac{m+1}{2}} \sqrt{2\pi}} e^{-\varepsilon^2 2^m} = b(\varepsilon), \end{aligned}$$

teniendo en cuenta que $1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ (ver ejercicio 14 del Cap. 7 de [3]). Elijamos ahora $\varepsilon = \varepsilon(m)$ para cada $m \geq 1$ de forma que $\sum_{m=1}^{\infty} b(\varepsilon(m)) < \infty$. Según el lema de Borel-Cantelli, para cada ω en un conjunto de probabilidad uno, a partir de un cierto $n(\omega)$ se cumple $\max_{0 \leq t \leq 1} |f_m(t)| \leq \varepsilon(m)$. Si elegimos $\varepsilon(m)$ de forma que, además de nuestro supuesto, se cumpla $\sum_{m=1}^{\infty} \varepsilon(m) < \infty$, del criterio de la mayorante numérica de Weierstrass se obtiene que la serie $\sum_{m=0}^{\infty} f_m(t)$ es uniformemente convergente. Es inmediato verificar que tomando $\varepsilon(m) = m2^{-m/2}$ ambas series convergen.

Una vez construido el proceso $Y_t = \sum_{m=0}^{\infty} f_m(t)$ ($0 \leq t \leq 1$), veremos que cumple las propiedades de la definición 34. La propiedad (a) es inmediata, y la (b) surge de la convergencia uniforme y la continuidad de las funciones $f_m(t)$ ($m \geq 1$). En vez de verificar (c) y (d) veamos (f).

Seguendo la definición 33, la condición (f) es equivalente a

$$\mathbf{E} e^{i \sum_{j=1}^n \alpha_j (W(t_j) - W(t_{j-1}))} = \exp\left(-\frac{1}{2} \sum_{j=1}^n \alpha_j^2 (t_j - t_{j-1})\right). \quad (6.7)$$

Para ver esto, definimos $f(t) = \sum_{j=1}^n \alpha_j \mathbf{1}_{(t_{j-1}, t_j]}(t)$ y vemos que

$$\lim_{M \rightarrow \infty} \sum_{m=0}^M \sum_{k=0}^{a(m)} \left(\sum_{j=1}^n \alpha_j \int_{t_{j-1}}^{t_j} H_{mk}(s) ds \right)^2 = \int_0^1 f(t)^2 dt = \sum_{j=1}^n \alpha_j^2 (t_j - t_{j-1}).$$

hecho que resulta de aplicar (6.5) con $f = g$. Tenemos entonces

$$\begin{aligned}
& \mathbf{E} \exp \left(i \sum_{j=1}^n \alpha_j (W(t_j) - W(t_{j-1})) \right) \\
&= \lim_{M \rightarrow \infty} \mathbf{E} \exp \left(i \sum_{j=1}^n \alpha_j \sum_{m=0}^M \sum_{k=0}^{a(m)} X_{mk} \int_{t_{j-1}}^{t_j} H_{mk}(s) ds \right) \\
&= \lim_{M \rightarrow \infty} \mathbf{E} \exp \left(i \sum_{m=0}^M \sum_{k=0}^{a(m)} X_{mk} \sum_{j=1}^n \alpha_j \int_{t_{j-1}}^{t_j} H_{mk}(s) ds \right) \\
&= \lim_{M \rightarrow \infty} \exp \left(-\frac{1}{2} \sum_{m=0}^M \sum_{k=0}^{a(m)} \left(\sum_{j=1}^n \alpha_j \int_{t_{j-1}}^{t_j} H_{mk}(s) ds \right)^2 \right) \\
&= \exp \left(-\frac{1}{2} \sum_{j=1}^n \alpha_j^2 (t_j - t_{j-1}) \right)
\end{aligned}$$

concluyendo (6.7).

La construcción del proceso de la Wiener en la semirrecta $t \geq 0$ se hace de la siguiente forma. Sea W^1, W^2, \dots una sucesión de procesos en el intervalo $[0, 1]$ independientes entre sí, cada uno de los cuales se construye con el procedimiento recién indicado. Consideremos el proceso $W = \{W_t\}_{t \geq 0}$, donde se define

$$W_t = W_1^1 + W_1^2 + \dots + W_1^{[t]} + W_{t-[t]}^{[t]+1},$$

(donde $[t]$ denota la parte entera de t). El proceso W cumple $W_0 = 0$ c.s., tiene trayectorias continuas (como se verifica por separado según t sea natural o no), tiene incrementos independientes y estacionarios (como resulta de la independencia de los incrementos de cada uno de los procesos $\{W_t^n\}_{0 \leq t \leq 1}$ y de la independencia mutua de estos procesos para distintos valores de $n \geq 1$, teniendo la variable aleatoria W_t distribución normal centrada y varianza t (aquí se utiliza que W_t es una suma de variables aleatorias independientes con distribución normal). En conclusión, el proceso W cumple la definición (34). \square

Concluimos la sección con algunas primeras propiedades simples. Existen una serie de transformaciones que conservan la ley del proceso.

Proposición 11. Sea $\{W_t\}$ un proceso de Wiener. También son procesos de Wiener: $W^1 = \{-W_t\}$, $W^2 = \{\sqrt{c}W_{t/c}\}$, $W^3 = \{W_{a+t} - W_a\}$, donde $a > 0$.

Demostración. La demostración en todos los casos se reduce a la verificación de la definición 34. \square

6.4. Propiedades de las trayectorias del Proceso de Wiener

Una de las características más interesantes del proceso de Wiener es la naturaleza de sus trayectorias. Consideramos un intervalo $[0, T]$, y una sucesión de particiones

$$\lambda^n = \{0 = t_0^n < t_1^n < \dots < t_{k(n)}^n = T\} \quad (n = 1, 2, \dots), \quad (6.8)$$

cuya norma $|\lambda^n| = \max\{t_k^n - t_{k-1}^n : k = 1, \dots, k(n)\}$ tiende a cero si $n \rightarrow \infty$, y tales que se verifica $\lambda^n \subset \lambda^{n+1}$ ($n = 1, 2, \dots$) es decir, cada partición se obtiene de la anterior agregando puntos. Para una función $f: [0, T] \rightarrow \mathbf{R}$ con derivada $f'(t)$ continua en $[0, T]$, cuando $n \rightarrow \infty$, tenemos

$$\sum_{k=1}^{k(n)} |f(t_k^n) - f(t_{k-1}^n)| = \sum_{k=1}^{k(n)} |f'(\theta_k^n)|(t_k^n - t_{k-1}^n) \rightarrow \int_0^T |f'(t)| dt.$$

(Aquí aplicamos el teorema del valor medio, $\theta_k^n \in [t_{k-1}^n, t_k^n]$ para cada n y cada k .) El límite obtenido es la *variación* de la función $f(t)$ en el intervalo $[0, T]$. En forma similar, si $|f'(t)| \leq M$ ($0 \leq t \leq T$), tenemos

$$\sum_{k=1}^{k(n)} (f(t_k^n) - f(t_{k-1}^n))^2 \leq \sum_{k=1}^{k(n)} M^2 (t_k^n - t_{k-1}^n)^2 \leq M^2 T |\lambda^n| \rightarrow 0,$$

si $n \rightarrow \infty$, y decimos, que la *variación cuadrática* de la función $f(t)$ en el intervalo $[0, T]$ es nula.

El siguiente teorema muestra que las trayectorias de un proceso de Wiener presentan un comportamiento diferente: su variación en un intervalo $[0, T]$ no existe, es infinita; y su variación cuadrática en un intervalo $[0, T]$ es igual a T .

Teorema 40 (Propiedades de las trayectorias).

Consideremos un proceso de Wiener $\{W_t\}$ y una sucesión creciente de particiones $\{\lambda^n\}$ como en (6.8), cuyas normas $\{|\lambda^n|\}$ tienden a cero si $n \rightarrow \infty$.

Se verifica

$$V_n = \sum_{k=1}^{k(n)} |W_{t_k^n} - W_{t_{k-1}^n}| \rightarrow \infty \quad (n \rightarrow \infty) \quad \text{c.s.} \quad (6.9)$$

$$Q_n = \sum_{k=1}^{k(n)} (W_{t_k^n} - W_{t_{k-1}^n})^2 \rightarrow T \quad (n \rightarrow \infty) \quad \text{en media cuadrática.} \quad (6.10)$$

Además, si $\sum_{n=1}^{\infty} |\lambda^n| < \infty$, la convergencia en (6.10) es casi segura.

Demostración. Comencemos con la demostración de (6.9). En primer lugar observemos que como las particiones son crecientes, aplicando la propiedad triangular, se obtiene que $V_n \leq V_{n+1}$ ($n = 1, 2, \dots$), es decir, la sucesión $\{V_n\}$ es no decreciente, casi seguramente. Queremos demostrar que $\mathbf{P}(V_n \rightarrow \infty) = 1$. Esto es equivalente a demostrar que dado $K > 0$ arbitrario, se verifica $\mathbf{P}(\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{V_m(\omega) > K\}) = 1$. Esta igualdad, tomando complementos, es equivalente a $\mathbf{P}(\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} \{V_m(\omega) \leq K\}) = 0$. Como la sucesión $\{V_n\}$ es no decreciente, tenemos

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{V_m(\omega) \leq K\}\right) \leq \mathbf{P}\left(\bigcup_{m=n}^{\infty} \{V_m(\omega) \leq K\}\right) = \mathbf{P}(V_n(\omega) \leq K).$$

En conclusión, para demostrar (6.9), verificamos que $\mathbf{P}(V_n(\omega) \leq K) \rightarrow 0$ ($n \rightarrow \infty$). Como $\{W_t\}$ tiene incrementos independientes, aplicando la fórmula de la varianza de una suma de variables independientes, tenemos

$$\text{var } V_n = \sum_{k=1}^{k(n)} \text{var } |W_{t_k^n} - W_{t_{k-1}^n}| \leq \sum_{k=1}^{k(n)} \mathbf{E} |W_{t_k^n} - W_{t_{k-1}^n}|^2 = T.$$

Por otra parte, tenemos $\mathbf{E} |W_{t_k^n} - W_{t_{k-1}^n}| = \sqrt{t_k^n - t_{k-1}^n} \mathbf{E} |Z|$, donde Z es una variable aleatoria con distribución normal estándar, y $\mathbf{E} |Z| = \sqrt{2/\pi}$. Entonces, como $\sqrt{t_k^n - t_{k-1}^n} \geq (t_k^n - t_{k-1}^n) / \sqrt{|\lambda^n|}$, tenemos

$$\mathbf{E} V_n = \sum_{k=1}^{k(n)} \mathbf{E} |W_{t_k^n} - W_{t_{k-1}^n}| = \mathbf{E} |Z| \sum_{k=1}^{k(n)} \sqrt{t_k^n - t_{k-1}^n} \geq \frac{T \mathbf{E} |Z|}{\sqrt{|\lambda^n|}} \rightarrow \infty,$$

si $n \rightarrow \infty$. Para n suficientemente grande se verifica $\mathbf{E} V_n > K$, y aplicando la desigualdad de Chebishev (??), obtenemos

$$\begin{aligned} \mathbf{P}(V_n \leq K) &\leq \mathbf{P}(|V_n - \mathbf{E} V_n| \geq \mathbf{E} V_n - K) \\ &\leq \frac{1}{(\mathbf{E} V_n - K)^2} \text{var } V_n \rightarrow 0 \quad (n \rightarrow \infty), \end{aligned}$$

concluyendo la demostración de (6.9).

Veamos ahora la demostración de (6.10). Las variables aleatorias $Y_n = (W_{t_k^n} - W_{t_{k-1}^n})^2 - (t_k^n - t_{k-1}^n)$ ($k = 1, \dots, k(n)$) son independientes, y verifican

$$\mathbf{E}Y_n = 0, \quad \text{var } Y_n = (t_k^n - t_{k-1}^n)^2 \mathbf{E}(Z^2 - 1)^2,$$

si Z designa una variable aleatoria con distribución normal estándar. Aplicando la fórmula de la varianza de una suma de variables independientes centradas, donde los momentos de orden dos coinciden con las varianzas, tenemos

$$\begin{aligned} \mathbf{E}(Q_n - T)^2 &= \mathbf{E} \left(\sum_{k=1}^{k(n)} Y_n \right)^2 = \sum_{k=1}^{k(n)} \mathbf{E}(Y_n)^2 = \sum_{k=1}^{k(n)} (t_k^n - t_{k-1}^n)^2 \mathbf{E}(Z^2 - 1)^2 \\ &\leq T |\lambda^n| \mathbf{E}(Z^2 - 1)^2 \rightarrow 0 \quad (n \rightarrow \infty), \end{aligned} \quad (6.11)$$

obteniendo la convergencia en media cuadrática.

La convergencia casi segura, bajo el supuesto $\sum_{n=1}^{\infty} |\lambda^n| < \infty$, se obtiene de la siguiente forma. Sabemos que $\mathbf{P}(Q_n \rightarrow T) = 1$ si (y solo si) para todo $\varepsilon > 0$, se cumple $\mathbf{P} \left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|Q_n - T| > \varepsilon\} \right) = 0$ (ver fórmula (??)). Tenemos

$$\begin{aligned} \mathbf{P} \left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|Q_n - T| > \varepsilon\} \right) &\leq \mathbf{P} \left(\bigcup_{n=m}^{\infty} \{|Q_n - T| > \varepsilon\} \right) \\ &\leq \sum_{n=m}^{\infty} \mathbf{P}(|Q_n - T| > \varepsilon) \\ &\leq \frac{1}{\varepsilon^2} \sum_{n=m}^{\infty} \mathbf{E}|Q_n - T|^2 \leq t \mathbf{E}(Z^2 - 1) \sum_{n=m}^{\infty} |\lambda^n| \rightarrow 0, \end{aligned}$$

si $m \rightarrow \infty$, donde utilizamos la acotación obtenida en (6.11). De aquí se obtiene la convergencia casi segura³, concluyendo la demostración. \square

El siguiente resultado está tomado de [1].

Teorema 41 (No diferenciabilidad de las trayectorias del Proceso de Wiener). *Sea $\{W(t) : t \geq 0\}$ un proceso de Wiener. Tenemos*

$$\mathbf{P}(\{\omega : W(\cdot, \omega) \text{ no es diferenciable c.t.p}\}) = 1.$$

³Hemos demostrado que $\sum_{n=1}^{\infty} \mathbf{P}(\mathbf{A}_n) < \infty$ implica $\mathbf{P}(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \mathbf{A}_n) = 0$, que es la primer parte del llamado *lema de Borel-Cantelli*.

Demostración. Sea

$$X_{nk} = \max \left\{ W \left(\frac{j+1}{2^n} \right) - W \left(\frac{j}{2^n} \right) : j = k, k+1, k+2 \right\}$$

Tenemos

$$\begin{aligned} \mathbf{P}(|X_{nk}| \leq \varepsilon) &= \mathbf{P}(|W(1)|2^{-n/2} \leq \varepsilon) = \left(\int_{-\varepsilon 2^{n/2}}^{\varepsilon 2^{n/2}} \phi(x) dx \right)^3 \\ &\leq \left(2^{n/2} \varepsilon 2\phi(0) \right)^3 = \left(\sqrt{\frac{2}{\pi}} \right)^3 \varepsilon^3 2^{3n/2}. \end{aligned}$$

Consideremos ahora

$$Y_n = \min\{X_{kn} : 0 \leq k \leq n2^n\}.$$

Tenemos

$$\begin{aligned} \mathbf{P}(Y_n \leq \varepsilon) &= \mathbf{P} \left(\bigcup_{k=1}^{n2^n} \{X_{nk} \leq \varepsilon\} \right) \leq n2^n \mathbf{P}(X_{nk} \leq \varepsilon) \leq n2^n \left(\sqrt{\frac{2}{\pi}} \right)^3 \varepsilon^3 2^{3n/2} \\ &= n \left(\sqrt{\frac{2}{\pi}} \right)^3 \varepsilon^3 2^{5n/2}. \end{aligned}$$

Las derivadas laterales son

$$D^W(t, \omega) = \limsup_{h \rightarrow 0^+} \frac{W(t+h) - W(t)}{h}, \quad D_W(t, \omega) = \liminf_{h \rightarrow 0^-} \frac{W(t+h) - W(t)}{h}.$$

Consieremos ahora el suceso:

$$E = \{\omega : \exists t \geq 0 \text{ con } D^W(t, \omega) \text{ y } D_W(t, \omega) \text{ finitos}\}$$

Sea $\omega \in E$ y $K > 0$ tal que

$$-K < D_W(t, \omega) \leq D^W(t, \omega) < K.$$

Existe un $\delta = \delta(\omega, t, K)$ tal que

$$t \leq t+s \leq t+\delta \implies |W(s, \omega) - W(t, \omega)| \leq K|t-s|.$$

Sea ahora n tal que

$$4 \times 2^{-n} < \delta, \quad 8K < n, \quad n > t.$$

Dado ese n elijo el k que verifica

$$\frac{k-1}{2^n} \leq t$$

Luego se verifica

$$\left| \frac{j}{2^n} - t \right| \leq \delta \text{ para } j = k, k+1, k+2.$$

Entonces

$$X_{nk} \leq 2K \frac{4}{2^n} \leq \frac{n}{2^n}$$

lo que implica que $Y_n \leq \frac{n}{2^n}$.

$$\mathbf{P} \left(Y_n \leq \frac{n}{2^n} \right) \leq C \left(\frac{n}{2^n} \right)^3 n 2^{5n/2}.$$

da una serie convergente. Por Borel-Cantelli existe un n_0 tal que $Y_n \geq n/2^n$ lo que da

$$\mathbf{P}(E) = 0.$$

□

6.5. La distribución del máximo del proceso de Wiener

Teorema 42. *La distribución del máximo del proceso de Wiener en un intervalo acotado se calcula mediante*

$$\mathbf{P} \left(\max_{0 \leq t \leq T} W_t \geq u \right) = 2 \mathbf{P}(W_T \geq u). \quad (6.12)$$

La propiedad (6.12) se conoce como el *principio de reflexión* y se demuestra habitualmente utilizando la propiedad fuerte de Markov del proceso de Wiener (ver por ejemplo [2]). La demostración que presentamos se basa en propiedades de martingala y el teorema del muestreo opcional de Doob.

Demostración. Sea n fijo. Consideremos para $k = 0, \dots, 2^n - 1$ las variables aleatorias

$$X_k = \Phi \left(\frac{W_{kT/2^n} - u}{\sqrt{T(1 - k/2^n)}} \right),$$

donde Φ es la distribución normal estándar. Considerando la filtración $\mathcal{F}_k = \sigma(W_{jT/2^n} : 0 \leq j \leq k)$ tenemos $\mathbf{E}(X_{k+1} | \mathcal{F}_k) = X_k$. En efecto, designando $Z = W_{kT/2^n}$ y $\Delta = -W_{(k+1)T/2^n} + W_{kT/2^n} \sim \mathcal{N}(0, \delta)$, con $\delta = T/2^n$, tenemos

$$\begin{aligned}
\mathbf{E}(X_{k+1} | \mathcal{F}_k) &= \mathbf{E} \left(\Phi \left(\frac{Z - u - \Delta}{\sqrt{T(1 - (k+1)/2^n)}} \right) \mid \mathcal{F}_k \right) \\
&= \int_{-\infty}^{\infty} \Phi \left(\frac{Z + x - u}{\sqrt{T(1 - (k+1)/2^n)}} \right) \frac{1}{2\pi\delta} \varphi \left(\frac{x}{\sqrt{\delta}} \right) dx \\
&= \int_{-\infty}^{\infty} \mathbf{P}(\mathcal{N}(0, T(1 - (k+1)/2^n)) + x \leq Z - u \mid Z) \frac{1}{2\pi\delta} \varphi \left(\frac{x}{\sqrt{\delta}} \right) dx \\
&= \mathbf{P}(\mathcal{N}(0, T(1 - (k+1)/2^n)) + \mathcal{N}(0, \delta) \leq Z - u \mid Z) \\
&= \mathbf{P}(\mathcal{N}(0, T(1 - (k+1)/2^n) + \delta) \leq Z - u \mid Z) \\
&= \mathbf{P}(\mathcal{N}(0, T(1 - k/2^n)) \leq Z - u \mid Z) = \Phi \left(\frac{Z - u}{\sqrt{T(1 - k/2^n)}} \right) = X_k.
\end{aligned}$$

Si consideramos el tiempo de parada

$$\tau^n = \inf\{k = 0, \dots, 2^n - 1 : W_{kT/2^n} \geq u\} \wedge (2^n - 1),$$

y aplicamos el Teorema del Muestreo Opcional de Doob tenemos en los tiempos $\sigma = 0$ y τ^n , tenemos

$$\begin{aligned}
\Phi \left(\frac{-u}{\sqrt{T}} \right) &= \mathbf{E} \Phi \left(\frac{W_{\tau^n} - u}{\sqrt{T(1 - \tau^n/2^n)}} \right) \\
&= \mathbf{E} \Phi \left(\frac{W_{\tau^n T/2^n} - u}{\sqrt{T(1 - \tau^n/2^n)}} \right) \mathbf{1}_{\{\tau^n < (2^n - 1)\}} \\
&\quad + \mathbf{E} \Phi \left(\frac{W_{T(1 - 2^{-n})} - u}{\sqrt{T/2^n}} \right) \mathbf{1}_{\{\tau^n = (2^n - 1)\}} \tag{6.13}
\end{aligned}$$

En primer lugar observamos que

$$\Phi \left(\frac{-u}{\sqrt{T}} \right) = \mathbf{P}(W_T > u).$$

Tomamos ahora límite cuando n tiene a infinito. En lo que respecta al segundo sumando en (6.13), tenemos

$$\Phi \left(\frac{W_{T(1 - 2^{-n})} - u}{\sqrt{T/2^n}} \right) \mathbf{1}_{\{\tau^n = (2^n - 1)\}} \rightarrow \Phi(-\infty) \mathbf{1}_{\{\max_{0 \leq t \leq T} W_t < u\}} = 0.$$

Por otra parte, como las trayectorias son continuas, tenemos $W_{\tau^n T/2^n} \rightarrow u$, y $\tau^n \rightarrow \inf\{t \in [0, T]: W_t = u\}$. Por eso

$$\begin{aligned} \Phi\left(\frac{W_{\tau^n T/2^n} - u}{\sqrt{T(1 - \tau^n/2^n)}}\right) \mathbf{1}_{\{\tau^n < (2^n - 1)\}} &\rightarrow \Phi(0) \mathbf{1}_{\{\max_{0 \leq t \leq T} W_t \geq u\}} \\ &= \frac{1}{2} \mathbf{1}_{\{\max_{0 \leq t \leq T} W_t \geq u\}}. \end{aligned}$$

Como $0 \leq \phi(x) \leq 1$, aplicando convergencia dominada en (6.13) obtenemos la tesis (6.12). \square

Estamos en condiciones de demostrar el siguiente resultado.

Proposición 12. *El proceso de Wiener verifica la ley fuerte de los grandes números, es decir*

$$\frac{W_t}{t} \rightarrow 0 \text{ c.s. } (t \rightarrow \infty)$$

Demostración. La demostración se basa en la ley fuerte para sucesiones de variables aleatorias i.i.d. Tenemos

$$\frac{W_t}{t} = \frac{W_t - W_{[t]}}{t} + \frac{[t]}{t} \times \frac{W_{[t]}}{[t]}. \quad (6.14)$$

Tenemos

$$\frac{W_{[t]}}{[t]} = \frac{1}{[t]} \sum_{k=1}^{[t]} (W_k - W_{k-1}) \rightarrow 0 \text{ (} t \rightarrow \infty \text{),}$$

dado que las variables $\{W_k - W_{k-1}\}$ son independientes e idénticamente distribuidas, centradas, por lo que se aplica la Ley fuerte de Kolmogorov. Como $[t]/t \rightarrow 1$ si $t \rightarrow \infty$, tenemos que el segundo sumando en (6.14) tiende a cero casi seguramente.

Veamos ahora que

$$\frac{W_t - W_{[t]}}{[t]} \rightarrow 0 \text{ c.s.} \quad (6.15)$$

En efecto, dada la homogeneidad de la distribución y la distribución del máximo calculada en el teorema 42, tenemos (poniendo n en lugar de $[t]$), que

$$\sum_{n=1}^{\infty} \mathbf{P}\left(\max_{n \leq t \leq n+1} \frac{|W_t - W_n|}{n} \geq \varepsilon\right) < \infty,$$

porque

$$\begin{aligned} \mathbf{P} \left(\max_{n \leq t \leq n+1} \frac{|W_t - W_n|}{n} \geq \varepsilon \right) &= \mathbf{P} \left(\max_{0 \leq t \leq 1} |W_t| \geq \varepsilon n \right) \\ &\leq 2 \mathbf{P} \left(\max_{0 \leq t \leq 1} W_t \geq \varepsilon n \right) = 4 \mathbf{P} (W_1 \geq \varepsilon n) \\ &\leq \frac{4}{\sqrt{2\pi\varepsilon n}} \exp(-(\varepsilon n)^2/2), \end{aligned}$$

es el término general de una serie convergente, lo que demuestra (6.15) aplicando Borel-Cantelli, concluyendo la demostración. \square

Capítulo 7

Parada óptima para el proceso de Wiener

- Tomar una decisión en un escenario de incertidumbre
- Modelamos la incertidumbre mediante un proceso estocástico $\{X_t\}$ en $(\Omega, \mathcal{F}, \mathbf{P})$
- Medimos la calidad de la decisión a través de una función real $g(x)$
- Pagamos por observar el proceso de acuerdo a una tasa $r \geq 0$.
- Las estrategias son tiempos aleatorios $\tau(\omega)$
- Aplicaciones: estadística secuencial, opciones americanas.

Un problema de parada óptima es el modelo matemático que corresponde a la toma de una decisión en un contexto de incertidumbre, modelada mediante un proceso estocástico $X_t(\omega)$

El objetivo del decisor es maximizar una cantidad o ganancia $g(X_t)$ donde g es una función positiva, y a veces puede existir un costo de observación, en cuyo caso la ganancia es $e^{-rt}g(X_t)$, donde $r > 0$ es la tasa de descuento.

La decisión es una estrategia $\tau = \tau(\omega)$ que depende de la observación en tiempo real del proceso, es decir, podemos decidir parar en $\tau = t$ con la observación del proceso hasta t . En términos matemáticos τ es un tiempo de parada. Las dos aplicaciones más importantes son en estadística, donde el proceso es la verosimilitud y la función está relacionada con la suma de los errores de tipo I y tipo II (estadística secuencial) y finanzas, donde se observa un precio X_t para vender, y se pretende maximizar la ganancia (o en forma parecida se constituye una opción).

7.1. El problema de parada óptima: Ejemplo Guía

Tenemos tres elementos:

- Un proceso estocástico $X = \{X_t\}$, que parte de $X_0 = x$.
- Una función de pago $g = g(x) \geq 0$
- Una tasa de descuento $r \geq 0$.

Nuestro objetivo

Elegir t para maximizar $g(X_t)$ con un costo de observación

Matemáticamente:

Elegir τ para maximizar $\mathbf{E}_x e^{-r\tau} g(X_\tau)$.

- El proceso es el movimiento Browniano: $X = \{x + W_t\}$.
- La función de pago es: $g(x) = x^+ = \max(x, 0)$,
- La tasa de descuento es $r > 0$.

El problema de optimización es entonces¹:

Hallar $V(x)$, τ^* tales que

$$V(x) = \sup_{\tau} \mathbf{E} e^{-r\tau} (x + W_{\tau})^+ = \mathbf{E} e^{-r\tau^*} (x + W_{\tau^*})^+$$

Busquemos estrategias de nivel, es decir, tiempos de parada de la forma

$$\tau(a) = \inf\{t \geq 0: x + W_t \geq a\}$$

Calculemos para cada $a > x$ cual es el valor de detenernos en el nivel a . Es decir, calculemos

$$V(x, a) = \mathbf{E} e^{-r\tau(a)} (x + W_{\tau(a)})^+$$

Lo sencillo de estas estrategias es que

$$(x + W_{\tau(a)})^+ = a \mathbf{1}_{\{\tau(a) < \infty\}}.$$

¹Taylor (1968), Annals of Math. Statistics

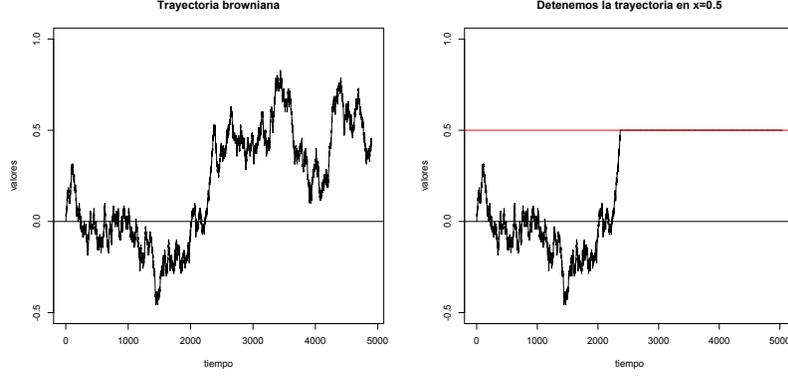


Figura 7.1: A la izquierda una trayectoria Browniana, a la derecha nos detenemos en el nivel $a = 0,5$

(Si $\tau(a) = \infty$ entonces el valor a recibir es cero.) Resulta entonces

$$V(x, a) = a \mathbf{E} e^{-r\tau(a)} \mathbf{1}_{\{\tau(a) < \infty\}}.$$

Calculemos ahora esta última esperanza Para eso observamos que

$$M_t = e^{\sqrt{2r}(x+W_t)-rt}$$

es una martingala. Aplicamos entonces el teorema del muestreo opcional de Doob con el tiempo de parada $\tau(a) \wedge T$ (que es acotado):

$$\mathbf{E} M_0 = \mathbf{E} M_{\tau(a) \wedge T} \quad (7.1)$$

Vamos a tomar límite si $T \rightarrow \infty$. Tenemos

$$M_{\tau(a) \wedge T} = e^{\sqrt{2r}(x+W_{\tau(a) \wedge T})-r(\tau(a) \wedge T)} \leq e^{\sqrt{2r}a}$$

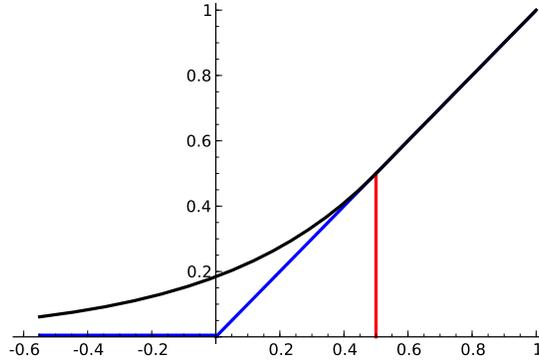
$$\lim_{T \rightarrow \infty} M_{\tau(a) \wedge T} = e^{\sqrt{2r}(x+W_{\tau(a)})-r\tau(a)} \mathbf{1}_{\{\tau(a) < \infty\}} = e^{\sqrt{2r}a} e^{-r\tau(a)} \mathbf{1}_{\{\tau(a) < \infty\}}.$$

Esto resulta porque si $\tau(a) = \infty$, como $W(T)/T \rightarrow 0$ ($T \rightarrow \infty$), tenemos

$$\lim_{T \rightarrow \infty} M_{\tau(a) \wedge T} \mathbf{1}_{\{\tau(a) = \infty\}} = 0.$$

Estamos en condiciones de aplicar el Teorema de Convergencia dominada en (7.1):

$$\begin{aligned} e^{\sqrt{2r}x} &= \mathbf{E} M_0 = \mathbf{E} M_{\tau(a) \wedge T} \rightarrow \mathbf{E} e^{\sqrt{2r}a} e^{-r\tau(a)} \mathbf{1}_{\{\tau(a) < \infty\}} \\ &= e^{\sqrt{2r}a} \mathbf{E} e^{-r\tau(a)} \mathbf{1}_{\{\tau(a) < \infty\}} \end{aligned}$$



Despejando obtenemos

$$V(x, a) = a \mathbf{E} e^{-r\tau(a)} \mathbf{1}_{\{\tau(a) < \infty\}} = a e^{\sqrt{2r}(x-a)}$$

Podemos ahora elegir el mejor nivel, es decir, el nivel a que maximiza $V(x, a)$
 Para eso maximizamos $f(a) = a e^{-a\sqrt{2r}}$

$$f'(a) = e^{-\sqrt{2r}a}(1 - a\sqrt{2r}) = 0,$$

y tenemos un nivel óptimo

$$a^* = \frac{1}{\sqrt{2r}}.$$

Para ese nivel, y para $x < a^*$, tenemos

$$V(x, 1/\sqrt{2r}) = \frac{1}{\sqrt{2r}} e^{\sqrt{2r}(x-1/\sqrt{2r})} = \frac{1}{e\sqrt{2r}} e^{\sqrt{2r}x}$$

Tenemos un candidato a solución (dibujo para $r = 2$):

- En azul la función de pago $g(x) = x^+$.
- En negro la función de valor $V(x)$.
- En rojo el nivel óptimo $a^* = 1/2$.

Que nos resta:

- Demostrar que para $x > 1/\sqrt{2r}$ lo mejor es detenerse en $\tau^* = 0$, resultando $V(x) = g(x)$

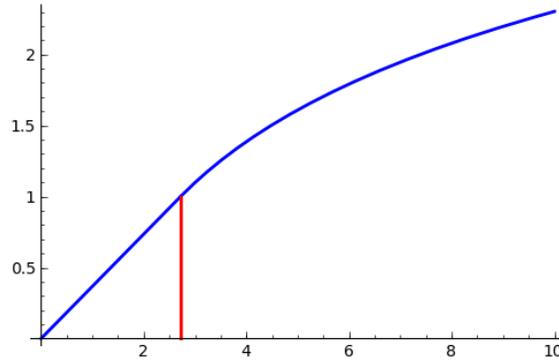


Figura 7.2: La función Φ es cóncava

- Demostrar que la estrategia es la mejor en el conjunto de todos los tiempos de parada, no solamente los del tipo de alcanzar un nivel

Reescribimos nuestra martingala.

$$M_t = e^{\sqrt{2r}(x+W_t)-rt} = e^{-rt}m(x + W_t)$$

donde introducimos

$$m(y) = e^{\sqrt{2r}y}, \quad X_t = x + W_t.$$

Buscamos ahora una función $\Phi(u)$ tal que

$$V(x) = \Phi(m(x)) \quad \text{para todo } x \text{ real}$$

Es decir, buscamos Φ que verifique

$$\Phi(e^{\sqrt{2r}y}) = \begin{cases} \frac{1}{e\sqrt{2r}}e^{\sqrt{2r}y}, & \text{cuando } y \leq \frac{1}{\sqrt{2r}} \\ y, & \text{cuando } y > \frac{1}{\sqrt{2r}} \end{cases}$$

No es difícil ver que

$$\Phi(u) = \begin{cases} \frac{1}{e\sqrt{2r}}u, & \text{cuando } u \leq e \\ \frac{1}{\sqrt{2r}} \log u, & \text{cuando } u > e \end{cases}$$

verifica la propiedad. Grafiquemos $\Phi(u)$: Si $\alpha < 1$ entonces

$$\alpha\Phi(u) = \alpha\Phi(u) + (1 - \alpha)\Phi(0) \leq \Phi(\alpha u)$$

Si $\beta = 1/\alpha > 1$, $v = \alpha u$,

$$\Phi(\beta v) \leq \beta \Phi(v).$$

Veamos que $\{e^{-rt}V(X_t)\}$ es una supermartingala. En efecto,

$$\begin{aligned} \mathbf{E}(e^{-r(t+h)}V(X_{t+h}) \mid \mathcal{F}_t) &= e^{-r(t+h)} \mathbf{E}(\Phi(m(x + W_{t+h})) \mid \mathcal{F}_t) \\ &\leq e^{-r(t+h)} \Phi(\mathbf{E}(m(X_{t+h})) \mid \mathcal{F}_t) \\ &= e^{-r(t+h)} \Phi\left(e^{r(t+h)} \mathbf{E}(e^{-r(t+h)}m(X_{t+h})) \mid \mathcal{F}_t\right) \\ &= e^{-r(t+h)} \Phi\left(e^{r(t+h)}e^{-rt}m(X_t)\right) \\ &= e^{-r(t+h)} \Phi\left(e^{rh}m(X_t)\right) \leq e^{-rt} \Phi(m(X_t)) = e^{-rt}V(X_t). \end{aligned}$$

Recapitulemos:

- Definiendo $\tau(a^*) = \inf\{t \geq 0: x + W_t \geq a^*\}$ resulta $\tau(a^*) = 0$ si $x \geq a$.
- Por eso, la función candidata verifica

$$V(x) = \mathbf{E} e^{-r\tau(a^*)}(x + W_{\tau(a^*)})^+$$

- Es suficiente probar que para cualquier tiempo de parada σ

$$V(x) \geq \mathbf{E} e^{-r\sigma}(x + W_\sigma)^+$$

para todo x real. Consideremos σ tiempo de parada y $\sigma \wedge T$, acotado. Como $e^{-rt}V(X_t)$ es super-martingala, y se verifica

$$V(x) \geq g(x),$$

tenemos

$$V(x) \geq \mathbf{E} e^{-r(\sigma \wedge T)}V(X_{\sigma \wedge T}) \geq \mathbf{E} e^{-r(\sigma \wedge T)}g(X_{\sigma \wedge T}).$$

Como $\lim_{T \rightarrow \infty} e^{-rT}g(X_T) = 0$, aplicando Fatou, obtenemos

$$\begin{aligned} V(x) &\geq \liminf_{T \rightarrow \infty} \mathbf{E} e^{-r(\sigma \wedge T)}g(X_{\sigma \wedge T}) \\ &\geq \mathbf{E} \liminf_{T \rightarrow \infty} e^{-r(\sigma \wedge T)}g(X_{\sigma \wedge T}) = \mathbf{E} e^{-r\sigma}g(X_\sigma) \end{aligned}$$

Esto concluye la demostración de que $V, \tau(a^*)$ es la solución del problema

Bibliografía

- [1] Billingsley, Patrick. Probability and measure. Second edition. Wiley, New York, 1986.
- [2] Karatzas, Ioannis; Shreve, Steven E. Brownian motion and stochastic calculus. Second edition. Graduate Texts in Mathematics, 113. Springer-Verlag, New York, 1991.
- [3] Petrov, V.V., Mordecki, E. Teoría de la Probabilidad. (2a. ed.) Facultad de Ciencias, DIRAC, 2008.