Probabilidad y Procesos Estocásticos

Onésimo Hernández-Lerma y Carlos G. Pacheco
CINVESTAV-IPN

Contenido

Parte 1: PROBABILIDAD **3**

1 Espacios de medida **3**

2 Espacios discretos y continuos **16**

3 Probabilidad condicional e independencia **25**

4 Variables aleatorias **34**

5 Vectores aleatorios **49**

6 Esperanza de vv.aa. discretas y continuas **58**

7 La integral de Lebesgue **67**

8 Covarianza e independencia **75**

9 Convergencia de vv.aa. **82**

10 Funciones características y el TLC **106**

11 Esperanza condicional **125**

Parte 2: PROCESOS ESTOCÁSTICOS **142**

12 Cadenas de Markov: conceptos básicos **142**

13 Clasificación de estados de una CM **163**

14 Distribución límite de una cadena de Markov **180**

15 Martingalas **196**

16 Procesos a tiempo continuo: introducción **214**

17 Proceso markoviano de saltos (PMS) **232**

18 La matriz generadora de un PMS **244**

19 Comportamiento asintótico de un PMS **257**

20 Procesos estacionarios **269**

21 Cálculo en L_2 **288**

22 Ecuaciones diferenciales en L_2 **301**

23 Movimiento Browniano **312**

24 La integral de Ito **322**

25 La regla diferencial de Ito **339**

26 Ecuaciones diferenciales estocásticas **354**

27 Apéndice **376**

28 Bibliografía **380**

Parte 1: PROBABILIDAD

1 Espacios de medida

Contenido: Espacios medibles, conjuntos de Borel, medidas, el teorema de extensión de Carathéodory.

Un espacio de medida es una terna $(\Omega, \mathcal{F}, \mu)$ cuyas componentes definimos en esta sección.

Definición 1.1. Sea Ω un conjunto no vacío y $\mathcal{F}$ una familia de subconjuntos de Ω. Decimos que $\mathcal{F}$ es una $\boldsymbol{\sigma}$**–álgebra** si:

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

(b) si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$,

(c) si $\{A_1, A_2, \ldots\}$ es una sucesión de conjuntos en $\mathcal{F}$, entonces $\cup A_n \in \mathcal{F}$.

Equivalentemente, $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω si, en la Definición 1.1, se cumple la condicion (b) pero las condiciones (a) y (c) se sustituyen por las respectivas condiciones de la siguiente proposicion.

Proposición 1.2. $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω ssi

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

(b) Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$,

(c) Si A_n es una sucesion en $\mathcal{F}$, entonces $\bigcap A_n \in \mathcal{F}$.

Definición 1.3. Si $\mathcal{F}$ es una σ–álgebra de Ω, se dice que el par $(\Omega, \mathcal{F})$ es un **espacio medible**. Si A es un conjunto en $\mathcal{F}$, decimos que A es $\mathcal{F}$–**medible** (o medible con respecto a $\mathcal{F}$).

Terminología de probabilidad. Si $(\Omega, \mathcal{F})$ es un espacio medible, en probabilidad se dice que Ω es el **espacio muestral** o **evento seguro**, y que $\mathcal{F}$ es una **familia de eventos**. A un conjunto $A \in \mathcal{F}$ se le llama **evento**. Al conjunto vacío $\phi \in \mathcal{F}$ se le llama **evento imposible**.

Ejemplo 1.4. (a) A la familia $\{\Omega, \phi\}$ se le llama la ***σ*–álgebra trivial**, y es la "mínima" σ–álgebra de Ω.

(b) La σ–álgebra que consiste de *todos* los subconjuntos de Ω se llama el **conjunto potencia** de Ω y se denota por 2^Ω. Esta es la "máxima" σ–álgebra de Ω. (El nombre *conjunto potencia* se debe a que si Ω consiste de n elementos, digamos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, entonces tiene 2^n subconjuntos. Vea el Ejemplo 2.8).

(c) Si $\{\mathcal{F}_i, i \in I\}$ es una familia de σ–álgebras de Ω, entonces la intersección $\cap \mathcal{F}_i$ también es una σ–álgebra. (**Nota**. En general, la unión $\cup \mathcal{F}_i$ **no** es una σ–álgebra porque la condición (c) en 1.1 no se cumple.) Recuerde que $\bigcap_{i \in I} \mathcal{F}_i = \{A \subset \Omega : (\forall i \in I) A \in \mathcal{F}_i\}$

(d) Sea $\mathcal{A}$ una familia arbitraria de subconjuntos de Ω, y sea $\sigma\{\mathcal{A}\}$ la intersección de todas las σ–álgebras de Ω que contienen a $\mathcal{A}$. Entonces, por (c), $\sigma\{\mathcal{A}\}$ es una σ–álgebra y, de hecho, es la *mínima* σ–álgebra que contiene a $\mathcal{A}$; es decir, si $\mathcal{F}$ es cualquier σ–álgebra de Ω que contiene a $\mathcal{A}$, entonces $\sigma\{\mathcal{A}\} \subset \mathcal{F}$. A $\sigma\{\mathcal{A}\}$ se le llama la ***σ*–álgebra generada por** $\mathcal{A}$. Por ejemplo, supóngase que $\mathcal{A}$ consiste de un único conjunto $B \subset \Omega$, es decir, $\mathcal{A} = \{B\}$. Entonces $\sigma\{\mathcal{A}\} = \{B, B^c, \Omega, \phi\}$. □

Un caso especial muy importante del Ejemplo 1.4(d) es el siguiente.

Definición 1.5. Sea $\Omega = \mathbb{R}$ y sea $\mathcal{A}$ la familia de todos los intervalos abiertos $(a, b) \subset \mathbb{R}$. Entonces la σ–álgebra $\sigma\{\mathcal{A}\}$ generada por $\mathcal{A}$ se llama la ***σ*–álgebra de Borel** de $\mathbb{R}$ y se denota por $\mathcal{B}(\mathbb{R})$. Si $B \in \mathcal{B}(\mathbb{R})$ se dice que B es un **conjunto de Borel** de $\mathbb{R}$.

Un *conjunto abierto* en $\mathbb{R}$ se puede expresar como una unión numerable de intervalos abiertos. Por lo tanto, por las Definiciones 1.5 y 1.1(c), cualquier conjunto abierto en $\mathbb{R}$ es un conjunto de Borel. Luego, por 1.1(b), también cualquier *conjunto cerrado* en $\mathbb{R}$ es un conjunto de Borel. Asimismo, los intervalos de la forma $(a, b]$ o $[a, b)$ son conjuntos de Borel. Por ejemplo,

$$(a, b] = (-\infty, b] \cap (a, \infty) \in \mathcal{B}(\mathbb{R}).$$

La Definición 1.5 se puede extender al caso "vectorial" $\Omega = \mathbb{R}^n$ como sigue. Si $\boldsymbol{a} = (a_1, \ldots, a_n)$ y $\boldsymbol{b} = (b_1, \ldots, b_n)$ son vectores en $\mathbb{R}^n$, decimos que

$\boldsymbol{a} < \boldsymbol{b}$ si $a_i < b_i$ para $i = 1, \ldots, n$. En este caso definimos el "rectángulo abierto" $(\boldsymbol{a}, \boldsymbol{b})$ en $\mathbb{R}^n$ como el producto cartesiano

$$\begin{aligned}(\boldsymbol{a}, \boldsymbol{b}) &:= (a_1, b_1) \times \cdots \times (a_n, b_n) \\ &= \{\boldsymbol{x} \in \mathbb{R}^n | a_i < x_i < b_i \quad \text{para} \quad i = 1, \ldots, n\}.\end{aligned}$$

Definición 1.6. Sea $\Omega = \mathbb{R}^n$ y sea $\mathcal{A}$ la familia de todos rectángulos abiertos $(\boldsymbol{a}, \boldsymbol{b})$ en $\mathbb{R}^n$. La σ–álgebra generada por $\mathcal{A}$ se llama la **$\boldsymbol{\sigma}$–álgebra de Borel** de $\mathbb{R}^n$ y se denota por $\mathcal{B}(\mathbb{R}^n)$. Si B está en $\mathcal{B}(\mathbb{R}^n)$ se dice que B es un **conjunto de Borel** en $\mathbb{R}^n$.

Al igual que en el caso "escalar" ($n = 1$) se puede ver que cualquier *conjunto abierto* — y por tanto cualquier *conjunto cerrado* — en $\mathbb{R}^n$ es un conjunto de Borel.

Medidas

Definición 1.7. Sea $(\Omega, \mathcal{F})$ un espacio medible y $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$ el conjunto "extendido" de los números reales. Se dice que una función $\mu : \mathcal{F} \to \overline{\mathbb{R}}$ es una **medida** sobre $\mathcal{F}$ (o sobre $(\Omega, \mathcal{F})$) si

(a) $\mu(\phi) = 0$,

(b) $\mu(A) \geq 0$ para cada $A \in \mathcal{F}$,

(c) μ es **$\boldsymbol{\sigma}$–aditiva** en el sentido de que si $\{A_n\}$ es una sucesión de conjuntos **ajenos** en $\mathcal{F}$ (es decir, $A_n \cap A_m = \phi$ para $n \neq m$), entonces

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{1.1}$$

En este caso se dice que $(\Omega, \mathcal{F}, \mu)$ es un **espacio de medida**. A $\mu(A)$ se le llama la **medida del conjunto** $A \in \mathcal{F}$ (con respecto a μ). Si $\mu(\Omega) < \infty$, se dice que μ es una **medida finita**. En particular, si $\mu(\Omega) = 1$, se dice que μ es una **medida de probabilidad** (en forma abreviada: m.p.).

Si μ es una m.p. se acostumbra escribir $\mu \equiv \mathrm{P}$ y decimos que $(\Omega, \mathcal{F}, \mathrm{P})$ es un **espacio de probabilidad**. Asimismo, se dice que $\mathrm{P}(A)$ es la **probabilidad del evento** $A \in \mathcal{F}$.

Por otra parte, para ver que P es una m.p. basta verificar las condiciones 1.7(b), 1.7(c) y que $\mathrm{P}(\Omega) = 1$, porque de aquí se deduce trivialmente 1.7(a), es decir $\mathrm{P}(\phi) = 0$. (Vea la Proposición 1.9(a).)

Ejemplo 1.8. Sea $(\Omega, \mathcal{F})$ un espacio medible.
(a) La **medida trivial:** $\mu(A) = 0$ para todo $A \in \mathcal{F}$.
(b) Sea $\#(A)$ la **cardinalidad** (o número de elementos) de $A \in \mathcal{F}$. La **medida de conteo** es $\mu(A) := \#(A)$. Es decir,

$$\mu(A) = \begin{cases} n & \text{si } \#(A) = n < \infty, \\ +\infty & \text{en c.c. (caso contrario).} \end{cases}$$

(c) Para cada punto $\omega \in \Omega$ definimos la **medida de Dirac** en $\omega \in \Omega$ como

$$\delta_\omega(A) := \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{en c.c.} \end{cases}$$

A δ_ω también se le llama la **m.p.** (o **medida unitaria**) **concentrada en** $\omega \in \Omega$.
(d) **Medida de Lebesgue.** Sea $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si I es un intervalo en $\mathbb{R}$ de la forma (a, b) ó $[a, b]$ ó $(a, b]$ ó $[a, b)$, definimos la **longitud** de I como $\ell(\mathrm{I}) := b - a$. Asimismo, si $\mathrm{I}_1, \ldots, \mathrm{I}_k$ son intervalos *ajenos*, definimos

$$\ell(\bigcup_{i=1}^{k} \mathrm{I}_i) := \ell(\mathrm{I}_1) + \cdots + \ell(\mathrm{I}_k). \tag{1.2}$$

• Un resultado de Análisis Real (vea el Teorema 1.20) asegura que la longitud ℓ se puede "extender" a una medida única λ sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que

$$\lambda(\mathrm{I}) = \ell(\mathrm{I}) \quad \forall \text{ intervalo I},$$

y a λ se llama la **medida de Lebesgue** sobre $\mathbb{R}$. Algunas propiedades de λ, para $B \in \mathcal{B}(\mathbb{R}^n)$, son las siguientes.
• Si B es *abierto*, entonces $\lambda(B) > 0$.
• Si B es un conjunto *acotado*, entonces $\lambda(B) < \infty$.
• Si $B = \{x\}$ consiste de un único punto x, entonces $\lambda(B) = 0$. (Compare con la medida de Dirac δ_x.)
• Si $B = \{x_1, x_2, \ldots\}$ es un conjunto *numerable*, entonces $\lambda(B) = 0$.
Cabría preguntarse si se cumple el *recíproco* de esta última propiedad, es decir, si $\lambda(B) = 0$ implica que B es *a lo más numerable*, o sea finito o infinito numerable. La respuesta es *no*; vea el Ejercicio 1.15.

La medida de Lebesgue sobre $\mathbb{R}^n$ $(n \geq 2)$ se define de manera similar. Si $\mathrm{I} = (\boldsymbol{a}, \boldsymbol{b}) = (a_1, b_1) \times \cdots \times (a_n, b_n)$ es un rectángulo en $\mathbb{R}^n$ definimos su **volumen** como

$$v(\mathrm{I}) := (b_1 - a_1) \cdot (b_2 - a_2) \cdots (b_n - a_n).$$

También se define una propiedad análoga a (2) con v en lugar de ℓ. Entonces "se puede demostrar" que existe una y sólo una medida λ sobre $\mathbb{R}^n$ tal que

$$\lambda(\mathrm{I}) = v(\mathrm{I}) \quad \forall \text{ rectángulo } \mathrm{I} \subset \mathbb{R}^n.$$

En este caso decimos que λ es la **medida de Lebesgue sobre** $\mathbb{R}^n$.

Proposición 1.9. (Propiedades de P) Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad, y A, B eventos en $\mathcal{F}$.

(a) $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$

(b) $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(B \cap A)$. En particular, $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A)$ si $A \subset B$.

(c) Propiedad de monotonía: $\mathrm{P}(A) \leq \mathrm{P}(B)$ si $A \subset B$.

(d) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$.

(e) Si $A_1, \ldots, A_n$ están en $\mathcal{F}$,

$$\mathrm{P}\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} \mathrm{P}(A_i).$$

Definición 1.10. (**Sucesiones monótonas de eventos**) Sea $\{A_n\}$ una sucesión de subconjuntos de Ω.

(a) Si $A_n \subset A_{n+1}$ para todo $n = 1, 2, \ldots$, decimos que $\{A_n\}$ es una sucesión **creciente** (o **no–decreciente**) y que converge al límite $A^+ := \bigcup_{n=1}^{\infty} A_n$. (En forma abreviada escribimos $A_n \uparrow A^+$.)

(b) Si $A_n \supset A_{n+1}$ para todo $n = 1, 2, \ldots$, decimos que $\{A_n\}$ es una sucesión **decreciente** (o **no–creciente**) y que converge al límite $A^- := \bigcap_{n=1}^{\infty} A_n$. (En forma abreviada: $A_n \downarrow A^-$.)

Ejemplo 1.11. (a) Sea $A_n := [0, n]$ para $n = 1, 2, \ldots$. Entonces $A_n \uparrow A^+ := [0, \infty)$.

(b) Si $A_n := [0, 1 - 1/n]$, entonces $A_n \uparrow A^+ := [0, 1)$.

(c) Si $A_n := [0, 1/n]$, entonces $A_n \downarrow \{0\}$. Pero los intervalos *abiertos* $A_n := (0, 1/n) \downarrow \phi$. $\square$

Proposición 1.12. (**Continuidad de** P **con respecto a sucesiones monótonas**)

(a) Si $A_n \uparrow A^+$, entonces $\lim \mathrm{P}(A_n) = \mathrm{P}(A^+)$; de hecho, $\mathrm{P}(A_n) \uparrow \mathrm{P}(A^+)$.

(b) Si $A_n \downarrow A^-$, entonces $\lim \mathrm{P}(A_n) = \mathrm{P}(A^-)$; de hecho, $\mathrm{P}(A_n) \downarrow \mathrm{P}(A^-)$.

Corolario 1.13. *(**Desigualdad de Boole**) Si $\{A_n\}$ es una sucesión de eventos, entonces*

$$\mathrm{P}(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

Observación 1.14. (Terminología de conjuntos vs. enunciados probabilísticos.) Sean A y B dos eventos

Conjuntos	**Enunciado probabilístico**
$A \cap B$	A y B ocurren ($\equiv$ ambos ocurren)
$A \cup B$	A ó B ocurren ($\equiv$ al menos uno de los dos eventos ocurre)
A^c	A no ocurre
$A \cap B = \phi$	A y B son mutuamente excluyentes
$A - B = A \cap B^c$	A ocurre y B no ocurre
$A \subset B$	A implica B ($\equiv$ si A ocurre, entonces también B ocurre)
$A \Delta B = (A - B) \cup (B - A)$	Ocurren A ó B pero no ambos
$(A \cup B)^c = A^c \cap B^c$	No ocurren A ni B

Ejemplo 1.15. Considere tres eventos A, B y C. Encuentre una expresión y represente en un diagrama de Venn los eventos siguientes:

(a) Ocurre exactamente uno de los tres eventos A, B, C.

(b) Ocurren a lo más dos de los tres eventos.

(c) Ocurren los tres eventos (simultáneamente).

(d) Ocurren exactamente dos de los tres eventos.

(e) Ocurren A ó B pero no C.

(f) Ocurre A únicamente.

Observación 1.16. Si $\{x_n\}$ es una sucesión de números reales, definimos

$$\limsup x_n := \inf_{n\geq 1} \sup_{k\geq n} x_k, \quad \liminf x_n := \sup_{n\geq 1} \inf_{k\geq n} x_k.$$

Observe que la sucesión $\sup_{k\geq n} x_k$ $(n = 1, 2, \ldots)$ es decreciente (o no-creciente) y, por lo tanto, su límite existe y coincide con $\limsup x_n$, i.e.

$$\limsup x_n = \lim_{n\to\infty} \sup_{k\geq n} x_k.$$

Análogamente, $\inf_{k\geq n} x_k$ es creciente (o no–decreciente) y

$$\liminf x_n = \lim_{n\to\infty} \inf_{k\geq n} x_k.$$

En general, $\liminf x_n \leq \limsup x_n$. Si se cumple la igualdad, es decir,

$$\liminf x_n = \limsup x_n =: x,$$

se dice que la sucesión $\{x_n\}$ converge a x y escribimos $\lim x_n = x$ ó $x_n \to x$. (En el Ejercicio 1.5 veremos la definicion de lim-inf y lim-sup de sucesiones de *conjuntos).*

Construcción de medidas

Definición 1.17. Sea $\mathcal{A}$ una familia de subconjuntos de un conjunto Ω. Decimos que es $\mathcal{A}$ es una **álgebra** si

(a) $\Omega \in \mathcal{A}$;

(b) si A está en $\mathcal{A}$, su complemento A^c también está en $\mathcal{A}$;

(c) si $A_1, \ldots, A_n$ está en $\mathcal{A}$, entonces su unión $\bigcup_{i=1}^n A_i$ también está en $\mathcal{A}$.

Comparando 1.16 con la Definición 1.1 vemos que la diferencia entre una álgebra y una σ–álgebra es que la primera es cerrada bajo uniones *finitas*, mientras que la segunda es cerrada bajo uniones *numerables*.

Por otra parte, en 1.7 definimos una medida sobre una σ–álgebra. A continuación definiremos el concepto de medida sobre una *álgebra*.

Definición 1.18. Sea $\mathcal{A}$ una álgebra de subconjuntos de un conjunto Ω. Una **medida** μ sobre $\mathcal{A}$ es una función $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ tal que

(a) $\mu(\phi) = 0$,

(b) $\mu(A) \geq 0$ para todo $A \in \mathcal{A}$, y

(c) Si $\{A_n\}$ es una sucesión de conjuntos ajenos en $\mathcal{A}$ y cuya unión está en $\mathcal{A}$, entonces

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

Ejemplo 1.19. Volviendo al Ejemplo 1.8(d), sea $\mathcal{A}$ la familia de todas las uniones *finitas* de intervalos de la forma

$$(a, b],\ (-\infty, b],\ (a, +\infty),\ (-\infty, +\infty). \tag{1.3}$$

Además, si $I_1, \ldots, I_n$ son conjuntos ajenos de la forma (1.3), definimos la longitud de su unión como en (1.2), es decir

$$\ell(I_1 \bigcup \cdots \bigcup I_n) := \ell(I_1) + \cdots + \ell(I_n).$$

Entonces $\mathcal{A}$ es una *álgebra* y *la longitud* ℓ *es una medida sobre* $\mathcal{A}$. (Una demostración de este hecho se puede ver, por ejemplo, en el Lema 9.3 del libro: R.G. Bartle, *The Elements of Integration and Lebesgue Measure*, Wiley, 1996.)

En vista del Ejemplo 1.19, la pregunta es cómo extender la longitud ℓ sobre la álgebra $\mathcal{A}$ a la medida de Lebesgue λ sobre $\mathcal{B}(\mathbb{R})$. Esto es consecuencia del siguiente resultado; donde se usa el concepto de medida $\sigma-$finita, lo cual significa que existe una colección numerable de conjuntos medibles $\{A_i,\ i = 1, 2, \ldots\}$ tales que $\Omega = \bigcup_{i=1}^{\infty} A_i$ y, además, cada A_i tiene medida finita.

Teorema 1.20. *(Teorema de extensión de Carathéodory) Una medida* $\sigma-$*finita sobre una álgebra* $\mathcal{A}$ *se puede extender de manera única a una medida sobre la* σ–*álgebra generada por* $\mathcal{A}$.

En particular, si $\mathcal{A}$ es la álgebra en el Ejemplo 1.19, la σ–álgebra $\sigma\{\mathcal{A}\}$ generada por $\mathcal{A}$ es precisamente la σ–álgebra de Borel $\mathcal{B}(\mathbb{R})$, y la extensión de la longitud ℓ es la medida de Lebesgue λ.

Una pregunta obvia es si la medida de "longitud" se puede extender a todos los subconjuntos de $\mathbb{R}$. La respuesta es **no**, de acuerdo con el siguiente resultado de S. Ulam (en el cual se supone la validez de la "hipótesis del continuo").

Teorema 1.21. *(Teorema de Ulam) Sea μ una medida definida sobre todos los subconjuntos de $\mathbb{R}$ tal que $\mu((n, n+1]) < \infty$ para todo entero n, y $\mu(\{x\}) = 0$ para todo $x \in \mathbb{R}$. Entonces $\mu(A) = 0$ para todo $A \subset \mathbb{R}$.*

Una demostración de 1.21 se puede ver en la sección 3.4 del libro: I.K. Rana, *An Introduction to Measure and Integration, Second Edition*, American Mathematical Society, 2002.

Observación 1.22. (a) Se dice que un espacio de medida $(\Omega, \mathcal{F}, \mu)$ es *completo* (o que la σ–álgebra $\mathcal{F}$ es completa con respecto a μ) si $\mathcal{F}$ contiene a todos los subconjuntos de conjuntos de medida cero; es decir, si $A \in \mathcal{F}$ es tal que $\mu(A) = 0$ y $N \subset A$, entonces N está en $\mathcal{F}$.
(b) La *completación* de una σ–álgebra $\mathcal{F}$ con respecto a una medida μ se define como la mínima σ–álgebra $\hat{\mathcal{F}}$ que contiene a $\mathcal{F}$ y tal que si $A \in \mathcal{F}$ tiene μ–medida cero y $N \subset A$ entonces $N \in \hat{\mathcal{F}}$. Equivalentemente, la completación de $\mathcal{F}$ es

$$\hat{\mathcal{F}} := \{A \cup N \mid A \in \mathcal{F} \quad \text{y} \quad N \subset B \quad \text{para algún} \quad B \in \mathcal{F} \ \text{ con } \ \mu(B) = 0\}.$$

En este caso, la medida μ se extiende de manera única a una medida $\hat{\mu}$ sobre $\hat{\mathcal{F}}$ definida como $\hat{\mu}(A \cup N) := \mu(A)$. A $\hat{\mu}$ se le llama la *completación* de μ.
(c) La σ–álgebra de Borel, $\mathcal{B}(\mathbb{R})$, *no* es completa. Su completación es la llamada *σ–álgebra de Lebesgue*.

Ejercicios § 1

1.1. Demuestre las leyes (o fórmulas) de De Morgan: si $\{A_i, i \in \mathrm{I}\}$ es una colección arbitraria de subconjuntos de Ω, entonces
(a) $(\bigcup_i A_i)^c = \bigcap_i A_i^c$,   (b) $(\bigcap_i A_i)^c = \bigcup_i A_i^c$

1.2. Sea $(\Omega, \mathcal{F})$ un espacio medible. Demuestre que si A y B están en $\mathcal{F}$, entonces la *diferencia* $A - B := A \cap B^c$ y la *diferencia simétrica* $A \Delta B := (A \cup B) - (A \cap B)$ también están en $\mathcal{F}$.

1.3. Sea $(\Omega, \mathcal{F})$ un espacio medible y $B \subset \Omega$ un conjunto $\mathcal{F}$–medible. La *σ–álgebra $\mathcal{F}$ restringida a B* se define como la familia

$$\mathcal{F}(B) := \{A \cap B | A \in \mathcal{F}\}.$$

Demuestre que, efectivamente, $\mathcal{F}(B)$ es una σ–álgebra y, por lo tanto, $(B, \mathcal{F}(B))$ es un espacio medible. (Nótese que un conjunto C está en $\mathcal{F}(B)$ si y sólo si existe $A \in \mathcal{F}$ tal que $C = A \cap B$.) Algunas veces $\mathcal{F}(B)$ se escribe como $\mathcal{F} \cap B$.

Como ejemplo, sea $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y sea B el intervalo $[a, b]$. Entonces la pareja

$$([a,b], \mathcal{B}([a,b])), \quad \text{con} \quad \mathcal{B}([a,b]) = \mathcal{B}(\mathbb{R}) \cap [a,b],$$

es un espacio medible.

1.4. Sean Ω y Ω' dos conjuntos arbitrarios y $f : \Omega \to \Omega'$ una función dada. Si B es un *subconjunto* de Ω', definimos la **imagen inversa de** B con respecto a f como

$$f^{-1}(B) := \{\omega \in \Omega | f(\omega) \in B\}.$$

Si $\mathcal{C}$ es una *familia de subconjuntos* de Ω', definimos la **imagen inversa de** $\mathcal{C}$ con respecto a f como la familia de subconjuntos de Ω dada por

$$f^{-1}(\mathcal{C}) := \{f^{-1}(B) | B \in \mathcal{C}\}.$$

Demuestre:

(a) $f^{-1}(\Omega') = \Omega$

(b) Si B y C son subconjuntos de Ω', entonces $f^{-1}(C - B) = f^{-1}(C) - f^{-1}(B)$. En particular, $f^{-1}(B^c) = [f^{-1}(B)]^c$ y $f^{-1}(\phi) = \phi$.

(c) Si $\{B_i, i \in \mathrm{I}\}$ es una familia arbitraria de subconjuntos de Ω', entonces

$$f^{-1}(\bigcup_i B_i) = \bigcup_i f^{-1}(B_i) \quad \text{y} \quad f^{-1}(\bigcap_i B_i) = \bigcap_i f^{-1}(B_i)$$

(d) Si $\mathcal{F}'$ es una σ–álgebra de Ω', entonces la familia

$$f^{-1}(\mathcal{F}') := \{f^{-1}(B) | B \in \mathcal{F}'\}$$

es una σ–álgebra de Ω.

1.5. (Compare con la Observación 1.16.) Si $\{A_n\}$ es una sucesión de subconjuntos de Ω, definimos

$$\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

y

$$\liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Si $\liminf A_n = \limsup A_n =: A$, decimos que $\{A_n\}$ **converge a** A y, en este caso, escribimos $\lim A_n = A$ ó $A_n \to A$. Demuestre:

(a) $\liminf A_n \subset \limsup A_n$. Dé un ejemplo en el que $\liminf A_n \neq \limsup A_n$.

(b) Si $\{A_n\}$ es creciente o decreciente (vea la Definición 1.10), entonces $A_n \to A^+$ ó $A_n \to A^-$, respectivamente, en donde $A^+ := \cup A_n$ y $A^- := \cap A_n$.

En los incisos siguientes suponga que $(\Omega, \mathcal{F})$ es un espacio medible.

(c) Si $\{A_n\}$ es una sucesión en $\mathcal{F}$, entonces $\liminf A_n$ y $\limsup A_n$ están en $\mathcal{F}$. Además,

$$\mathrm{P}(\liminf A_n) \leq \liminf \mathrm{P}(A_n) \leq \limsup \mathrm{P}(A_n) \leq \mathrm{P}(\limsup A_n). \quad (*)$$

(*Sugerencia:* para demostrar la primera desigualdad en (*) primero note que la sucesión $\bigcap_{k=n}^{\infty} A_k$, para $n = 1, 2, \ldots$, es creciente; después use la definición de $\liminf A_n$ y la Proposición 1.12(a). La demostración de la tercera desigualdad en (*) es similar. Por último, observe que la segunda desigualdad se sigue de 1.16.)

(d) **Continuidad de** P: Deduzca de (*) que si $A_n \to A$, entonces $\mathrm{P}(A_n) \to \mathrm{P}(A)$.

1.6. Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida y B un conjunto $\mathcal{F}$–medible. Para cada $A \in \mathcal{F}$ definimos $\mu_B(A) := \mu(A \cap B)$. Demuestre:

(a) μ_B es una medida sobre $\mathcal{F}$, llamada **la restricción de** μ a B;

(b) si $0 < \mu(B) < \infty$, entonces

$$\mathrm{P}_B(A) := \mu_B(A)/\mu(B) = \mu(A \cap B)/\mu(B) \quad \forall\, A \in \mathcal{F}$$

es una m.p. sobre $\mathcal{F}$.

1.7. Sea $p > 0$ una constante y $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad en el que $\mathcal{F}$ es el conjunto potencia 2^{Ω}, y $\mathrm{P}(A) = p$ para cada conjunto $A = \{\omega\}$ con un solo punto $\omega \in \Omega$. Demuestre:

(a) Ω tiene sólo un número finito de puntos; de hecho $\#(\Omega) \leq 1/p$.

(b) Si $\#(\Omega) = n$, entonces $p = 1/n$.

1.8. Sean $\mathrm{P}_1, \ldots, \mathrm{P}_n$ m.p.'s sobre $(\Omega, \mathcal{F})$, y $\alpha_1, \ldots, \alpha_n$ números no negativos con $\alpha_1 + \cdots + \alpha_n = 1$. Demuestre que la "combinación convexa" $\mathrm{P} := \sum_{i=1}^{n} \alpha_i \mathrm{P}_i$ es una m.p.

1.9. Demuestre que $|\mathrm{P}(A) - \mathrm{P}(B)| \leq \mathrm{P}(A \Delta B)$ para cualesquiera dos eventos A y B. ($A \Delta B$ es la *diferencia simétrica* definida en el Ejercicio 2.)

Observación 1.23. Dadas dos medidas μ_1, μ_2 sobre un espacio medible $(\Omega, \mathcal{F})$, se dice que $\mu_1 \leq \mu_2$ si $\mu_1(A) \leq \mu_2(A)$ para todo $A \in \mathcal{F}$. La igualdad $\mu_1 = \mu_2$ se define análogamente.

1.10. Demuestre que si $\mathrm{P}_1, \mathrm{P}_2$ son dos medidas de probabilidad sobre $(\Omega, \mathcal{F})$ y $\mathrm{P}_1 \leq \mathrm{P}_2$, entonces $\mathrm{P}_1 = \mathrm{P}_2$.

1.11. Sea $\{\mu_n\}$ una *sucesión creciente* de medidas sobre $\mathcal{F}$, es decir, $\mu_n \leq \mu_{n+1}$ para todo n. Defina, para cada $A \in \mathcal{F}$,

$$\mu(A) := \lim_{n \to \infty} \mu_n(A)$$

Demuestre que μ es una medida sobre $\mathcal{F}$ y que $\mu(A) = \sup_{n \geq 1} \mu_n(A)$ para todo $A \in \mathcal{F}$.

1.12. Demuestre que para cualquiera n eventos $A_1, \ldots, A_n$, con $n \geq 2$,

$$\mathrm{P}(A_1 \cap \ldots \cap A_n) \geq \mathrm{P}(A_1) + \cdots + \mathrm{P}(A_n) - n + 1.$$

1.13. Demuestre: si $\{A_n\}$ es una sucesión de eventos tales que $\mathrm{P}(A_n) = 1$ para todo n, entonces $\mathrm{P}(\bigcap\limits_{n=1}^{\infty} A_n) = 1$.

1.14. (a) Demuestre el **Lema de Borel–Cantelli** que dice lo siguiente: si $\{A_n,\ n \geq 1\}$ es una sucesión de eventos tales que $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$, entonces $\mathrm{P}(\limsup A_n) = 0$.
(b) Sea $(\Omega, \mathcal{F}, \mathrm{P})$ el espacio "unitario" $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$ y $\mathrm{P} = \lambda$ la medida de Lebesgue. Considere la sucesión de eventos $A_n := [0, 1/n]$ para $n = 1, 2, \ldots$. Calcule $\limsup A_n$, $\mathrm{P}(\limsup A_n)$, y $\sum_{n=1}^{\infty} \mathrm{P}(A_n)$. Diga si se cumple el *recíproco* del Lema de Borel–Cantelli.

1.15. Sea $C \subset [0,1]$ el *conjunto de Cantor* que se define como sigue. Tómese el intervalo [0,1] y elimine el intervalo abierto que consiste del "tercio medio" (1/3, 2/3). De cada una de las dos partes restantes [0,1/3] y [2/3,1] se elimina el tercio medio abierto, o sea, (1/9,2/9) y (7/9,8/9). De cada una de las cuatro partes restantes se eliminan los tercios medios abiertos (1/27,2/27), (7/27,8/27), (19/27,20/27), y (25/27,26/27). Procediendo de manera inductiva, de los subintervalos que quedan en la n–ésima etapa se eliminan los 2^{n-1} tercios medios abiertos, cada uno de longitud $1/3^n$. El conjunto de Cantor es lo que resulta del procedimiento anterior cuando $n \to \infty$. Demuestre que C es un conjunto no–numerable que tiene medida de Lebesgue cero.

2 Espacios discretos y continuos

Contenido: Función de densidad discreta, permutaciones, combinaciones, función de densidad continua.

En esta sección, como ejemplos de espacios de probabilidad introducimos los espacios discretos y continuos que son de uso muy común en probabilidad.

Definición 2.1. Decimos que un espacio medible $(\Omega, \mathcal{F})$ es **discreto** si Ω es un conjunto finito o infinito numerable, en cuyo caso la σ–álgebra $\mathcal{F}$ es el conjunto potencia de Ω, es decir, $\mathcal{F} = 2^{\Omega}$.

Sea $(\Omega, \mathcal{F})$ un espacio medible discreto y $f : \Omega \to [0,1]$ una función tal que

$$\text{(i)} \quad f(x) \geq 0 \quad \forall\, x \in \Omega, \quad \text{y} \quad \text{(ii)} \quad \sum_{x\in\Omega} f(x) = 1. \tag{2.1}$$

En este caso decimos que f es una **función de densidad discreta** y definimos la **m.p. asociada** a f como

$$\mathrm{P}_f(A) := \sum_{x\in A} f(x) \quad \forall\, A \subset \Omega. \tag{2.2}$$

Podemos escribir (2.2) en varias formas equivalentes. Por ejemplo, usando las medidas de Dirac δ_x introducidas en el Ejemplo 1.8(c), podemos expresar (2.2) como

$$\mathrm{P}_f(A) = \sum_{x\in\Omega} f(x)\delta_x(A).$$

Asimismo, si $\mathrm{I}_A := \Omega \to \mathbb{R}$ es la **función indicadora del evento** A, que se define como

$$\mathrm{I}_A(x) := \begin{cases} 1 & \text{si } \ x \in A, \\ 0 & \text{si } \ x \notin A, \end{cases} \tag{2.3}$$

vemos que (2.2) resulta

$$\mathrm{P}_f(A) = \sum_{x\in\Omega} f(x)\mathrm{I}_A(x).$$

Algunas funciones de densidad discretas muy comunes son las siguientes.

Ejemplo 2.2. (a) La **densidad uniforme.** Supóngase que Ω es un conjunto finito, digamos $\Omega = \{x_1, \ldots, x_n\}$, y sea $f(x_i) := 1/n$ para todo $i = 1, \ldots, n$. Entonces (2.2) resulta

$$\mathrm{P}_f(A) = \#(A)/n \quad \forall\, A \subset \Omega.$$

Por ejemplo, en el lanzamiento de una moneda, $\Omega = \{a, s\}$ y $f(x) = 1/2$. En el lanzamiento de un dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$ y $f(x) = 1/6$. En este último caso, $\mathrm{P}_f(A) = 1/2$ si A es cualquier subconjunto con tres elementos.

(b) La **densidad binomial**. Supóngase que $\Omega = \{0, 1, \ldots, n\}$. Si k es un entero no negativo, el **factorial** de k es $k! := 1 \cdot 2 \cdots k$ (con $0! := 1$) y el **coeficiente binomial**

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} \quad \text{para} \quad k = 0, 1, \ldots, n.$$

Entonces, dado un número $0 < p < 1$, definimos la **densidad binomial** $f : \Omega \to \mathbb{R}$ como

$$f(k) := \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para} \quad k = 0, \ldots, n. \tag{2.4}$$

Usando el **Teorema del Binomio**

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \, b^{n-k} \tag{2.5}$$

se puede verificar que, efectivamente, f es una función de densidad discreta.

Caso especial: si $n = 1$, entonces $\Omega = \{0, 1\}$ y a f se le llama **densidad Bernoulli**, que sólo toma los valores $f(1) = p$ y $f(0) = 1 - f(1) = 1 - p$.

(c) La **densidad geométrica.** Supóngase que $\Omega = \{0, 1, \ldots\}$ y sea $p \in (0, 1)$ un número dado. Entonces

$$f(k) := p \cdot (1-p)^k \quad \text{para} \quad k = 0, 1, \ldots$$

es la **función de densidad geométrica.** Para verificar que f satisface (2.1) use la **serie geométrica**

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r} \quad \text{si} \quad |r| < 1.$$

(d) La **densidad de Poisson.** Sea $\Omega = \{0, 1, \ldots\}$ y $\lambda > 0$ un número dado. La función

$$f(k) := e^{-\lambda} \lambda^k / k! \quad \text{para} \quad k = 0, 1, \ldots,$$

se llama la **densidad de Poisson con parámetro** λ. Para verificar (2.1) use la **serie exponencial**

$$e^r = \sum_{k=0}^{\infty} r^k/k! \quad \forall\, r \in \mathbb{R}.$$

Los siguientes conceptos son útiles para calcular probabilidades sobre conjuntos finitos.

Definición 2.3. Sea Ω un conjunto que consiste de n elementos y sea k un entero entre 0 y n.

(a) Una **permutación** de orden k (de los elementos de Ω) es una selección ordenada, sin repeticiones, de k elementos de Ω.

(b) Una **combinación** de orden k (de los elementos de Ω) es un subconjunto de Ω con k elementos.

Por ejemplo, sea $\Omega = \{a, b, c\}$ y $k = 2$. Entonces las *permutaciones* de orden 2 son

$$(a, b), (a, c), (b, a), (b, c), (c, a) \;\text{ y }\; (c, b),$$

mientras que las *combinaciones* de orden 2 son

$$\{a, b\}, \{a, c\} \;\text{ y }\; \{b, c\}.$$

Denotaremos por $p(n, k)$ el número de permutaciones de orden k de un conjunto con n elementos, y por $c(n, k)$ el número de combinaciones. Para calcular $p(n, k)$ usaremos el siguiente "principio".

Principio 2.4. (**de la multiplicación**) Considérense k tareas, digamos $A_1, \ldots, A_k$, tales que A_1 se puede realizar en n_1 formas y, para $i \geq 2$, A_i se puede realizar en n_i formas una vez que $A_1, \ldots, A_{i-1}$ se han realizado. Entonces la tarea total $(A_1, A_2, \ldots, A_k)$ en sucesión se puede realizar en $n_1 \cdot n_2 \cdots n_k$ formas.

Usando 2.4 vemos que el número de permutaciones de orden k es

$$p(n, k) = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!} \quad \text{para} \quad k = 1, \ldots, n. \tag{2.6}$$

Por otra parte, para cada combinación de orden k tenemos $k!$ permutaciones de orden k, o sea

$$p(n,k) = c(n,k) \cdot k!$$

Por lo tanto, el número de combinaciones de orden k es

$$c(n,k) = \frac{p(n,k)}{k!} = \frac{n!}{k!(n-k)!} \qquad \text{[por (6)]},$$

es decir,

$$c(n,k) = \binom{n}{k}. \tag{2.7}$$

Ejemplo 2.5. Se desea formar un comité de 4 profesores de un grupo que consiste de 10 profesores adjuntos y 6 titulares.

(a) ¿Cuántos comités se pueden formar en total? Respuesta: $\binom{16}{4}$.

(b) ¿Cuál es la probabilidad de que todos los miembros del comité sean profesores adjuntos? Respuesta: $\binom{10}{4} / \binom{16}{4}$.

Ejemplo 2.6. En un lote de 100 artículos hay 6 defectuosos. Si se toman del lote 5 artículos **al azar**, ¿cuál la probabilidad de exactamente 3 de ellos sean defectuosos? Respuesta: $\binom{6}{3}\binom{94}{2} / \binom{100}{5}$.

Ejemplo 2.7. Considérese un código binario de sucesiones de $0's$ y $1's$. ¿Cuántas "palabras" se pueden codificar usando exactamente n símbolos? Respuesta: 2^n.

Ejemplo 2.8. Demuestre que si un conjunto A consiste de n elementos, entonces A tiene 2^n subconjuntos.

Solución. El número de subconjuntos de A

$$\begin{aligned} &= \sum_{k=0}^{n} \text{ (número de subconjuntos con } k \text{ elementos)} \\ &= \sum_{k=0}^{n} \binom{n}{k} \qquad \text{(por (2.7))} \\ &= (1+1)^n = 2^n \quad \text{(por (2.5))}. \quad \square \end{aligned}$$

Definición 2.9. Sea $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Se dice que $f : \mathbb{R} \to [0, 1]$ es una **función de densidad continua** si $f(x) \geq 0$ para todo $x \in \mathbb{R}$ y

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

En este caso, podemos considerar la integral de Lebesgue que veremos posteriormente, y definir la **m.p. asociada** a f es la m.p. P_f sobre $\mathcal{B}(\mathbb{R})$

$$\mathrm{P}_f(A) := \int_A f(x)dx \quad \forall\, A \in \mathcal{B}(\mathbb{R}). \tag{2.8}$$

Nota 2.10. La m.p. P_f en (2.8) es "continua" en un sentido que especificaremos posteriormente, pero la función de densidad f **no** necesariamente es continua; vea el Ejemplo 2.11(a) ó (b).

Equivalentemente, (2.8) se puede expresar como

$$\mathrm{P}_f(A) = \int_{-\infty}^{\infty} f(x)\, \mathrm{I}_A(x)dx, \tag{2.9}$$

en donde I_A es la función indicadora de A definida en (2.3). La condición de σ–aditividad en la Definición 1.7(c) significa que si $\{A_n\}$ es una sucesión de conjuntos *ajenos* en $\mathcal{B}(\mathbb{R})$ y $A = \cup A_n$, entonces

$$\mathrm{P}_f(A) = \sum_n \mathrm{P}_f(A_n) = \sum_n \int_{A_n} f(x)dx.$$

Esta igualdad se puede demostrar escribiendo la función indicadora de $A = \cup A_n$ como

$$\mathrm{I}_A = \sum_n \mathrm{I}_{A_n} \quad \text{(pues los } A_n \text{ son ajenos)}$$

y (9) resulta

$$\begin{aligned}
\mathrm{P}_f(A) &= \int_{-\infty}^{\infty} f(x) \left[\sum_n \mathrm{I}_{A_n}(x)\right] dx \\
&= \int_{-\infty}^{\infty} \left[\sum_n f(x)\mathrm{I}_{A_n}(x)\right] dx \\
&= \sum_n \int_{-\infty}^{\infty} f(x)\mathrm{I}_{A_n}(x)(dx) \quad \text{(explique)} \\
&= \sum_n \mathrm{P}_f(A_n).
\end{aligned}$$

Ejemplo 2.11. (a) La **densidad uniforme** sobre un intervalo $[a, b]$ es la función

$$\begin{aligned} f(x) &:= 1/(b-a) \quad \text{si} \quad x \in [a, b], \\ &:= 0 \quad \text{en c.c.} \end{aligned}$$

Es fácil ver que f satisface la Definición 2.11.
(b) La **densidad exponencial** con parámetro λ ($\lambda > 0$) se define como

$$\begin{aligned} f(x) &:= \lambda e^{-\lambda x} \quad \text{si} \quad x \geq 0, \\ &:= 0 \quad \text{en c.c.} \end{aligned}$$

Nótese que $f(\cdot) \geq 0$ y

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_0^{\infty} f(x)dx \\ &= \lim_{y\to\infty} \int_0^y \lambda e^{-\lambda x} dx \\ &= \lim_{y\to\infty} (1 - e^{-\lambda y}) = 1. \end{aligned}$$

(c) La **densidad de Cauchy** se define para todo $x \in \mathbb{R}$ como

$$f(x) := \frac{1}{\pi(1 + x^2)}.$$

Entonces $f(\cdot) \geq 0$ y

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dx}{1 + x^2} \\ &= \frac{1}{\pi} \cdot \text{ arc tan } x|_{x=-\infty}^{x=+\infty} \\ &= \frac{1}{\pi} \cdot [\frac{\pi}{2} - (-\frac{\pi}{2})] = 1. \end{aligned}$$

(d) La **densidad normal** (o **gaussiana**) con parámetros $m \in \mathbb{R}$ y $\sigma^2 > 0$ es la función

$$f(x) := (2\pi\sigma^2)^{-1/2} \, e^{-(x-m)^2/2\sigma^2} \quad \forall \, x \in \mathbb{R}. \tag{2.10}$$

En particular, si $m = 0$ y $\sigma^2 = 1$ se obtiene la **densidad normal estándar**

$$\varphi(x) := (2\pi)^{-1/2} \, e^{-x^2/2}.$$

Sea
$$\alpha := \int_{-\infty}^{\infty} e^{-x^2/2} dx.$$
Entonces, usando coordenadas polares (r, θ)
$$\begin{aligned}
\alpha^2 &= \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx\, dy \\
&= \int_{0}^{\infty} \int_{-\pi}^{\pi} e^{-r^2/2}\, r\, d\theta\, dr \\
&= 2\pi \int_{0}^{\infty} e^{-r^2/2}\, r\, dr \\
&= -2\pi\, e^{-r^2/2} \Big|_{r=0}^{r=\infty} \\
&= 2\pi.
\end{aligned}$$
Luego $\alpha = \sqrt{2\pi}$ y de aquí se sigue que $\int_{-\infty}^{\infty} \varphi(x) dx = 1$. El resultado análogo para la función f en (2.10) se obtiene usando el hecho de que
$$f(x) = \sigma^{-1} \varphi((x - m)/\sigma).$$
Para concluir, observe que $\alpha = \sqrt{2\pi}$ se puede escribir explícitamente como
$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}. \tag{2.11}$$

Nota 2.12. Hay medidas de probabilidad que no necesariamente están definidas por una densidad discreta o continua como en (2.1) ó (2.9). Por ejemplo, sea δ_0 la medida de Dirac concentrada en $x = 0$, y sea f la densidad normal estándar. Entonces
$$\mu(A) := \frac{1}{2} \left(\delta_0(A) + \int_A f(x) dx \right)$$
es una “mezcla” de una densidad discreta y una continua. Posteriormente veremos que una medida de probabilidad es una mezcla de varios tipos de distribuciones. $\square$

Ejercicios § 2

2.1. Sea $f(\cdot)$ una función de densidad discreta. Si

$$\sum_x |x| f(x) < \infty, \tag{2.12}$$

entonces el **valor medio** (o "centro de gravedad") $\overline{f}$ de f se define como

$$\overline{f} := \sum_x x\, f(x).$$

Si no se cumple la condición (2.12) (es decir, $\sum_x |x| f(x) = \infty$), se dice que el valor medio de f **no** existe.

(a) Calcule el valor medio de las densidades en el Ejemplo 2.2(b), (c) y (d).

(b) Sea $r > 1$, y $f(n) := c/n^r$ para $n = 1, 2, \ldots$, en donde $c > 0$ es una constante tal que $\sum_{n=1}^{\infty} f(n) = 1$. Demuestre que la densidad de f tiene valor medio ssi $r > 2$. (**Nota.** La serie $\sum_{n=1}^{\infty} n^{-k}$ converge ssi $k > 1$.)

2.2. Si $f(\cdot)$ es una función de densidad continua tal que

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty, \tag{2.13}$$

entonces el **valor medio** (o "centro de gravedad") $\overline{f}$ de f se define como

$$\overline{f} := \int_{-\infty}^{\infty} x\, f(x) dx.$$

(a) Calcule el valor medio de las densidades en el Ejemplo 2.11(a), (b) y (d).

(b) Demuestre que el valor medio de la densidad de Cauchy en el Ejemplo 2.11(c) **no** existe, es decir, la condición (2.13) no se satisface.

2.3. Suponga que de un conjunto de n objetos se eligen k al azar ($k < n$), uno tras otro, *con* sustitución. Calcule la probabilidad de que ningún objeto sea elegido más de una vez.

2.4. De un conjunto que consiste de $n > 1$ números positivos y $m > 1$ negativos se selecciona al azar un subconjunto de dos números y se multiplican. ¿Cuál es la probabilidad de que el producto sea positivo?

2.5. Sea f la densidad normal con parámetros m y σ^2, definida en (2.10), y sea

$$\begin{aligned} g(x) &:= 0 \;\text{ si }\; x \leq 0, \\ &:= \frac{1}{x} f(\log x) \;\text{ si }\; x > 0. \end{aligned}$$

Demuestre que g es una densidad de probabilidad — se le llama la *densidad lognormal con parámetros* m y σ^2. (Vea el Ejercicio 6.17.)

2.6. (a) Demuestre que la *función gama*, definida como

$$\Gamma(p) := \int_0^\infty x^{p-1} e^{-x} dx \quad \text{para} \quad p > 0,$$

satisface que $\Gamma(p+1) = p \cdot \Gamma(p)$.
(b) Sean p y λ números positivos y defina

$$f(x) := \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \quad \text{para} \quad x > 0, \quad \text{y} \quad f(x) := 0 \quad \text{para} \quad x \leq 0.$$

Demuestre que f es una función de densidad de probabilidad. A f se le llama la *densidad gama con parámetros* p y λ. En particular, si $p = 1$ la función f es la *densidad exponencial* con parámetro λ en el Ejemplo 2.11(b).

3 Probabilidad condicional e independencia

Contenido: Probabilidad condicional, regla de la multiplicación, ley de la probabilidad total, fórmula de Bayes, independencia de eventos y de σ–álgebras, el lema de Borel–Cantelli, la ley 0–1 de Kolmogorov.

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y sean A y B dos eventos en $\mathcal{F}$, con $\mathrm{P}(B) > 0$. Definimos la **probabilidad condicional de** A **dado** B como

$$\mathrm{P}(A|B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}. \tag{3.1}$$

En este caso se dice que B es el **evento condicionante**. En expresiones como (3.1) siempre supondremos que el evento condicionante tiene probabilidad positiva.

En el Ejercicio 3.1 se pide demostrar la siguiente proposición.

Proposición 3.1. Sea B tal que $\mathrm{P}(B) > 0$. La función $A \mapsto \mathrm{P}(A|B)$ definida como en (1), para todo $A \in \mathcal{F}$, es una m.p.

Por ejemplo, en el lanzamiento de tres monedas "honestas" el espacio muestral es

$$\Omega = \{aaa, aas, asa, saa, ass, sas, ssa, sss\}.$$

Calcule la probabilidad de que "a lo más una moneda cae águila" dado que "en la primera moneda cae águila". Tomamos

$$A = \{sss, ass, sas, ssa\}, \quad B = \{aaa, aas, asa, ass\}.$$

Entonces $\mathrm{P}(A|B) = 1/4$.

De (3.1) tenemos $\mathrm{P}(A \cap B) = \mathrm{P}(B)\mathrm{P}(A|B)$ si $\mathrm{P}(B) > 0$. Este es un caso particular (para dos eventos) de la siguiente "regla de la multiplicación".

Proposición 3.2. (**Regla de la multiplicación**) Sean $A_1, A_2, \ldots, A_n$ eventos tales que

$$\mathrm{P}(A_1 \cap \ldots \cap A_{n-1}) > 0. \tag{3.2}$$

Entonces

$$\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathrm{P}(A_1) \cdot \mathrm{P}(A_2|A_1) \cdot \mathrm{P}(A_3|A_1 \cap A_2) \cdots \mathrm{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Obsérvese que las probabilidades condicionales en esta última expresión están bien definidas porque

$$\mathrm{P}(A_1) \geq \mathrm{P}(A_1 \cap A_2) \geq \cdots \geq \mathrm{P}(A_1 \cap \ldots \cap A_{n-1}) > 0 \quad \text{(por (3.2))}.$$

Ejemplo 3.3. En un laboratorio de computación hay 14 computadoras de las cuales 8 son de la marca x. Se seleccionan tres computadoras al azar, una tras otra. ¿Cuál es la probabilidad de que las tres sean de la marca x?
Solución. Sea A_i $(i = 1, 2, 3)$ el evento "la i–ésima computadora seleccionada es de la marca x". Entonces

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\mathrm{P}(A_3|A_1 \cap A_2) = \frac{8}{14} \cdot \frac{7}{13} \cdot \frac{6}{12}. \quad \Box$$

Definición 3.4. Una familia $\{A_i,\ i \in \mathrm{I}\}$ de eventos es una **partición** de Ω si

(a) los eventos son ajenos (es decir, $A_i \cap A_j = \phi$ para $i \neq j$), y

(b) $\cup_i A_i = \Omega$.

El ejemplo más simple de una partición es $\{A, A^c\}$. En este caso es evidente que para cualquier evento B se tiene

$$\mathrm{P}(B) = \mathrm{P}(A)\mathrm{P}(B|A) + \mathrm{P}(A^c)\mathrm{P}(B|A^c) \quad \text{(explique).}$$

Este es un caso particular de la "ley de la probabilidad total" en el siguiente teorema.

Teorema 3.5. *Sea* $\{A_1, \ldots, A_n\}$ *una partición de* Ω *tal que* $\mathrm{P}(A_i) > 0$ *para* $i = 1, \ldots, n$. *Entonces*

(a) **ley de la probabilidad total:** *para cualquier evento* B

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(A_i)\mathrm{P}(B|A_i). \tag{3.3}$$

(b) **Fórmula de Bayes:** *si* $\mathrm{P}(B) > 0$, *entonces*

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(A_i \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A_i)\mathrm{P}(B|A_i)}{\sum_j \mathrm{P}(A_j)\mathrm{P}(B|A_j)} \tag{3.4}$$

para $i = 1, \ldots, n$.

En el teorema anterior las probabilidades $\mathrm{P}(A_i)$ y $\mathrm{P}(A_i|B)$ se llaman probabilidades **a priori** y **a posteriori**, respectivamente.

Ejemplo 3.6. En una fábrica, tres máquinas M_1, M_2 y M_3 elaboran, respectivamente, el 30%, el 50% y el 20% de la producción total. Los porcentajes de artículos defectuosos producidos por estas máquinas son 1%, 3% y 2%, respectivamente. Si se selecciona un artículo al azar, calcule la probabilidad de que el artículo

(a) sea defectuoso,

(b) no sea defectuoso,

(c) haya sido producido en la máquina M_i $(i = 1, 2, 3)$ dado que resultó ser defectuoso.

Definición 3.7. (**Independencia**)

(a) Dos *eventos* A y B son **independientes** si

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B).$$

Dos *σ–álgebras* $\mathcal{F}_1$ y $\mathcal{F}_2$ de Ω son **independientes** si cualquiera dos eventos $A_1 \in \mathcal{F}_1$ y $A_2 \in \mathcal{F}_2$ son independientes.

(b) Una colección $\{A_i,\ i \in I\}$ de eventos es **independiente** si para cada entero positivo k y cada selección de índices distintos $i_1, \ldots, i_k$ en I se cumple que

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \mathrm{P}(A_{i_1}) \cdot \mathrm{P}(A_{i_2}) \ldots \mathrm{P}(A_{i_k}). \tag{3.5}$$

Una colección $\{\mathcal{A}_i,\ i \in I\}$ de familias de eventos — en particular, σ–álgebras — es **independiente** (o las familias $\mathcal{A}_i$, $i \in I$, son independientes) si para cada entero positivo k, cada selección de índices distintos $i_1, \ldots, i_k$ en I, y todo $A_1 \in \mathcal{A}_{i_1}, \ldots, A_k \in \mathcal{A}_{i_k}$, se cumple la condición (3.5).

Nótese que si $\mathrm{P}(B) > 0$, entonces A y B son independientes ssi $\mathrm{P}(A|B) = \mathrm{P}(A)$.

Observación 3.8. Para $n \geq 3$ eventos, independencia "por parejas" **no** implica independencia. Por ejemplo, sea $\Omega = \{a, b, c, d\}$ un espacio muestral equiprobable, y sean $A_1 = \{a, b\}$, $A_2 = \{b, c\}$ y $A_3 = \{a, c\}$. Entonces A_1, A_2, A_3 son independientes "por parejas" porque

$$\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i) \cdot \mathrm{P}(A_j) \quad \text{para } i \neq j.$$

Sin embargo, los eventos no son independientes (en el sentido de la Definición 3.7(b)) porque

$$0 = \mathrm{P}(A_1 \cap A_2 \cap A_3) \neq \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3) = \frac{1}{8}.$$

También puede haber eventos $A_1, \ldots, A_n$ que **no** son independientes pero $\mathrm{P}(A_1 \cap \ldots \cap A_n) = \mathrm{P}(A_1) \cdots \mathrm{P}(A_n)$; vea el Ejercicio 3.2.

Ejemplo 3.9. Considérese un "experimento de Bernoulli", es decir, un experimento que tiene sólo dos resultados posibles, éxito (1) ó fracaso (0), con probabilidades p y $q := 1 - p$, respectivamente, con $0 < p < 1$. Suponga que se realizan n repeticiones independientes del experimento y calcule la probabilidad de que ocurran exáctamente k éxitos ($0 \leq k \leq n$).
Solución. El espacio muestral Ω consiste de todos los vectores

$$(x_1, x_2, \ldots, x_n) \quad \text{con} \quad x_i = 1 \quad \text{ó} \quad x_i = 0 \quad \text{para} \quad i = 1, \ldots, n.$$

Sea A_i el evento "la i–ésima repetición del experimento tiene éxito". Por ejemplo, el evento "en las primeras k repeticiones ocurren éxitos y en las restantes $n - k$ fracasos" es

$$A_1 \cap \ldots \cap A_k \cap A_{k+1}^c \cap \ldots \cap A_n^c$$

y consiste del vector $(1, \ldots, 1$ (k veces), $0 \ldots, 0$ ($n - k$ veces)). Por independencia, la probabilidad de dicho evento es igual a

$$\mathrm{P}(A_1) \cdots \mathrm{P}(A_k) \cdot \mathrm{P}(A_{k+1}^c) \cdots \mathrm{P}(A_n^c) = p^k q^{n-k}.$$

Por otra parte, nótese que el número total de vectores $(x_1, \ldots, x_n)$ en los que exáctamente k componentes toman el valor 1 es el número de combinaciones

$$c(n, k) = \binom{n}{k}.$$

Por lo tanto,

$$\mathrm{P}(\text{ocurren exáctamente } k \text{ éxitos}) = \binom{n}{k} p^k q^{n-k} \tag{3.6}$$

para cualquier $k = 0, 1, \ldots, n$, que coincide con la densidad binomial del Ejemplo 2.2(b).

Ejemplo 3.10. (Caso especial del Ejemplo 3.9) Por experiencia, la administración de un restaurante sabe que el 20% de las personas que hacen reservación no se presentan. Si el restaurante tiene 50 mesas y toma 52 reservaciones, ¿cuál es la probabilidad de que haya lugar para todos los clientes que se presentan?
Solución. El ejemplo se puede expresar como $n = 52$ repeticiones independientes de un experimento de Bernoulli con probabilidad de "éxito" (no se presenta un cliente) $p = 0.20$. Por lo tanto, usando (3.6) la probabilidad que se desea calcular resulta

$$\mathrm{P}(\text{No. de éxitos } \geq 2) = \sum_{k=2}^{52} \binom{n}{k} p^k q^{n-k}$$

con $q = 1 - p = 0.80$. De hecho, es más fácil calcular

$$\begin{aligned} \mathrm{P}(\text{No. de éxitos } \geq 2) &= 1 - \mathrm{P}(\text{No. de éxitos } \leq 1) \\ &= 1 - q^{52} - 52\, p \cdot q^{51}. \quad \Box \end{aligned}$$

En el Ejercicio 1.14(a) vimos el Lema de Borel–Cantelli, de acuerdo con el cual $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$ implica que $\mathrm{P}(\limsup A_n) = 0$. En la parte (b) del mismo ejercicio se ve que el recíproco de este resultado no se cumple. La versión completa de dicho lema es como sigue.

Lema 3.11. (**de Borel–Cantelli**) Sea $\{A_n\}$ una sucesión de eventos.

(a) Si $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$, entonces $\mathrm{P}(\limsup A_n) = 0$.

(b) Si $\sum_{n=1}^{\infty} \mathrm{P}(A_n) = \infty$ y los eventos A_n son *independientes*, entonces $\mathrm{P}(\limsup A_n) = 1$.

Demostración. (a)

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0.$$

(b) Como $\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, se tiene que $(\limsup A_n)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$. Así, usando que los complementos A_k^c son también eventos independientes (vea Ejercicio 3.3),

$$\begin{aligned}
P((\limsup A_n)^c) &= \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty}\lim_{m\to\infty} P\left(\bigcap_{k=n}^{m} A_k^c\right) \\
&= \lim_{n\to\infty}\lim_{m\to\infty} \Pi_{k=n}^{m}(1-P(A_k)) \le \lim_{n\to\infty} \Pi_{k=n}^{\infty} e^{-P(A_k)} \\
&= \lim_{n\to\infty} e^{-\sum_{k=n}^{\infty} P(A_k)} = 0,
\end{aligned}$$

donde hemos usado que $e^{-x} \ge 1-x$ para $x \ge 0$. Luego $P(\limsup A_n) = 1$.
$\square$

Una aplicación de este lema la veremos en el Ejemplo 9.23 en el contexto de variables aleatorias.

El Lema de Borel–Cantelli afirma que si $\{A_n\}$ es una sucesión de eventos independientes, entonces $\mathrm{P}(\limsup A_n)$ es *cero* o *uno*. Este es un caso especial de una clase de resultados llamados *leyes cero–uno*. A continuación veremos el resultado más prominente dentro de esta clase, la *ley cero–uno de Kolmogorov*. Esto requiere introducir el siguiente concepto.

Dada una sucesión de eventos $\{A_n\}$, para cada $n = 1, 2, \ldots$, sea $\sigma\{A_n, A_{n+1}, \ldots\}$ la σ–álgebra generada por $\{A_n, A_{n+1}, \ldots\}$. La σ–álgebra

$$\mathcal{C} := \bigcap_{n=1}^{\infty} \sigma\{A_n, A_{n+1}, \ldots\}$$

se llama la **$\boldsymbol{\sigma}$–álgebra cola** asociada a $\{A_n\}$. Los conjuntos en $\mathcal{C}$ se llaman **eventos cola**. Por ejemplo, $\limsup A_n$ y $\liminf A_n$ son eventos cola.

Teorema 3.12. ***(Ley cero–uno de Kolmogorov)***
Sea $\mathcal{C}$ la σ–álgebra cola asociada a una sucesión $\{A_n\}$. Si los eventos A_n son independientes y $A \in \mathcal{C}$, entonces

$$\mathrm{P}(A) = 0 \;\; \text{ó} \;\; 1.$$

Una posible demostración ocupa el siguiente lema, cuya demostración es el Ejercicio 4.17.

Lema 3.13. *si $A_1, A_2, \ldots, B_1, B_2, \ldots$ son eventos independientes, entonces las σ–álgebras $\sigma\{A_1, A_2, \ldots\}$ y $\sigma\{B_1, B_2, \ldots\}$ son independientes.*

Demostración. Idea 1 de la demostración de 3.12. Sea $A \in \mathcal{C}$ un evento cola. En particular, para cada $n = 1, 2, \ldots,$ A está en la $\sigma-$álgebra $\sigma\{A_n, A_{n+1}, \ldots\}$. Luego, por el Lema 3.13, A es *independiente* de $\sigma\{A_1, \ldots, A_{n-1}\}$. Como esto se cumple para todo $n \geq 1$, se sigue que A es independiente de $\{A_1, A_2, \ldots\}$. Por otra parte, $A \in \mathcal{C} \subset \sigma\{A_1, A_2, \ldots\}$ y por lo tanto, A *es independiente de sí mismo.* Esto implica que $\mathrm{P}(A \cap A) = \mathrm{P}(A)\mathrm{P}(A)$, i.e., $\mathrm{P}(A) = \mathrm{P}(A)^2$, así que $\mathrm{P}(A) = 0$ ó 1. $\square$

Demostración. Idea 2 de la demostración de 3.12. Sea $A \in \mathcal{C}$ un evento cola. En particular, para cada $n = 1, 2, \ldots,$ A está en la $\sigma-$álgebra $\sigma\{A_n, A_{n+1}, \ldots\}$. Luego, por el Lema 3.13, A es *independiente* de $\mathcal{A}_n := \sigma\{A_1, \ldots, A_{n-1}\}$. Entonces para todo $n \geq 1$ y $B_n \in \mathcal{A}_n$, $P(A \bigcap B_n) = P(A)P(B_n)$. Por otro lado, como A está en la $\sigma-$álgebra $\bigcup_{n \geq 1} \mathcal{A}_n$, entonces hay $C_n \in \mathcal{A}_n,\ n = 1, 2, \ldots$ tal que $A \subset C_n$ y $\lim_{n \to \infty} P(C_n) = P(A)$. Pero como ya mencionamos, $P(C_n \bigcap A) = P(C_n)P(A)$ para todo $n \geq 1$. Así

$$P(A) = P(C_n \bigcap A) = P(C_n)P(A) \to P(A)^2,\ n \to \infty.$$

Por lo tanto $P(A) = 0$ o 1.

$\square$

La ley 0-1 de Kolmogorov se usa mucho para variables aleatorias independientes, que veremos más adelante.

Ejercicios § 3

3.1. Demuestre la Proposición 3.1.

3.2. En el lanzamiento de dos dados honestos el espacio muestral es

$$\Omega = \{(i,j) \mid i,j = 1,\ldots,6\}.$$

Considere los eventos $A_1 = \{(i,j)|j = 1,2 \text{ ó } 5\}$, $A_2 = \{(i,j)|j = 4,5 \text{ ó } 6\}$ y $A_3 = \{(i,j)|i + j = 9\}$. Demuestre que

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3),$$

pero los eventos **no** son independientes porque $\mathrm{P}(A_i \cap A_j) \neq \mathrm{P}(A_i)\mathrm{P}(A_j)$ para $i \neq j$.

3.3. Demuestre que si A y B son eventos independientes, entonces

(a) A^c y B^c son independientes,

(b) A^c y B son independientes,

(c) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) \cdot \mathrm{P}(A^c)$.

3.4. Demuestre: si $A_1, \ldots, A_n$ son eventos independientes, entonces

$$\mathrm{P}(\bigcup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} \mathrm{P}(A_i^c).$$

Si además $\mathrm{P}(A_i) = p_i$ para $i = 1, \ldots, n$, entonces la probabilidad de que ninguno de tales eventos ocurra es $\prod_{i=1}^{n}(1 - p_i)$.

3.5. Demuestre:

(a) Si A es un evento independiente de sí mismo, entonces $\mathrm{P}(A) = 0$ ó $\mathrm{P}(A) = 1$.

(b) Si $\mathrm{P}(A) = 0$ ó $\mathrm{P}(A) = 1$, entonces A y cualquier otro evento son independientes.

3.6. Un estudiante toma un examen de opción múltiple en el que cada pregunta tiene 5 respuestas posibles. Si el estudiante conoce la respuesta correcta, la selecciona; en caso contrario selecciona al azar una de las 5 respuestas posibles. Suponga que el estudiante conoce la respuesta del 70% de las preguntas.

(a) ¿Cuál es la probabilidad de que de una pregunta dada el estudiante dé la respuesta correcta?

(b) Si el estudiante obtiene la respuesta correcta a una pregunta, ¿cuál es la probabilidad de que efectivamente conozca la respuesta?

3.7. Sean A, B y C tres eventos dados. Demuestre:

(a) Si los eventos son independientes y $\mathrm{P}(A \cap B) > 0$, entonces $\mathrm{P}(C|A \cap B) = \mathrm{P}(C)$.

(b) Si $\mathrm{P}(A \cap B \cap C) > 0$ y $\mathrm{P}(C|A \cap B) = \mathrm{P}(C|B)$, entonces $\mathrm{P}(A|B \cap C) = \mathrm{P}(A|B)$.

3.8. Por la Definición 3.7(b), n eventos $A_1, \ldots, A_n$ son independientes si la condición (5) se cumple para $k = 2, 3, \ldots, n$. Demuestre que el número total de condiciones de la forma (5) que se deben verificar para que $A_1, \ldots,$ A_n sean independientes es $2^n - n - 1$.

3.9. Demuestre que si A y B son ajenos, entonces A y B *no* pueden ser independientes a menos que $\mathrm{P}(A) = 0$ ó $\mathrm{P}(B) = 0$.

3.10. Suponga que hay una prueba para detectar cáncer con la propiedad de que 90% de los individuos con cáncer reaccionan positivamente, mientras que 5% de aquellos que *no* tienen cáncer reaccionan positivamente. Suponga que el 1% de los individuos en una cierta población tiene cáncer. Calcule la probabilidad de que en verdad tenga cáncer un paciente seleccionado al azar de dicha población y que reacciona positivamente a la prueba.

3.11. Muestre que $\limsup A_n$ está en la σ-álgebra cola. Sugerencia: muestre que

$$\bigcap_{n \geq 1} \bigcup_{k \geq n} A_n = \bigcap_{n \geq 2} \bigcup_{k \geq n} A_n.$$

4 Variables aleatorias

Contenido: Función medible, variable aleatoria (v.a.), función de distribución, medida de Lebesgue–Stieltjes, v.a. discreta, v.a. continua.

A partir de esta sección usaremos con mucha frecuencia los conceptos en el Ejercicio 1.4 sobre la *imagen inversa* de una función. Por tal motivo, se recomienda que el lector repase dicho ejercicio. En particular, recuérdese que si Ω y Ω' son conjuntos arbitrarios y f es una función de Ω en Ω', entonces para cualquier conjunto $B \subset \Omega'$ definimos la *imagen inversa* de B como

$$f^{-1}(B) := \{\omega \in \Omega \mid f(\omega) \in B\}. \tag{4.1}$$

Si $\Omega' = \mathbb{R}$ y B es el intervalo $(-\infty, x]$, escribimos (4.1) como

$$f^{-1}(-\infty, x] = \{\omega \in \Omega \mid f(\omega) \leq x\} \equiv \{f \leq x\}.$$

Definición 4.1. Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida. Se dice que $\mathrm{X} : \Omega \to \mathbb{R}$ es una función $\mathcal{F}$–**medible** (o medible con respecto a $\mathcal{F}$) si

$$\mathrm{X}^{-1}(-\infty, x] = \{\omega \in \Omega \mid \mathrm{X}(\omega) \leq x\} \in \mathcal{F} \quad \forall\, x \in \mathbb{R}. \tag{4.2}$$

(Nótese que si $\mathrm{X}^{-1}(-\infty, x]$ está en $\mathcal{F}$, entonces su medida $\mu(\mathrm{X}^{-1}(-\infty, x])$ está "bien definida".) Casos especiales:

(a) Si $\mu = \mathrm{P}$ es una m.p., decimos que X es una **variable aleatoria** (abreviado: v.a.)

(b) Si $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y $\mathrm{X} = \mathbb{R} \to \mathbb{R}$ es $\mathcal{B}(\mathbb{R})$–medible, decimos que X es una **función de Borel** o **Borel–medible**.

Es **importante** notar lo siguiente: para verificar que X es $\mathcal{F}$–medible, en (4.2) podemos sustituir $(-\infty, x]$ por cualquier otro intervalo en $\mathbb{R}$, es decir, de la forma $(-\infty, x)$ ó (x, ∞) ó $[x, \infty)$ ó $[x, y)$ ó $(x, y]$ ó (x, y) ó $[x, y]$. Asimismo, podemos sustituir $(-\infty, x]$ por cualquier conjunto *abierto* $B \subset \mathbb{R}$ o cualquier *cerrado* $B \subset \mathbb{R}$. En particular, por razones teóricas es conveniente enfatizar que

$$\mathrm{X} \text{ \textit{es} } \mathcal{F}\text{\textit{–medible} ssi } \mathrm{X}^{-1}(B) \in \mathcal{F} \ \forall \text{ \textit{abierto} } B \subset \mathbb{R}. \tag{4.3}$$

Ejemplo 4.2. Demuestre que las siguientes proposiciones son equivalentes:

(a) X es $\mathcal{F}$–medible (en el sentido de (4.2)).

(b) $\mathrm{X}^{-1}(x,\infty) \in \mathcal{F} \quad \forall\, x \in \mathbb{R}$.

(c) $\mathrm{X}^{-1}(x,y) \in \mathcal{F}$ para cualquier intervalo abierto (x,y) en $\mathbb{R}$.

(d) $\mathrm{X}^{-1}(B) \in \mathcal{F}$ para cualquier conjunto abierto $B \subset \mathbb{R}$.

Demostración. (a) $\Leftrightarrow$ (b). Como $(-\infty,x]^c = (x,\infty)$, por el Ejercicio 1.4(b) tenemos que

$$(\mathrm{X}^{-1}(-\infty,x])^c = \mathrm{X}^{-1}((-\infty,x]^c) = \mathrm{X}^{-1}(x,\infty).$$

Por lo tanto, para cualquier $x \in \mathbb{R}$, si $\mathrm{X}^{-1}(-\infty,x]$ está en $\mathcal{F}$, entonces también $\mathrm{X}^{-1}(x,\infty)$ está en $\mathcal{F}$. Es decir, (a) $\Rightarrow$ (b). En forma análoga se obtiene el recíproco, (b) $\Rightarrow$ (a).

(a) $\Rightarrow$ (c). Primero observe que $(x,y) = (x,\infty) \cap (-\infty,y)$ y que, por el Ejercicio 1.4(c),

$$\mathrm{X}^{-1}(x,y) = \mathrm{X}^{-1}((x,\infty) \cap (-\infty,y)) = \mathrm{X}^{-1}(x,\infty) \cap \mathrm{X}^{-1}(-\infty,y). \tag{4.4}$$

Por el párrafo anterior, (a) implica que $\mathrm{X}^{-1}(x,\infty)$ está en $\mathcal{F}$. Por lo tanto, en vista de (4.4), para demostrar (c) basta ver que $\mathrm{X}^{-1}(-\infty,y)$ está en $\mathcal{F}$, lo cual se obtiene notando que

$$(-\infty,y) = \bigcup_{n=1}^{\infty} (-\infty, y-1/n]$$

y que, por (a) y el Ejercicio 1.4(c),

$$\mathrm{X}^{-1}(-\infty,y) = \bigcup_{n=1}^{\infty} \mathrm{X}^{-1}(-\infty, y-1/n] \in \mathcal{F}.$$

La demostración de (c) $\Rightarrow$ (a) es similar.

(c) $\Leftrightarrow$ (d). Es evidente que (d) $\Rightarrow$ (c) porque (x,y) es un conjunto abierto. Para probar el recíproco, (c) $\Rightarrow$ (d), recuérde que si $B \subset \mathbb{R}$ es un conjunto abierto, entonces existe una colección numerable de intervalos abiertos I_n tales que $B = \bigcup_n \mathrm{I}_n$. Por lo tanto, si (c) se cumple vemos que $\mathrm{X}^{-1}(B) = \bigcup_n \mathrm{X}^{-1}(\mathrm{I}_n)$ está en $\mathcal{F}$. $\square$

Ejemplo 4.3. Si $(\Omega,\mathcal{F})$ es un espacio discreto, entonces *cualquier función* $\mathrm{X} : \Omega \to \mathbb{R}$ *es una v.a.* porque el conjunto potencia $\mathcal{F} = 2^{\Omega}$ consiste de *todos* los posibles subconjuntos de Ω; luego (2) se cumple trivialmente.

Para ser más específicos, considérese el lanzamiento de dos dados. En tal caso el espacio muestral Ω consiste de todas las parejas $\omega = (i, j)$ de números enteros i, j entre 1 y 6. Algunos ejemplos de vv.aa. son:

$\mathrm{X}(i, j) := i + j =$ suma de los resultados de ambos dados;

$Y(i, j) := i =$ resultado del primer dado;

$Z(i, j) := \min(i, j) =$ mínimo de los resultados de ambos dados. $\square$

La probabilidad del evento $\mathrm{X}^{-1}(-\infty, x]$ en (4.2), para todo $x \in \mathbb{R}$, se llama la *función de distribución* de la v.a. X. Para introducir formalmente este concepto usaremos la siguiente **notación:**

$$\{\mathrm{X} \leq x\} := \mathrm{X}^{-1}(-\infty, x] = \{\omega \in \Omega \mid \mathrm{X}(\omega) \leq x\} \quad \text{para} \quad x \in \mathbb{R}. \tag{4.5}$$

Asimismo, para cualquier conjunto $B \subset \mathbb{R}$ escribimos

$$\{\mathrm{X} \in B\} := \mathrm{X}^{-1}(B) = \{\omega \in \Omega \mid \mathrm{X}(\omega) \in B\}.$$

Definición 4.4. Sea X una v.a. La **función de distribución** (abreviado: f.d.) de X es la función $F_{\mathrm{X}} : \mathbb{R} \to [0, 1]$ dada por

$$F_{\mathrm{X}}(x) := \mathrm{P}\{\mathrm{X} \leq x\} \quad \forall\, x \in \mathbb{R}. \tag{4.6}$$

La **m.p. inducida por** X es la m.p. sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definida como

$$\mathrm{P}_{\mathrm{X}}(B) := \mathrm{P}\{\mathrm{X} \in B\} \quad \forall\, B \in \mathcal{B}(\mathbb{R}). \tag{4.7}$$

Nótese que $\mathrm{P}_{\mathrm{X}}(B) = \mathrm{P}[\mathrm{X}^{-1}(B)]$. Además, la f.d. de X y la m.p. inducida por X están relacionadas como

$$F_{\mathrm{X}}(x) = \mathrm{P}_{\mathrm{X}}(-\infty, x]. \tag{4.8}$$

Ejemplo 4.5. (a) Se dice que $\mathrm{X} := \Omega \to \mathbb{R}$ es una v.a. **constante** si existe $c \in \mathbb{R}$ tal que $\mathrm{X}(\omega) = c$ para todo $\omega \in \Omega$. Es claro que tal función es una v.a. porque

$$\begin{aligned} \{\mathrm{X} \leq x\} &= \phi \quad \text{si} \quad x < c, \\ &= \Omega \quad \text{si} \quad x \geq c. \end{aligned}$$

De aquí se sigue que la f.d. de $\mathrm{X} = c$ es la función escalonada

$$\begin{aligned} F_{\mathrm{X}}(x) &= 0 \quad \text{si} \quad x < c, \\ &= 1 \quad \text{si} \quad x \geq c. \end{aligned}$$

(b) Si X es una v.a. que toma únicamente dos valores, digamos 0 y 1, se dice que X es una v.a. **Bernoulli**. Si $\mathrm{P}\{\mathrm{X} = 1\} = p$, entonces $\mathrm{P}\{\mathrm{X} = 0\} = 1 - \mathrm{P}\{\mathrm{X} = 1\} = 1 - p$ y la f.d. de X es la función escalonada

$$\begin{aligned} F_{\mathrm{X}}(x) &= 0 \quad \text{si} \quad x < 0, \\ &= 1 - p \quad \text{si} \quad 0 \leq x < 1, \\ &= 1 \quad \text{si} \quad x \geq 1. \end{aligned}$$

Un caso particular de v.a. Bernoulli es la **función indicadora** I_A de un evento $A \in \mathcal{F}$, definida en (2.3), i.e.

$$\begin{aligned} \mathrm{I}_A(\omega) &:= 1 \quad \text{si} \quad \omega \in A, \\ &:= 0 \quad \text{si} \quad \omega \notin A. \end{aligned}$$

Nótese que

$$\begin{aligned} \{\mathrm{I}_A \leq x\} &= \phi \quad \text{si} \quad x < 0, \\ &= A^c \quad \text{si} \quad 0 \leq x < 1, \\ &= \Omega \quad \text{si} \quad x \geq 1. \end{aligned}$$

Además, $\mathrm{P}\{\mathrm{I}_A = 1\} = \mathrm{P}(A)$ y $\mathrm{P}\{\mathrm{I}_A = 0\} = 1 - \mathrm{P}\{\mathrm{I}_A = 1\} = \mathrm{P}(A^c)$. □

Proposición 4.6. (**Propiedades de** F_{X}) Si X es una v.a., su f.d. F_{X} satisface que

(a) F_{X} es *no–decreciente*, es decir si $x < y$ entonces $F_{\mathrm{X}}(x) \leq F_{\mathrm{X}}(y)$;

(b) $F_{\mathrm{X}}(+\infty) := \lim_{x \to \infty} F_{\mathrm{X}}(x) = 1$ y $F_{\mathrm{X}}(-\infty) := \lim_{x \to -\infty} F_{\mathrm{X}}(x) = 0$;

(c) F_{X} es *continua por la derecha*, es decir, si $F_{\mathrm{X}}(x+) := \lim_{y \downarrow x} F_{\mathrm{X}}(y)$ es el límite de F_{X} en el punto x por la derecha, entonces $F_{\mathrm{X}}(x+) = F_{\mathrm{X}}(x)$ para todo $x \in \mathrm{X}$.

Demostración. (a) Si $x < y$, entonces $\{X \leq x\} \subset \{X \leq y\}$. Por lo tanto (a) se sigue de la propiedad de monotonía 1.9(c).

(b) Puesto que $\{X \leq n\} \uparrow \Omega$ cuando $n \to \infty$, la propiedad de continuidad 1.11(a) da que $F_X(n) = P\{X \leq n\} \to 1$. Análogamente, por 1.12(b), $F_X(-n) = P\{X \leq -n\} \to 0$ pues $\{X \leq -n\} \downarrow \phi$.

(c) Como $\{x < X \leq x + y\} \downarrow \phi$ cuando $y \downarrow 0$,

$$F_X(x + y) - F_X(x) = P\{x < X \leq x + y\} \downarrow 0. \quad \square$$

$\square$

Definición 4.7. Cualquier función $F : \mathbb{R} \to [0, 1]$ que satisface las propiedades (a), (b) y (c) de la Proposición 4.6 se dice que es una **función de distribución de probabilidad** (abreviado: f.d.p.)

Observación 4.8. (a) Por la Proposición 4.6 la f.d. de una v.a. X es una f.d.p. El recíproco también es cierto, en el siguiente sentido: Si $F : \mathbb{R} \to [0, 1]$ es una f.d.p., entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y una v.a. X sobre Ω cuya f.d. es F, es decir, $F_X = F$. (Vea la Proposición 4.10.)

(b) Si X y Y son vv.aa. y $X = Y$, entonces es claro que $F_X = F_Y$. Sin embargo, el recíproco es **falso**, i.e.

$$F_X = F_Y \not\Rightarrow X = Y.$$

De hecho, puede ocurrir que $F_X = F_Y$ aunque las vv.aa. X, Y ni siquiera estén definidas en el mismo espacio de probabilidad. (Explique.) $\square$

Definición 4.9. Sea μ una medida sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Decimos que μ es una **medida de Lebesgue–Stieltjes** (abreviado: medida de LS) si $0 \leq \mu(I) < \infty$ para cualquier intervalo $I \subset \mathbb{R}$ acotado.

Por ejemplo, la medida de Lebesgue λ y una m.p. μ sobre $\mathbb{R}$ son medidas de LS. En particular, la m.p. P_X inducida por una v.a. X es de LS (ver (4.7)).

Si μ es una m.p. sobre $\mathbb{R}$, entonces la función $F : \mathbb{R} \to \mathbb{R}$ definida como

$$F(x) := \mu(-\infty, x] \quad \forall\, x \in \mathbb{R} \tag{4.9}$$

es una f.d.p. (Compare (4.9) con (4.8).) Recíprocamente, *si F es una f.d.p., entonces existe una única medida de LS μ_F*. De hecho, esta m.p. se caracteriza por la propiedad de que

$$\mu_F(a, b] = F(b) - F(a) \quad \forall \text{ intervalo } (a, b] \subset \mathbb{R}. \tag{4.10}$$

(La demostración de este resultado es muy parecida a la construcción de la medida de Lebesgue, basándose en el Teorema de extensión de Carathéodory 1.20 y en el Ejemplo 1.19 con la fórmula (4.10) en lugar de la $\ell(a,b] := b-a$. Para más detalles vea, por ejemplo, el libro de Ash (1972), Teorema 1.4.4.) A μ_F se le llama la **m.p. inducida por** F.

Proposición 4.10. Si F es una f.d.p., entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$ y una v.a. X sobre Ω cuya f.d. F_{X} coincide con F, es decir, $F_{\mathrm{X}}(x) = F(x)$ para todo $x \in \mathrm{X}$.

Demostración. Sea $(\Omega, \mathcal{F}) := (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y sea P la medida de LS definida (o inducida) por F sobre $\mathcal{B}(\mathbb{R})$, es decir, como en (10). Finalmente, sea X la v.a. sobre $\Omega = \mathbb{R}$ definida como $\mathrm{X}(\omega) := \omega$ para todo $\omega \in \mathbb{R}$. Entonces $F_{\mathrm{X}}(x) := \mathrm{P}\{\mathrm{X} \leq x\} = F(x)$ para todo $x \in \mathbb{R}$. $\square$

Definición 4.11. Se dice que una v.a. X es **discreta** si existe un conjunto finito o infinito numerable $S \subset \mathbb{R}$ tal que X toma valores en S únicamente.

Las vv.aa. en el Ejemplo 4.5 son discretas. En el siguiente ejemplo se presentan otras vv.aa. discretas.

Ejemplo 4.12. (a) Como continuación del Ejemplo 3.9, suponga que se realizan n repeticiones independientes de un experimento que tiene probabilidad p $(0 < p < 1)$ de "éxito" y $q := 1-p$ de "fracaso". Sea X := número de éxitos. Entonces X es una v.a. discreta con valores en el conjunto $S = \{0, 1, \ldots, n\}$ y, además,

$$\mathrm{P}\{\mathrm{X} = k\} = \binom{n}{k} p^k q^{n-k} \quad \forall\, k = 0, 1, \ldots, n.$$

En este caso se dice que X tiene **distribución binomial** con parámetros n y p, y en símbolos escribimos $\mathrm{X} \sim \mathrm{Bin}(n, p)$.

(b) (Recuérdese la densidad geométrica en el Ejemplo 2.2(c).) Supóngase que el experimento en el inciso (a) se repite hasta que ocurre el primer éxito. Sea Y la v.a. que cuenta el número de repeticiones que ocurren *antes* del primer éxito. En particular, $Y = 0$ ssi ocurre éxito en la primera realización del experimento. En general, Y toma valores en el conjunto $S = \{0, 1, \ldots\}$ y, por independencia,

$$\mathrm{P}\{Y = k\} = q^k \cdot p \quad \forall\, k = 0, 1, \ldots.$$

En este caso se dice que Y tiene **distribución geométrica** con parámetro p, y escribimos $Y \sim \mathrm{Geo}(p)$.

Definición 4.13. Se dice que una v.a. X es **absolutamente continua** si existe una función de Borel $f : \mathbb{R} \to \mathbb{R}$, no–negativa, y tal que

$$F_{\mathrm{X}}(x) = \int_{-\infty}^{x} f(y)dy \quad \forall\, x \in \mathbb{R}. \tag{4.11}$$

En este caso se dice que f es la **densidad de probabilidad** (o simplemente la **densidad**) de X.

En el Ejemplo 2.11 vimos algunos ejemplos de densidades continuas. En algunos de tales casos se utiliza una nomenclatura especial. Por ejemplo, si X tiene la **densidad uniforme** en $[a, b]$, simbólicamente escribimos $\mathrm{X} \sim \mathrm{Uni}[a, b]$. Asimismo, para el caso de la **densidad exponencial** con parámetro λ escribimos $\mathrm{X} \sim \mathrm{Exp}(\lambda)$, y para la **densidad normal** con parámetros m y σ^2 escribimos $\mathrm{X} \sim N(m, \sigma^2)$. En particular, $\mathrm{X} \sim N(0, 1)$ significa que X tiene **densidad normal estándar**.

Observación 4.14. Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida. Decimos que una cierta propiedad $\mathcal{P}$ se cumple μ–**casi donde quiera** (abreviado: μ–c.d.q.) si $\mathcal{P}$ se cumple en todo Ω excepto en un conjunto de medida μ igual a cero. En otras palabras, existe un conjunto $A \in \mathcal{F}$ tal que

(a) $\mu(A) = 0$, y

(b) $\mathcal{P}$ se cumple para todo $x \notin A$.

Por ejemplo, sea $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ y supóngase que F_{X} satisface (4.11). Entonces se puede demostrar que la derivada F'_{X} existe y coincide con la densidad f λ–c.d.q. Para ser más concretos, supóngase, por ejemplo, que $\mathrm{X} \sim \mathrm{Uni}[a, b]$ de modo que X tiene la densidad

$$f(x) := \frac{1}{b-a} \quad \text{si} \quad x \in [a, b]$$

y $f(x) := 0$ para $x \notin [a, b]$. Entonces, calculando la integral en (4.11) vemos que

$$\begin{aligned} F_{\mathrm{X}}(x) = \int_{-\infty}^{x} f(y)dy &= 0 \quad \text{si} \quad x < a, \\ &= \frac{x-a}{b-a} \quad \text{si} \quad a \le x < b, \\ &= 1 \quad \text{si} \quad x \ge b. \end{aligned} \tag{4.12}$$

Entonces se tiene que la derivada $F'_X(x) = f(x)$ para todo $x \in \mathbb{R}$ *excepto* en el conjunto $A = \{a, b\}$, el cual tiene medida de Lebesgue $\lambda(A) = 0$. (Del Ejemplo 1.8(d), recuerde que $\lambda(B) = 0$ si B es un conjunto numerable.)

Otro ejemplo: si $X \sim \text{Exp}(\lambda)$, entonces

$$\begin{aligned} F_X(x) &= 1 - e^{-\lambda x} \quad \text{si} \quad x \geq 0, \\ &= 0 \quad \text{si} \quad x < 0, \end{aligned}$$

de modo que la derivada $F'_X(x) = f(x) =$ la densidad exponencial para todo $x \in \mathbb{R}$, *excepto* en $x = 0$. Es decir, $F'_X = f$ λ–c.d.q.

Definición 4.15. Sea F una f.d.p. y μ_F la medida de LS inducida por F. Se dice que F es **continua singular** si

(a) F es continua y

(b) existe un conjunto de Borel $S \subset \mathbb{R}$ que tiene medida de Lebesgue $\lambda(S) = 0$, pero $\mu_F(S) = 1$.

En otras palabras, la condición (b) dice que μ_F *está concentrada* en un conjunto nulo con respecto a λ.

Ejemplo 4.16. (**La distribución de Cantor.**)
Sea $C_1 := [0, 1/3] \cup [2/3, 1]$. Sea f_1 la función de densidad uniforme sobre C_1 y F_1 su f.d., es decir,

$$f_1(x) = \begin{cases} c_1 & \text{si } x \in C_1, \\ 0 & \text{en c.c.,} \end{cases}$$

con $c_1 = 3/2$, y

$$F_1(x) := \int_{-\infty}^{x} f_1(y)dy = \begin{cases} 0 & \text{si } x < 0, \\ (3/2)x & \text{si } 0 \leq x < 1/3, \\ 1/2 & \text{si } 1/3 \leq x < 2/3, \\ (3/2)x - 1/2 & \text{si } 2/3 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Análogamente, sean $C_2 := [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, f_1 la densidad uniforme sobre C_2, y F_2 su correspondiente f.d.p.

En general, para $n \geq 2$, sea C_n la unión de los 2^n intervalos ajenos de longitud $(1/3)^n$ que se obtienen al eliminar los "tercios medios" de los 2^{n-1} intervalos ajenos de longitud $(1/3)^{n-1}$ en C_{n-1}.

Sea

$$f_n(x) := \begin{cases} c_n & \text{si } x \in C_n, \\ 0 & \text{en c.c.}, \end{cases}$$

con $c_n = (3/2)^n$, y

$$F_n(x) := \int_{-\infty}^{x} f_n(y)dy.$$

Las f.d.p. F_n son continuas y convergen uniformemente porque, si $m < n$, entonces

$$|F_m(x) - F_n(x)| \leq \left(\frac{1}{2}\right)^m \quad \forall\, x.$$

De aqui se sigue que la sucesion F_n converge porque recuérdese que el espacio de funciones continuas es completo con la norma del supremo. Es decir, existe una f.d.p. F continua, tal que $F_n(x) \to F(x)$ para todo $x \in \mathbb{R}$. Además, si $C := \bigcap_{n=1}^{\infty} C_n$ es el conjunto Cantor, es claro que la medida μ_F del complemento de C es cero, de modo que $\mu_F(C) = 1$. Por lo tanto, F es continua singular porque $\lambda(C) = 0$. $\square$

Concluimos esta sección con varias proposiciones elementales. Vea también el Ejercicio 4.8.

Proposición 4.17. Si X y Y son vv.aa. y $a \in \mathbb{R}$, entonces $a\mathrm{X}$, X^2, $\mathrm{X} + Y$, $\mathrm{X}Y$, $\max(\mathrm{X}, Y), \min(\mathrm{X}, Y)$ y $|\mathrm{X}|$ son vv.aa.

Demostración. Las primeras dos son sencillas. La tercera surge de

$$\begin{aligned} \{X + Y < z\} &= \{X < z - Y\} = \bigcup_{\alpha \in \mathbb{R}} \{X < \alpha < z - Y\} \\ &= \bigcup_{r \in \mathbb{Q}} \{X < r < z - Y\} = \bigcup_{r \in \mathbb{Q}} \{X < r\} \bigcap \{r < z - Y\} \in \mathcal{F}. \end{aligned}$$

Para el producto sabemos que $XY = \frac{1}{2}(X+Y)^2 - \frac{1}{2}(X-Y)^2$. Para el máximo se tiene que

$$\{\max(X, Y) < z\} = \{X < z\} \bigcap \{Y < z\}.$$

Para el mínimo se usa que $\min(X, Y) = -\max(-X, -Y)$. La última se tienes pues $|X| = \max(X, 0) - \min(0, X)$. $\square$

Observación 4.18. Sea V la familia de todas las vv.aa. definidas sobre un espacio de probabilidad dado. Entonces de 4.17 vemos que V es un **espacio vectorial**, es decir $a\mathrm{X}$ y $\mathrm{X}+Y$ están en V para todo $a \in \mathbb{R}$ y X, Y en V.

Proposición 4.19. X es una v.a. discreta ssi existe una sucesión $\{x_1, x_2, \ldots\}$ en $\mathbb{R}$ y una sucesión $\{A_1, A_2, \ldots\}$ de eventos que forman una partición de Ω tales que

$$\mathrm{X} = \sum_k x_k \, \mathrm{I}_{A_k}. \tag{4.13}$$

En particular, X es una v.a. discreta con un número **finito** de valores ssi X se puede expresar como una combinación lineal (finita) de funciones indicadoras de conjuntos ajenos.

Observación 4.20. En Análisis Real, una función medible que toma sólo un número finito de valores (como la v.a. discreta en (4.13)) se dice que es una función **simple**.

Proposición 4.21. Si X es una v.a. no negativa, entonces existe una sucesión $\{\mathrm{X}_n\}$ de vv.aa. discretas con un número finito de valores tales que X_n converge puntualmente a X.

La sucesión $\{\mathrm{X}_n\}$ en 4.21 se puede construir explícitamente. Véase, por ejemplo, el libro de Ash (1972), Teorema 1.5.5, o el libro de R.G. Bartle (1995), *The Elements of Integration and Lebesgue Measure*, Lema 2.11.

Funciones medibles: caso general

Sean $(\Omega, \mathcal{F})$ y $(\Omega', \mathcal{F}')$ dos espacios medibles. Se dice que una función $\mathrm{X} : \Omega \to \Omega'$ es **medible** con respecto a $\mathcal{F}$ y $\mathcal{F}'$ si

$$\mathrm{X}^{-1}(B) \in \mathcal{F} \quad \forall \, B \in \mathcal{F}'. \tag{4.14}$$

Equivalentemente, usando la notación introducida en el Ejercicio 1.4, la función X es medible ssi

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

La siguiente proposición da un criterio muy útil para verificar medibilidad.

Proposición 4.22. Sea $\mathcal{C}$ una familia de subconjuntos de Ω' que genera a $\mathcal{F}'$, es decir, $\sigma\{\mathcal{C}\} = \mathcal{F}'$. Entonces $\mathrm{X} : \Omega \to \Omega'$ es medible con respecto a $\mathcal{F}$ y $\mathcal{F}'$ ssi $\mathrm{X}^{-1}\{\mathcal{C}\} \subset \mathcal{F}$.

Demostración. La necesidad es obvia. Para demostrar la suficiencia, supóngase que

$$\mathrm{X}^{-1}(C) \in \mathcal{F} \quad \forall\, C \in \mathcal{C}. \tag{4.15}$$

Deseamos demostrar (4.14). Con este fin considere $\mathcal{D} := \{B \in \mathcal{F}' | \mathrm{X}^{-1}(B) \in \mathcal{F}\}$. Entonces $\mathcal{C} \subset \mathcal{D}$ y como $\mathcal{D}$ es una σ–álgebra

$$\sigma\{\mathcal{C}\} \subset \mathcal{D} \subset \mathcal{F}'.$$

Finalmente, como $\sigma\{\mathcal{C}\} = \mathcal{F}'$, concluimos que $\mathcal{F}' = \mathcal{D}$ y por lo tanto $\mathrm{X}^{-1}(\mathcal{F}') = \mathrm{X}^{-1}(\mathcal{D}')$. $\square$

Tomando $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y recordando que la familia $\mathcal{C}$ de todos los intervalos $(-\infty, x]$, con $x \in \mathbb{R}$, genera la σ–álgebra de Borel $\mathcal{B}(\mathbb{R})$, vemos que la condición (2) es un caso especial de (4.15).

Ejercicios § 4

4.1. (Compare con el Ejemplo 4.2.) Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida y $\mathrm{X} : \Omega \to \mathbb{R}$ una función dada. Demuestre que las siguientes proposiciones son equivalentes:

(a) X es $\mathcal{F}$–medible.

(b) $\mathrm{X}^{-1}[x, y] \in \mathcal{F}$ para cualquier intervalo cerrado $[x, y]$ en $\mathbb{R}$.

(c) $\mathrm{X}^{-1}(C) \in \mathcal{F}$ para cualquier conjunto cerrado C en $\mathbb{R}$.

4.2. Sea X una v.a. sobre $(\Omega, \mathcal{F}, \mathrm{P})$. Demuestre que la función $\mathrm{P_X}$ definida en (4.7) es efectivamente una m.p. sobre $\mathbb{R}$. Nótese que, más explícitamente, podemos escribir (4.7) como

$$\mathrm{P_X}(B) := \mathrm{P}[\mathrm{X}^{-1}(B)] \quad \forall\, B \in \mathcal{B}(\mathbb{R}).$$

4.3. Sea F_{X} la f.d. de una v.a. X y sea $F_{\mathrm{X}}(x-)$ el límite de F_{X} en el punto x **por la izquierda**, i.e.

$$F_{\mathrm{X}}(x-) := \lim_{y \uparrow x} F_{\mathrm{X}}(y)$$

Demuestre que F_{X} satisface:

(a) $\mathrm{P}\{\mathrm{X} > x\} = 1 - F_{\mathrm{X}}(x)$,

(b) $\mathrm{P}\{\mathrm{X} < x\} = F_{\mathrm{X}}(x-)$,

(c) $\mathrm{P}\{y < \mathrm{X} \leq x\} = F_{\mathrm{X}}(x) - F_{\mathrm{X}}(y)$,

(d) $\mathrm{P}\{y \leq \mathrm{X} \leq x\} = F_{\mathrm{X}}(x) - F_{\mathrm{X}}(y-)$,

(e) $\mathrm{P}\{y < \mathrm{X} < x\} = F_{\mathrm{X}}(x-) - F_{\mathrm{X}}(y)$,

(f) $\mathrm{P}\{y \leq \mathrm{X} < x\} = F_{\mathrm{X}}(x-) - F_{\mathrm{X}}(y-)$,

(g) $\mathrm{P}\{\mathrm{X} = x\} = F_{\mathrm{X}}(x) - F_{\mathrm{X}}(x-)$; por lo tanto, F_{X} es continua en x ssi $\mathrm{P}\{\mathrm{X} = x\} = 0$.

Nota 4.23. Observe que si F_{X} es continua en x, entonces $F_{\mathrm{X}}(x-) = F_{\mathrm{X}}(x+) = F_{\mathrm{X}}(x)$ con $F_{\mathrm{X}}(x+)$ como en la Proposición 4.6(c). Por lo tanto, si F_{X} es continua en x, en (b) se tiene $\mathrm{P}\{\mathrm{X} < x\} = F_{\mathrm{X}}(x)$. Asimismo, si F_{X} es continua en x y y, entonces las probabilidades en (d)–(f) son iguales a la probabilidad en (c).

4.4. Sea X una v.a. y sean a, b números reales con $a > 0$. Considere la v.a. $Y := a\mathrm{X} + b$. Demuestre que $F_Y(x) = F_{\mathrm{X}}(\frac{x-b}{a})$ para todo $x \in \mathrm{X}$. Calcule F_Y cuando $a < 0$.

4.5. Sea $\mathrm{X} : \Omega \to \mathbb{R}$ una v.a. y $h : \mathbb{R} \to \mathbb{R}$ una función de Borel. Demuestre que la **composición** $h \circ \mathrm{X} : \Omega \to \mathbb{R}$ es una v.a. (Recuerde que $h \circ \mathrm{X}(\omega) := h(\mathrm{X}(\omega))$.)

Nota 4.24. (a) Frecuentemente la composición $h \circ \mathrm{X}$ se escribe como $h(\mathrm{X})$. (b) Es fácil ver que, por ejemplo, una función continua $h : \mathbb{R} \to \mathbb{R}$ es de Borel. De hecho, prácticamente todas las funciones "usuales" (e.g. las funciones que se estudian en cursos de cálculo) son de Borel. En este curso sólo estudiaremos funciones $h(\mathrm{X})$ que son vv.aa.

4.6. Sea X una v.a. sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$. Del Ejercicio 1.4(d) deduzca que, con $\mathcal{B} \equiv \mathcal{B}(\mathbb{R})$,

$$\mathrm{X}^{-1}(\mathcal{B}) := \{\mathrm{X}^{-1}(B) \mid B \in \mathcal{B}\}$$

es una σ–álgebra de Ω. Además, demuestre que $\mathrm{X}^{-1}(\mathcal{B})$ es una **sub–$\boldsymbol{\sigma}$–álgebra** de $\mathcal{F}$, es decir, $\mathrm{X}^{-1}(\mathcal{B}) \subset \mathcal{F}$.

Nota 4.25. En probabilidad, a $\mathrm{X}^{-1}(\mathcal{B})$ usualmente se le denota por $\sigma\{\mathrm{X}\}$ y se le llama la σ–**álgebra generada** (o **inducida)** por X. Es claro que si $\mathcal{G}$ es *cualquier* σ–álgebra de Ω con respecto a la cual X es medible (i.e. $\mathrm{X}^{-1}(B) \in \mathcal{G}$ para todo $B \in \mathcal{B}(\mathbb{R})$), entonces $\mathcal{G}$ contiene a $\mathrm{X}^{-1}(\mathcal{B}) \equiv \sigma\{\mathrm{X}\}$. Por tal motivo también se dice que $\sigma\{\mathrm{X}\}$ es la **mínima** σ–álgebra de Ω con respecto a la cual X es medible.

4.7. Sean X y Y vv.aa. y $h : \mathbb{R} \to \mathbb{R}$ una función de Borel. Demuestre que si $Y = h(\mathrm{X})$, entonces $\sigma\{Y\} \subset \sigma\{\mathrm{X}\}$.

4.8. Sea $\{\mathrm{X}_n\}$ una sucesión de funciones $\mathcal{F}$–medibles (Definición 4.1). Demuestre:

(a) $\sup \mathrm{X}_n, \inf \mathrm{X}_n, \limsup \mathrm{X}_n$ y $\liminf \mathrm{X}_n$ son funciones $\mathcal{F}$–medibles.

(b) Si X_n converge puntualmente a X, entonces X es $\mathcal{F}$–medible.

4.9. Demuestre que una **combinación convexa** de f.d.p.'s es una f.d.p. Es decir, si $F_1, \ldots, F_n$ son f.d.p.'s y $\alpha_1, \ldots, \alpha_n$ son números no negativos tales que $\alpha_1 + \cdots + \alpha_n = 1$, entonces $\alpha_1 F_1 + \cdots + \alpha_n F_n$ es una f.d.p.

4.10. Sea X $\sim$Uni$[0,1]$. (La f.d. de X está dada por (12) con $a = 0$ y $b = 1$.) Sea G una f.d.p. que es continua y estrictamente creciente, y considere la v.a. $Y := G^{-1}(\mathrm{X})$. Demuestre que la f.d. de Y es $F_Y = G$.

4.11. Sea X una v.a. continua con densidad

$$f(x) := \frac{1}{2}\, e^{-|x|} \quad \forall\, x \in \mathbb{R}.$$

Calcule $\mathrm{P}(\mathrm{X} \geq 0)$ y $\mathrm{P}(|\mathrm{X}| \leq 2)$.

4.12. Supóngase que X $\sim$Exp(λ) y sea $c > 0$ una constante dada. Calcule la densidad de $Y := c\mathrm{X}$.

4.13. Sea X una v.a. con f.d.

$$\begin{aligned} F_{\mathrm{X}}(x) &= 0 && \text{si } x < 0, \\ &= x/3 && \text{si } 0 \leq x < 1, \\ &= x/2 && \text{si } 1 \leq x < 2, \\ &= 1 && \text{si } x \geq 2. \end{aligned}$$

Calcule $\mathrm{P}(1/2 \leq \mathrm{X} \leq 3/2)$, $\mathrm{P}(1/2 \leq \mathrm{X} \leq 1)$, $\mathrm{P}(\mathrm{X} > 1)$, $\mathrm{P}(\mathrm{X} \geq 1)$ y $\mathrm{P}(1 \leq \mathrm{X} \leq 3/2)$. (*Sugerencia*: use el Ejercicio 17.3.)

4.14. Considere la **función gama** (definida en el Ejercicio 2.6)

$$\Gamma(p) := \int_0^\infty x^{p-1}\, e^{-x}\, dx \quad \text{para } p > 0. \tag{$*$}$$

Demuestre que: (a) $\Gamma(1/2) = \sqrt{\pi}$. (*Sugerencia*: en (*) haga el cambio de variable $x = y^2/2$; después use la expresión (2.11).)

(b) $\Gamma(p+1) = p \cdot \Gamma(p) \quad \forall\, p > 0$.

(c) $\Gamma(n+1) = n!$ si n es un entero positivo.

(d) $\Gamma(n/2) = (n/2-1)\cdot\Gamma(n/2-1)$ para cualquier entero $n \geq 3$. En particular, del inciso (a),

$$\Gamma(3/2) = \frac{1}{2}\sqrt{\pi},\ \Gamma(5/2) = \frac{3}{4}\sqrt{\pi}, \ldots$$

4.15. Demuestre que si X tiene distribución normal estándar $N(0, 1)$, entonces X^2 tiene densidad gama con parámetros $p = \lambda = 1/2$. (*Sugerencia:* Use los Ejercicios 4.14(a) y 2.6(b).)

4.16. Sean α y β números positivos. Demuestre que

$$\begin{aligned} f(x) &:= 0 \quad \text{ si } x < 0, \\ &:= \alpha\beta^\alpha x^{\alpha-1} e^{-(\beta x)^\alpha} \ \text{ si } x \geq 0 \end{aligned}$$

es una función de densidad de probabilidad. A f se le llama la **densidad Weibull con parámetros** α, β. (Observe que si $\alpha = 1$ se obtiene una densidad exponencial con parámetro β.)

4.17. Use la idea de la Proposición 4.22 para demostrar el Lema 3.13.

5 Vectores aleatorios

Contenido: Vector aleatorio, distribución conjunta, densidad conjunta, distribución marginal, densidad marginal, independencia de vv.aa.

Si $\mathrm{X}_1, \ldots, \mathrm{X}_n$ son vv.aa. sobre $(\Omega, \mathcal{F}, \mathrm{P})$ decimos que

$$\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n) : \Omega \to \mathbb{R}^n$$

es un **vector aleatorio** (de dimensión n). La función de distribución (f.d.) de $\mathbf{X}$ se define, para cada vector $\mathbf{x} = (x_1, \ldots, x_n)$ en $\mathbb{R}^n$, como

$$F_{\mathbf{X}}(\mathbf{x}) := \mathrm{P}(\mathbf{X} \leq \mathbf{x}) = \mathrm{P}\{\mathrm{X}_1 \leq x_1, \ldots, \mathrm{X}_n \leq x_n\}, \tag{5.1}$$

en donde

$$\{\mathrm{X}_1 \leq x_1, \ldots, \mathrm{X}_n \leq x_n\} := \bigcap_{i=1}^{n} \{\mathrm{X}_i \leq x_i\}.$$

A la función $F_{\mathbf{X}}$ también se le llama la **distribución conjunta** de las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$. Por otra parte, si $F_i \equiv F_{\mathrm{X}_i}$ es la f.d. de X_i ($i = 1, \ldots, n$), se dice que F_i es la i–ésima f.d. **marginal** del vector $\mathbf{X}$.

La distribución conjunta en (5.1) tiene esencialmente las mismas propiedades que una f.d. unidimensional; vea la Proposición 4.6. En particular, en lugar de 4.6(b) y 4.6(c) tenemos, respectivamente,

- $F_{\mathbf{X}}(x_1, \ldots, x_n) \to 1$ si $x_i \uparrow \infty$ *para todo* $i = 1, \ldots, n$, y
  $F_{\mathbf{X}}(x_1, \ldots, x_n) \to 0$ si $x_i \downarrow -\infty$ *para alguna* i;
- $F_{\mathbf{X}}(x_1, \ldots, x_n)$ es continua por arriba en cada argumento x_i ($i = 1, \ldots, n$).

Sin embargo, la condición de que $F_{\mathbf{X}}$ sea "no decreciente" (ver 4.6(a)) es un poco más elaborada y la explicaremos sólo en el caso $n = 2$.

Si $\boldsymbol{a} = (a_1, a_2)$ y $\boldsymbol{b} = (b_1, b_2)$ son dos vectores en $\mathbb{R}^2$ con $\boldsymbol{a} < \boldsymbol{b}$ (es decir $a_i < b_i$) definimos el "intervalo" $(\boldsymbol{a}, \boldsymbol{b}] := \{\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2 \mid a_i < x_i \leq b_i$ para $i = 1, 2\}$ y, escribiendo $F(x_1, x_2) := F_{\mathrm{X}}(x_1, x_2)$,

$$\mu_F(\boldsymbol{a}, \boldsymbol{b}] := F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

(Compare esta expresión con (4.10).) Entonces F es "no decreciente" en el sentido de que

- $\mu_F(\boldsymbol{a}, \boldsymbol{b}] \geq 0 \quad \forall\, \boldsymbol{a} < \boldsymbol{b}$.

Por otra parte, se sabe que μ_F se puede "extender" a una única medida de LS sobre $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, es decir una medida μ_F tal que $\mu_F(\mathrm{I}) < \infty$ para cualquier "intervalo" acotado $\mathrm{I} \subset \mathbb{R}^2$. (De hecho, μ_F es una m.p.)

Dada la distribución conjunta $F_{\mathbf{X}}$, para calcular la i–ésima f.d. marginal F_i en un punto arbitrario x_i se toma el límite en (1) cuando $x_j \to \infty$ para todo $j \neq i$, es decir

$$F_i(x_i) = \lim_{x_j \to \infty} F_{\mathbf{X}}(\boldsymbol{x}) \quad \forall\, 1 \leq j \leq n, \;\text{ con }\; j \neq i.$$

Por ejemplo, consideremos el caso $n = 2$, de modo que $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2)$ y (5.1) se reduce a

$$F_{\mathbf{X}}(x_1, x_2) = \mathrm{P}\{\mathrm{X}_1 \leq x_1, \mathrm{X}_2 \leq x_2\}.$$

Si $x_2 \uparrow \infty$, entonces $\{\mathrm{X}_2 \leq x_2\} \uparrow \Omega$, y se sigue que

$$\{\mathrm{X}_1 \leq x_1, \mathrm{X}_2 \leq x_2\} \uparrow \{\mathrm{X}_1 \leq x_1\} \cap \Omega = \{\mathrm{X}_1 \leq x_1\}.$$

Por lo tanto, por la continuidad de P (vea la Proposición 1.12(a)),

$$\lim_{x_2 \to \infty} F_{\mathbf{X}}(x_1, x_2) = \mathrm{P}\{\mathrm{X}_1 \leq x_1\} = F_1(x_1). \tag{5.2}$$

Análogamente,

$$\lim_{x_1 \to \infty} F_{\mathbf{X}}(x_1, x_2) = \mathrm{P}\{\mathrm{X}_2 \leq x_2\} = F_2(x_2). \tag{5.3}$$

Se dice que el vector aleatorio $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ es **discreto** si las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$ son discretas. En este caso la función

$$f(x_1, \ldots, x_n) := \mathrm{P}\{\mathrm{X}_1 = x_1, \ldots, \mathrm{X}_n = x_n\} \tag{5.4}$$

se llama la **función de densidad** del vector $\mathbf{X}$, o también se dice que f es la **densidad conjunta** de las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$. Algunas veces escribimos $f_{\mathbf{X}}$ en lugar de f. Por otra parte, a la densidad $f_i(x) := \mathrm{P}\{\mathrm{X}_i = x\}$ de X_i se le llama la i–ésima **densidad marginal**. La densidad marginal f_i se obtiene sumando la función en (4) sobre todos los valores x_j con $j \neq i$. Por ejemplo, si $n = 2$, podemos escribir (4) como $f(x, y) = \mathrm{P}\{\mathrm{X}_1 = x,\ \mathrm{X}_2 = y\}$ y entonces

$$f_1(x) = \sum_y f(x, y) \quad \forall\, x \in \mathrm{X}, \tag{5.5}$$

$$f_2(y) = \sum_x f(x, y) \quad \forall\, y \in Y. \tag{5.6}$$

La Definición 4.13 se extiende en forma natural a vectores aleatorios: decimos que el vector $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ es **absolutamente continuo** si existe una función de Borel $f : \mathbb{R}^n \to \mathbb{R}$ no negativa y tal que

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(y_1, \ldots, y_n) dy_1, \ldots, dy_n \tag{5.7}$$

para todo vector $(x_1, \ldots, x_n)$ en $\mathbb{R}^n$. Al igual que en el caso discreto, a f se le llama la **función de densidad** de $\mathbf{X}$ o la **densidad conjunta** de las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$. La densidad f_i de la v.a. X_i $(i = 1, \ldots, n)$ se llama la i–ésima **densidad marginal** de $\mathbf{X}$ y se calcula como en (5.5) y (5.6) cambiando las sumas por integrales. Es decir, para $n = 2$, las densidades marginales de $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2)$ están dadas por

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \forall\, x \in \mathbb{R}, \tag{5.8}$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad \forall\, y \in \mathbb{R}. \tag{5.9}$$

Por último, una f.d.p. $F : \mathbb{R}^n \to \mathbb{R}$ se dice que es **continua singular** si es continua y, además, existe un conjunto de Borel $S \subset \mathbb{R}^n$ tal que $\mu_F(S) = 1$, pero $\lambda(S) = 0$, donde λ es la medida de Lebesgue.

Independencia

En 3.7 definimos el concepto de independencia de *eventos* y de *σ–álgebras*. En particular, podemos recordar lo siguiente.

Sea $(\Omega, \mathcal{F})$ un espacio medible e I un conjunto arbitrario de índices. Para cada $i \in \mathrm{I}$, sea $\mathcal{F}_i$ una sub–σ–álgebra de $\mathcal{F}$. Decimos que las σ–álgebras en la familia $\{\mathcal{F}_i, i \in \mathrm{I}\}$ son **independientes** si para cualquier subconjunto finito J de I, y cualesquiera conjuntos $A_i \in \mathcal{F}_i$, se cumple que

$$\mathrm{P}\left(\cap_{i \in \mathrm{J}} A_i\right) = \Pi_{i \in \mathrm{J}}\, \mathrm{P}(A_i).$$

Esta definición de independencia de σ–álgebras se extiende a vv.aa. de la siguiente manera.

Definición 5.1. Para cada $i \in \mathrm{I}$, sea $(\Omega_i, \mathcal{F}_i)$ un espacio medible y

$$\mathrm{X}_i : (\Omega, \mathcal{F}) \to (\Omega_i, \mathcal{F}_i)$$

una v.a., es decir $\mathrm{X}_i^{-1}(B) \in \mathcal{F}$ para cada $B \in \mathcal{F}_i$. Sea $\sigma\{\mathrm{X}_i\} \equiv \mathrm{X}_i^{-1}(\mathcal{F}_i)$ la sub–σ–álgebra de $\mathcal{F}$ generada por X_i (vea el Ejercicio 4.6). Se dice que las vv.aa. $\{\mathrm{X}_i, i \in \mathrm{I}\}$ son **independientes** si la σ–álgebras $\{\mathrm{X}_i^{-1}(\mathcal{F}_i), i \in \mathrm{I}\}$ son independientes.

En la Definición 5.1 es importante notar que las vv.aa. X_i pueden tomar valores en conjuntos *distintos* $(\Omega_i, \mathcal{F}_i)$. En el caso especial en el que $(\Omega_i, \mathcal{F}_i) \equiv (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ y, además, I es un conjunto finito, obtenemos trivialmente el siguiente hecho.

Teorema 5.2. *Sea* $\mathbf{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *un vector aleatorio con distribución conjunta* $F(x_1, \ldots, x_n)$ *y, para cada* $i = 1, \ldots, n$*, sea* F_i *la distribución marginal de* X_i*. Entonces las siguientes proposiciones son equivalentes.*

(a) *Las vv.aa.* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *son independientes.*

(b) $\mathrm{P}\{\mathrm{X}_1 \in B_1, \ldots, \mathrm{X}_n \in B_n\} = \mathrm{P}\{\mathrm{X}_1 \in B_1\} \cdots \mathrm{P}\{\mathrm{X}_n \in B_n\} \; \forall \, B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$.

(c) $F(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n) \; \forall \, (x_1, \ldots, x_n) \in \mathbb{R}^n$.

(d) *Las vv.aa.* $h_1(\mathrm{X}_1), \ldots, h_n(\mathrm{X}_n)$ *son independientes para cualquier conjunto de funciones de Borel*

$$h_i : \mathbb{R} \to \mathbb{R}, \; i = 1, \ldots, n.$$

Con respecto a la parte (d) en el teorema anterior, vea el Ejercicio 5.2.

Para vv.aa. discretas o continuas el concepto de independencia se puede expresar usando densidades. El resultado preciso es el siguiente, cuya demostración se puede ver, por ejemplo, en el libro de Ash (Teorema 5.8.4).

Teorema 5.3. *Sean* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *vv.aa. discretas o continuas con densidad conjunta* $f(x_1, \ldots, x_n)$ *y, para cada* $i = 1, \ldots, n$*, sea* f_i *la densidad marginal de* X_i*. Entonces* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *son independientes ssi*

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n) \quad \forall \, (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{5.10}$$

Sean X y Y vv.aa. discretas con valores en $\{1,2\}$ y $\{1,2,3,4\}$, respectivamente. Suponga que la densidad conjunta está dada como en la siguiente tabla.

	1	2	3	4
1	1/4	1/8	1/16	1/16
2	1/16	1/16	1/4	1/8

Por (5.5), la densidad marginal $f_1(x) = \sum_y f(x,y)$ de X se obtiene sumando cada fila de la tabla lo cual da

$$f_1(1) = f_1(2) = 1/2.$$

Análogamente, la marginal $f_2(y) = \sum_x f(x,y)$ de Y se obtiene sumando cada columna:

$$f_2(1) = f_2(3) = 5/16, \quad f_2(2) = f_2(4) = 3/16.$$

De aquí se sigue que X y Y **no** son independientes porque no se satisface (5.10). Por ejemplo, $f(1,1) = 1/4$ pero $f_1(1) \cdot f_2(1) = (1/2)(5/16) \neq f(1,1)$.

Ejemplo 5.4. (**Distribución normal bivariada.**) Sea $|\rho| < 1$ un número dado y $r := (1-\rho^2)^{1/2}$. Se dice que el vector aleatorio (X, Y) tiene *distribución normal bivariada estándar* si su densidad conjunta $f := \mathbb{R}^2 \to \mathbb{R}$ está dada por

$$f(x,y) = \frac{1}{2\pi r}\, e^{-(x^2-2\rho xy+y^2)/2r^2} \quad \forall\,(x,y) \in \mathbb{R}^2. \tag{5.11}$$

(a) Demuestre que las densidades marginales de X y de Y son ambas la densidad normal estándar, i.e. $\mathrm{X} \sim N(0,1)$ y $Y \sim N(0,1)$.

(b) X y Y son independientes ssi $\rho = 0$.

Solución. (a) Primero observe que "completando cuadrados" el numerador del exponente en (5.11) se puede escribir como

$$x^2 - 2\rho xy + y^2 = (x-\rho y)^2 + y^2(1-\rho^2) = (x-\rho y)^2 + y^2 r^2. \tag{5.12}$$

Por lo tanto, podemos expresar (5.11) en la forma

$$f(x,y) = \frac{1}{2\pi r} e^{-(x-\rho y)^2/2r^2} \cdot e^{-y^2/2}. \tag{5.13}$$

Nótese también que por el Ejemplo 2.11(c), la densidad normal $N(\rho y, r^2)$ tiene densidad

$$g(x) = (2\pi r^2)^{-1/2} e^{-(x-\rho y)^2/2r^2} \tag{5.14}$$

de modo que, como $\int_{-\infty}^{\infty} g(x)dx = 1$, tenemos

$$\int_{-\infty}^{\infty} e^{-(x-\rho y)^2/2r^2} dx = (2\pi r^2)^{1/2}.$$

Luego, por (5.9), integrando ambos lados de (5.13) con respecto a x obtenemos la densidad marginal de Y:

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx = (2\pi)^{-1/2} e^{-y^2/2} \quad \forall\, y \in \mathbb{R},$$

es decir, $Y \sim N(0,1)$. Análogamente, intercambiando el papel de x, y en las ecuaciones (5.12)–(5.14) se obtiene que X $\sim N(0,1)$, i.e.

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = (2\pi)^{-1/2} e^{-x^2/2} \quad \forall\, x \in \mathbb{R}.$$

(b) Por el inciso (a),

$$f_1(x) \cdot f_2(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \quad \forall\, (x,y) \in \mathbb{R}^2.$$

Comparando esta expresión con (5.11) se obtiene (b), es decir, $f(x,y) = f_1(x) \cdot f_2(y)$ ssi $\rho = 0$.

Observación 5.5. Hasta ahora, al hablar de la distribución normal o gaussiana $N(m, \sigma^2)$ hemos supuesto que σ^2 es un número positivo. Sin embargo, por razones técnicas es conveniente considerar también $\sigma^2 = 0$. En este caso, $N(m, 0)$ se interpreta como la distribución de la v.a. *constante* m y algunas veces decimos que $N(m, 0)$ es una distribución normal *degenerada*. Por supuesto, la distribución $N(m, \sigma^2)$ tiene la densidad

$$f(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-m)^2/2\sigma^2}$$

ssi $\sigma^2 > 0$. A esta distribución se le llama distribución normal (o gaussiana) *univariada* para distinguirla de la distribución normal *multivariada* que consideramos a continuación. (En particular, vea el caso bivariado en el Ejemplo 5.4.)

Definición 5.6. Un vector aleatorio $(X_1,\ldots,X_n)$ es **gaussiano**, o que tiene **distribución normal multivariada**, si cualquier combinación lineal $a_1X_1+\cdots+a_nX_n$ tiene distribución normal (posiblemente degenerada, por ejemplo $N(0,0)$ si todos los coeficientes $a_1,\ldots,a_n$ son cero).

En la Sección 10 veremos una caracterización de la distribución normal multivariada usando "funciones características".

De la Definición 5.6 se sigue de manera obvia que si $(X_1,\ldots,X_n)$ es un vector gaussiano, entonces las distribuciones marginales también son gaussianas. Sin embargo, **el recíproco es falso**. Es decir, existen vectores aleatorios que no son gaussianos pero cuyas marginales sí son gaussianas, como en el siguiente ejemplo.

Ejemplo 5.7. Considérese la densidad normal bivariada en (5.11) con $\rho=0$, i.e.

$$f(x,y)=\frac{1}{2\pi}e^{-(x^2+y^2)/2} \quad \forall\ (x,y)\in\mathbb{R}^2.$$

Ahora sea (X_1,X_2) el vector con densidad conjunta

$$g(x,y):=\begin{cases} 2f(x,y) & \text{si} \quad xy\geq 0,\\ 0 & \text{en} \quad \text{c.c.}\end{cases}$$

Entonces se puede demostrar (ver Ejercicio 5.7) que las densidades marginales de X_1 y X_2 son ambas $N(0,1)$. Sin embargo, la densidad bivariada g no es gaussiana porque su soporte no es todo el plano ni tampoco es un subespacio de dimensión 1.

Volviendo a la Definición 5.6, si el vector $(X_1,\ldots,X_n)$ es gaussiano, entonces cualquier combinación lineal $a_1X_1+\cdots+a_nX_n$ es una v.a. gaussiana. Sin embargo, una combinación lineal de vv.aa. gaussianas no necesariamente es gaussiana si las variables *no* tienen una distribución normal multivariada (es decir, cada una de las vv.aa. $X_1,\ldots,X_n$ es gaussiana, pero el *vector* $(X_1,\ldots,X_n)$ no es gaussiano). Vea el Ejercicio 5.8.

Ejercicios § 5

5.1. Demuestre que si $X_1, \ldots, X_n$ son vv.aa. independientes, entonces también lo son $X_1, \ldots, X_k$ para $k < n$. (De hecho, se puede ver que cualquier subcolección $X_{i_1}, \ldots, X_{i_k}$ de $X_1, \ldots, X_n$ son independientes.)

5.2. Sean $h_1, \ldots, h_n : \mathbb{R} \to \mathbb{R}$ funciones de Borel. Demuestre que si $X_1, \ldots, X_n$ son vv.aa. independientes, entonces también lo son las vv.aa. $h_1(X_1), \ldots, h_n(X_n)$.

5.3. Sea (X, Y) un vector aleatorio que tiene densidad "uniforme" sobre el disco unitario $D := \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$, es decir, la densidad conjunta de X y Y es

$$f(x, y) = 1/\pi \quad \text{si} \quad (x, y) \in D, \quad \text{y} \quad f(x, y) = 0 \quad \text{si} \quad (x, y) \notin D.$$

(a) Encuentre las densidades marginales de X y Y.

(b) Demuestre que X y Y no son independientes.

5.4. Sea (X, Y) un vector aleatorio con densidad conjunta

$$f(x, y) = \frac{1}{4}[1 + xy(x^2 - y^2)] \quad \text{para} \quad |x| < 1, \ |y| < 1$$

y $f(x, y) = 0$ en c.c. Demuestre que, efectivamente, f es una función de densidad. Calcule las densidades marginales de X y Y.

5.5. Sean X y Y vv.aa. con funciones de distribución F y G, respectivamente, y sea $H(x, y) := P\{X \leq x, Y \leq y\}$ la f.d. conjunta. Para cada $(x, y) \in \mathbb{R}^2$, sean

$$M(x, y) := \min(F(x), G(y)) \quad \text{y} \quad W(x, y) := \max(F(x) + G(y) - 1, 0).$$

Demuestre que:

(a) $W(x, y) \leq H(x, y) \leq M(x, y)$ para todo $(x, y) \in \mathbb{R}^2$. A las funciones M y W se les llama las *cotas de Fréchet–Hoeffding* de la distribución conjunta H.

(b) M y W son funciones de distribución conjunta.

5.6. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$ vv.aa. independientes con funciones de distribución $F_1, \ldots, F_n$, respectivamente. Sean $Y := \max\{\mathrm{X}_1, \ldots, \mathrm{X}_n\}$ y $Z := \min \{\mathrm{X}_1, \ldots, \mathrm{X}_n\}$. Demuestre que Y y Z tienen distribución

$$\begin{aligned} \mathrm{P}\{Y \leq x\} &= F_1(x) \cdots F_n(x), \\ \mathrm{P}\{Z \leq x\} &= 1 - \tilde{F}_1(x) \cdots \tilde{F}_n(x), \end{aligned}$$

para todo $x \in \mathbb{R}$, donde $\tilde{F}_k(x) := 1 - F_k(x)$ para $k = 1, \ldots, n$.

5.7. Sea $g(x, y)$ la densidad conjunta en el Ejemplo 5.7. Demuestre que las densidades marginales de g son ambas $N(0, 1)$.

5.8. Sea X una v.a. $N(0, 1)$, y B una v.a. Bernoulli con $\mathrm{P}(B = i) = 1/2$ para $i = 0, 1$. Supóngase que X y B son independientes. Sea $Y := (-1)^B \mathrm{X}$. Demuestre que Y tiene distribución $N(0, 1)$, pero $\mathrm{X}+Y$ **no** tiene distribución normal. (Observe que $\mathrm{P}(\mathrm{X} + Y = 0) = \mathrm{P}(B = 1) = 1/2$, mientras que $\mathrm{X} + Y = 2\mathrm{X}$ cuando $B = 0$.)

6 Esperanza de vv.aa. discretas y continuas

Contenido: Momentos de una v.a., varianza, desigualdad de Chebyshev.

En este capítulo y el siguiente estudiaremos la **esperanza** o **valor esperado** de una v.a. y algunas de sus propiedades. Primeramente, veremos el caso "elemental" en el que las vv.aa. son discretas o continuas, de modo que sus esperanzas se pueden calcular mediante técnicas elementales. Después, en el Capítulo 7, veremos la esperanza de vv.aa. generales para lo cual se usa la integral de Lebesgue.

Sea X una v.a. discreta con valores en el conjunto $\{x_1, x_2, \ldots\}$ y función de densidad $f_{\mathrm{X}}(x_i) := \mathrm{P}(\mathrm{X} = x_i)$. Si se cumple que

$$\sum_i |x_i| \, f_{\mathrm{X}}(x_i) < \infty, \tag{6.1}$$

definimos la **esperanza** de X como

$$E\mathrm{X} := \sum_i x_i \, f_{\mathrm{X}}(x_i). \tag{6.2}$$

En forma análoga, sea X una v.a. absolutamente continua con densidad $f_{\mathrm{X}} : \mathbb{R} \to \mathbb{R}$. Si

$$\int_{-\infty}^{\infty} |x| \, f_{\mathrm{X}}(x) \, dx < \infty, \tag{6.3}$$

entonces la **esperanza** de X se define como

$$E\mathrm{X} := \int_{-\infty}^{\infty} x \, f_{\mathrm{X}}(x) \, dx. \tag{6.4}$$

A la esperanza de X se le conoce con varios nombres, por ejemplo, **valor esperado** o **valor medio** o **media** de X.

Definición 6.1. Denotaremos por $L_1 \equiv L_1(\Omega, \mathcal{F}, \mathrm{P})$ la familia de vv.aa. sobre $(\Omega, \mathcal{F}, \mathrm{P})$ que tienen esperanza finita (es decir, vv.aa. X que satisfacen (6.1) en el caso discreto, ó (6.3) en el caso absolutamente continuo).

Ejemplo 6.2. Sea X una v.a. continua con densidad de Cauchy

$$f_{\mathrm{X}}(x) = \frac{1}{\pi(1 + x^2)} \quad \forall\, x \in \mathbb{R}.$$

En este caso la integral en (6.3) resulta

$$\int_{-\infty}^{\infty} |x|\, f_{\mathrm{X}}(x)\, dx = \frac{1}{\pi}\int_{0}^{\infty} \frac{2x dx}{1+x^2} \quad \text{(explique)}.$$

Por lo tanto, haciendo el cambio de variable $u := 1 + x^2$ vemos que

$$\int_{-\infty}^{\infty} |x|\, f_{\mathrm{X}}(x) dx = \frac{1}{\pi}\int_{1}^{\infty} \frac{du}{u} = \frac{1}{\pi} \ln u \Big|_{u=1}^{u=\infty} = \infty.$$

Es decir, (6.3) **no** se cumple de modo que la v.a. X **no** está en L_1.

Sea X una v.a. y $h : \mathbb{R} \to \mathbb{R}$ una función de Borel tal que

$$\sum_i |h(x_i)|\, f_{\mathrm{X}}(x_i) < \infty \quad \text{ó} \quad \int_{-\infty}^{\infty} |h(x)|\, f_{\mathrm{X}}(x) dx < \infty, \tag{6.5}$$

si X es discreta o continua, respectivamente. En tal caso veremos posteriormente que la esperanza de $h(\mathrm{X})$ se puede calcular como en (6.2) ó (6.4) sustituyendo x por $h(x)$, i.e.

$$Eh(\mathrm{X}) = \sum_i h(x_i) f_{\mathrm{X}}(x_i) \quad \text{si} \quad \mathrm{X} \quad \text{es discreta}$$

$$Eh(\mathrm{X}) = \int_{-\infty}^{\infty} h(x) f_{\mathrm{X}}(x) dx \quad \text{si} \quad \mathrm{X} \quad \text{es continua.}$$

Observación 6.3. (a) Si $h(x) = x^k$ para algún $k > 0$, a la esperanza $Eh(\mathrm{X}) = E(\mathrm{X}^k)$ se le llama el **momento de orden** k de X. Además, en lugar de decir que X^k está en L_1 frecuentemente diremos que X **está en** $L_k \equiv L_k(\Omega, \mathcal{F}, \mathrm{P})$. Para $k = 1$, el momento de orden 1 de X coincide con la esperanza de X.
(b) Sea $m_{\mathrm{X}} := E\mathrm{X}$ y sea $h(x) := (x - m_{\mathrm{X}})^k$. Entonces, suponiendo que (6.5) se cumple,

$$Eh(\mathrm{X}) := E(\mathrm{X} - m_{\mathrm{X}})^k$$

se llama el **momento central de orden** k de la v.a. X. En particular, para $k = 2$ el momento central de orden 2 se llama la **varianza** de X y se denota por Var(X) ó σ_{X}^2, es decir

$$\mathrm{Var}(\mathrm{X}) \equiv \sigma_{\mathrm{X}}^2 = E(\mathrm{X} - m_{\mathrm{X}})^2. \tag{6.6}$$

Nótese que $(\mathrm{X} - m_{\mathrm{X}})^2 = \mathrm{X}^2 + m_{\mathrm{X}}^2 - 2m_{\mathrm{X}}\mathrm{X}$ de modo que

$$\mathrm{Var}(\mathrm{X}) \equiv \sigma_{\mathrm{X}}^2 = E(\mathrm{X}^2) - m_{\mathrm{X}}^2 \tag{6.7}$$

porque para cualquier constante c y cualquier v.a. $Y \in L_1$ se cumple que

$$E(c) = c \quad \text{y} \quad E(cY) = c\,EY. \quad \text{(Explique.)}$$

La raíz cuadrada positiva de la varianza se llama la **desviación estándar** y se denota por σ_{X}, i.e.

$$\sigma_{\mathrm{X}} := +\sqrt{\mathrm{Var}(\mathrm{X})}.$$

Interpretación física de $E\mathrm{X}$ **y** $\mathrm{Var}(\mathrm{X})$. Sea X una v.a. discreta con valores $x_1, x_2, \ldots$ y densidad $f_{\mathrm{X}}(x_1),\ f_{\mathrm{X}}(x_2), \ldots$ Entonces $E\mathrm{X}$ es el *centro de gravedad* o *punto de equilibrio* del sistema de "masas" $f_{\mathrm{X}}(x_i)$ en los puntos con coordenadas x_i $(i = 1, 2, \ldots)$, y $\mathrm{Var}(\mathrm{X})$ es el *momento de inercia* de dicho sistema con respecto a su centro de gravedad $E\mathrm{X}$. Para una v.a. continua $E\mathrm{X}$ y $\mathrm{Var}(\mathrm{X})$ tienen una interpretación física similar para la "densidad de masa" $f_{\mathrm{X}}(x)$.

Interpretación probabilística (o "frecuencial") de $E\mathrm{X}$. Sea X una v.a. (discreta o continua) en L_1, y sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ vv.aa. con la misma distribución que X y que, además, son independientes. Para cada $n = 1, 2, \ldots,$ sea

$$\mathrm{P}_n := \frac{1}{n}(\mathrm{X}_1 + \cdots + \mathrm{X}_n)$$

el promedio (o "frecuencia") de $\mathrm{X}_1, \ldots, \mathrm{X}_n$. Entonces de acuerdo con la *Ley de los Grandes Números* (que veremos en una sección posterior)

$$\mathrm{P}_n \to E\mathrm{X} \quad \text{"casi seguramente".}$$

Ejemplo 6.4. Sea X una v.a. discreta. Demuestre:

(a) Si $f_{\mathrm{X}}(k) = 1/n$ para $k = 1, \ldots, n$, entonces

$$E\mathrm{X} = \frac{1}{2}(n+1) \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = \frac{1}{12}(n^2 - 1).$$

(b) Si $f_{\mathrm{X}}(k) = \frac{1}{k(k+1)}$ para $k = 1, 2, \ldots$, entonces $E\mathrm{X}$ no existe (i.e. $\mathrm{X} \notin L_1$).

Solución. (a) Usando la fórmula

$$\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$$

vemos que (6.2) resulta

$$EX = \sum_{k=1}^{n} k f_X(k) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{1}{2}(n+1).$$

Para calcular la varianza de X, primero calcularemos el segundo momento $E(X^2)$ de X y después usaremos (6.7) con $m_X := EX$. Para calcular $E(X^2)$ recuérdese que

$$\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1).$$

Por lo tanto,

$$E(X^2) = \sum_{k=1}^{n} k^2 f_X(k) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{6}(n+1)(2n+1).$$

Así pues, de (6.7),

$$\mathrm{Var}(X) = \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \frac{1}{12}(n^2-1).$$

(b) Esto se sigue del hecho de que la serie

$$EX = \sum_{k=1}^{\infty} k f_X(k) = \sum_{k=1}^{\infty} \frac{1}{k+1} = \sum_{k=2}^{\infty} \frac{1}{k}$$

no converge. (**Nota**. La serie $\sum_{k=1}^{\infty} 1/k^p$ converge si $p > 1$ y diverge si $p \leq 1$.)

Teorema 6.5. *La familia $L_k \equiv L_k(\Omega, \mathcal{F}, P)$ de las vv.aa. con momento de orden k finito ($k \geq 1$) es un espacio vectorial; es decir, si $a \in \mathbb{R}$ y $X, Y \in L_k$ entonces aX y $X+Y$ están en L_k. Más generalmente, si $a_1, \ldots, a_n$ son números reales y $X_1, \ldots, X_n$ están en L_k, entonces*

$$a_1X_1 + \cdots + a_nX_n \in L_k.$$

Demostración. La v.a. $a\mathrm{X}$ está en L_k porque $E|a\mathrm{X}|^k = |a|^k E|\mathrm{X}|^k < \infty$. Además, $E(a\mathrm{X})^k = a^k E(\mathrm{X}^k)$.

Para ver que $\mathrm{X} + Y$ está en L_k nótese primero que

$$|x+y|^k \leq 2^k(|x|^k + |y|^k) \quad \forall\, x, y \in \mathbb{R},\ k > 0. \tag{6.8}$$

En efecto, como $|x+y| \leq |x| + |y| \leq 2 \cdot \max\{|x|, |y|\}$, se sigue que

$$|x+y|^k \leq 2^k \cdot (\max\{|x|, |y|\})^k \leq 2^k \cdot (|x|^k + |y|^k).$$

Por lo tanto, usando el hecho de que

$$E(\mathrm{X}_1 + \mathrm{X}_2) = E\mathrm{X}_1 + E\mathrm{X}_2 \quad \text{si} \quad \mathrm{X}_1, \mathrm{X}_2 \in L_1,$$

vemos de (6.8) que

$$E|\mathrm{X} + Y|^k \leq 2^k(E|\mathrm{X}|^k + E|Y|^k) < \infty.$$

□

Teorema 6.6. *(***Desigualdad de Chebyshev***.) Sea $\mathrm{X} \in L_1$ una v.a. no negativa, y $g : [0, \infty) \to [0, \infty)$ una función no decreciente, con $g(x) > 0$ si $x > 0$, y tal que $g(\mathrm{X}) \in L_1$. Entonces para cada $\varepsilon > 0$*

$$\mathrm{P}\{\mathrm{X} \geq \varepsilon\} \leq Eg(\mathrm{X})/g(\varepsilon). \tag{6.9}$$

Casos especiales: (a) Con $\mathrm{X} \in L_k$ y $g(x) = x^k$ $(k \geq 1)$

$$\mathrm{P}\{|\mathrm{X}| \geq \varepsilon\} \leq E|\mathrm{X}|^k/\varepsilon^k.$$

(b) En (a) tómese $k = 2,\ \mathrm{X} \in L_2,\ g(x) = x^2$; además tómese $\mathrm{X} - m_{\mathrm{X}}$ en lugar de X. Entonces

$$\mathrm{P}\{|\mathrm{X} - m_{\mathrm{X}}| \geq \varepsilon\} \leq E(\mathrm{X} - m_{\mathrm{X}})^2/\varepsilon^2,$$

i.e.

$$\mathrm{P}\{|\mathrm{X} - m_{\mathrm{X}}| \geq \varepsilon\} \leq \sigma_{\mathrm{X}}^2/\varepsilon^2 \tag{6.10}$$

o equivalentemente

$$\mathrm{P}\{|\mathrm{X} - m_{\mathrm{X}}| < \varepsilon\} \geq 1 - \sigma_{\mathrm{X}}^2/\varepsilon^2. \tag{6.11}$$

Demostración. Considere la v.a. discreta

$$Y := \begin{cases} 0 & \text{si X} < \varepsilon \\ g(\epsilon) & \text{si X} \geq \varepsilon \end{cases}$$

Entonces (como $\mathrm{X} \geq \varepsilon \Rightarrow g(\mathrm{X}) \geq g(\varepsilon)$)

$$Eg(\mathrm{X}) \geq EY = g(\varepsilon) \cdot \mathrm{P}\{Y = g(\varepsilon)\} = g(\varepsilon)\mathrm{P}\{\mathrm{X} \geq \varepsilon\}. \quad \square$$

$\square$

Ejemplo 6.7. (a) Sea X una v.a. discreta con densidad

$$\begin{aligned} f(x) &:= 1/18 \quad \text{si} \quad x = 1, 3, \\ &:= 16/18 \quad \text{si} \quad x = 2. \end{aligned}$$

Entonces $m_{\mathrm{X}} = 2$ y $\sigma_{\mathrm{X}}^2 = 1/9$. Luego, de (6.10),

$$\mathrm{P}\{|\mathrm{X} - 2| \geq \varepsilon\} \leq 1/9\varepsilon^2 \quad \forall\, \varepsilon > 0. \tag{6.12}$$

En particular, si $\varepsilon = 1$ se tiene *igualdad* en (6.12), porque

$$\mathrm{P}\{|\mathrm{X} - 2| \geq 1\} = \mathrm{P}\{\mathrm{X} = 1 \quad \text{ó} \quad \mathrm{X} = 3\} = 1/9,$$

de modo que, *en general, la desigualdad de Chebyshev no se puede mejorar.*

(b) A pesar de la conclusión en el inciso (a), en algunos casos la estimación que se obtiene de la desigualdad de Chebyshev (6.10) puede ser muy "pobre". Por ejemplo, supóngase que $\mathrm{X} \sim \mathrm{Exp}(\lambda)$, con $\lambda = 1/2$. Entonces, por el Ejercicio 6.2(b), $m_{\mathrm{X}} = 1/\lambda = 2$ y $\sigma_{\mathrm{X}}^2 = 1/\lambda^2 = 4$ de modo que con $\varepsilon = 4$, la desigualdad (6.10) da

$$\mathrm{P}\{|\mathrm{X} - m_{\mathrm{X}}| \geq 4\} \leq \sigma_{\mathrm{X}}^2/16 = 0.25.$$

Sin embargo, si usamos la distribución exponencial $F_{\mathrm{X}}(x) = 1 - e^{-\lambda x}$ (para $x \geq 0$) vemos que

$$\begin{aligned} \mathrm{P}\{|\mathrm{X} - m_{\mathrm{X}}| \geq 4\} &= \mathrm{P}\{\mathrm{X} \geq m_{\mathrm{X}} + 4 \quad \text{ó} \quad \mathrm{X} \leq m_{\mathrm{X}} - 4\} \\ &= \mathrm{P}\{\mathrm{X} \geq 6\} \\ &= 1 - F_{\mathrm{X}}(6) \\ &= e^{-3} = 0.0494 << 0.25. \end{aligned}$$

En otras palabras, usando la distribución *exacta* de X se obtiene una probabilidad mucho menor que la que se obtiene usando (6.10) como una aproximación.

Ejercicios § 6

6.1. En cada uno de los siguientes casos verifique que se cumple el valor dado de $E\mathrm{X}$ y $\mathrm{Var}(\mathrm{X})$.

(a) **Distribución binomial:** $\mathrm{X} \sim \mathrm{Bin}(n, p)$, $f(k) := \binom{n}{k} p^k \; q^{n-k}$ para $k = 0, 1, \ldots, n$, con $q := 1 - p$.

$$E\mathrm{X} = np \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = npq.$$

(b) **Distribución geométrica:** $\mathrm{X} \sim \mathrm{Geo}(p)$, $f(k) := pq^k$ para $k = 0, 1, \ldots,$ con $q := 1 - p$.

$$E\mathrm{X} = q/p \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = q/p^2.$$

(c) **Distribución de Poisson:** $\mathrm{X} \sim \mathrm{Poi}(\lambda)$, $f(k) := e^{-\lambda}\lambda^k/k!$ para $k = 0, 1, \ldots$.

$$E\mathrm{X} = \mathrm{Var}(\mathrm{X}) = \lambda.$$

6.2. Repita el problema anterior para cada una de las siguientes distribuciones continuas.

(a) **Distribución uniforme:** $\mathrm{X} \sim \mathrm{Uni}[a, b]$, $f(x) := 1/(b - a)$ para $a \leq x \leq b$.

$$E\mathrm{X} = \frac{a+b}{2} \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = \frac{(b-a)^2}{12}.$$

(b) **Distribución exponencial:** $\mathrm{X} \sim \mathrm{Exp}(\lambda)$, $f(x) := \lambda e^{-\lambda x}$ para $x \geq 0$.

$$E\mathrm{X} = 1/\lambda \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = 1/\lambda^2.$$

(c) **Distribución normal:** $\mathrm{X} \sim N(m, \sigma^2)$, $f(x) := (2\pi\sigma^2)^{-1/2} e^{-(x-m)^2/2\sigma^2}$ para todo $x \in \mathbb{R}$.

$$E\mathrm{X} = m \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = \sigma^2.$$

(d) **Distribución gama:** $\mathrm{X} \sim \Gamma(p, \lambda)$, $f(x) := \lambda^p x^{p-1} e^{-\lambda x}/\Gamma(p)$ para $x > 0$ y $f(x) := 0$ para $x \leq 0$ (p y λ son parámetros positivos; vea el Ejercicio 2.6).

$$E\mathrm{X} = p/\lambda \quad \text{y} \quad \mathrm{Var}(\mathrm{X}) = p/\lambda^2.$$

6.3. Demuestre que si $\mathrm{X} \in L_2$, entonces la función $f(a) := E(\mathrm{X} - a)^2$ alcanza su mínimo en $a = E\mathrm{X}$, es decir

$$\min_{a \in \mathbb{R}} E(\mathrm{X} - a)^2 = f(E\mathrm{X}) = \mathrm{Var}(\mathrm{X}).$$

6.4. Sea M un entero positivo y $\mathrm{X} \sim \mathrm{Geo}(p)$. Calcule la media de $Y := \min\{\mathrm{X}, M\}$.

6.5. Calcule la esperanza de $Y := 1/(1+\mathrm{X})$, en donde $\mathrm{X} \sim \mathrm{Poi}(\lambda)$.

6.6. Sea N un entero positivo, y sea $f(k) := 2k/N(N+1)$ para $k = 1, \ldots, N$, y $f(k) := 0$ en caso contrario. Demuestre que f es una densidad discreta y calcule su media.

6.7. Sea $\mathrm{X} \sim N(0,1)$. Calcule la media y la varianza de $|\mathrm{X}|, \mathrm{X}^2$, y $e^{t\mathrm{X}}$ para $t \in \mathbb{R}$ fijo.

6.8. Sean $k \leq r$ dos enteros positivos. Demuestre que

(a) $|x|^k \leq |x|^r + 1 \quad \forall\, x \in \mathbb{R}$.

(b) Si $\mathrm{X} \in L_1$ y $\mathrm{X} \geq 0$, entonces $E\mathrm{X} \geq 0$. (Por lo tanto, si $\mathrm{X}, Y \in L_1$ y $\mathrm{X} \geq Y$, entonces $E\mathrm{X} \geq EY$.)

(c) Si $\mathrm{X} \in L_r$ entonces $\mathrm{X} \in L_k$; en otras palabras, si $1 \leq k \leq r$ entonces $L_r \subset L_k$. (*Sugerencia:* use (a) y (b).)

6.9. Sea X una v.a. y M una constante tal que $\mathrm{P}\{|\mathrm{X}| \leq M\} = 1$. Demuestre que

$$|E\mathrm{X}| \leq E|\mathrm{X}| \leq M.$$

(*Sugerencia:* para obtener la primera desigualdad use el Ejercicio 6.8(b).)

6.10. Sea X una v.a. no negativa. Demuestre:

(a) Si X es discreta con valores en $\{0, 1, \ldots\}$, entonces $E\mathrm{X} = \sum_{n=1}^{\infty} \mathrm{P}\{\mathrm{X} \geq n\}$.

(b) Si X es absolutamente continua, entonces $E\mathrm{X} = \int_0^\infty \mathrm{P}\{\mathrm{X} > x\}dx$.

6.11. Sea $\mathrm{X} \sim \mathrm{Geo}(p)$, y sea $q := 1 - p$.

(a) Demuestre que $\mathrm{P}(\mathrm{X} \geq n) = q^n$ para todo $n = 0, 1, \ldots$.

(b) Use el Ejercicio 6.10(a) para demostrar que $E\mathrm{X} = q/p$.

6.12. Sea $\mathrm{X} \sim \mathrm{Exp}(\lambda)$.

(a) Demuestre que $\mathrm{P}(\mathrm{X} > x) = e^{-\lambda x}$ para todo $x > 0$.

(b) Use el Ejercicio 8(b) para demostrar que $E\mathrm{X} = 1/\lambda$.

6.13. Se dice que una v.a. X es **simétrica** si X y $-$X tienen la misma f.d. Por otra parte, una función de densidad f es **simétrica** si f es "par" (o simétrica con respecto al origen), i.e. $f(-x) = f(x)$ para todo $x \in \mathbb{R}$.

(a) Demuestre que X es simétrica ssi $\mathrm{P}\{\mathrm{X} \leq x\} = \mathrm{P}\{\mathrm{X} \geq -x\}$ para todo $x \in \mathbb{R}$.

(b) Si X es una v.a. continua con densidad f_{X}, demuestre que X es simétrica ssi f_{X} es simétrica.

(c) Dé al menos dos ejemplos de vv.aa. simétricas.

6.14. Demuestre que si $\mathrm{X} \sim N(m, \sigma^2)$ y $b \neq 0$, entonces

$$a + b\mathrm{X} \sim N(a + bm,\ b^2\sigma^2).$$

6.15. Demuestre que $\mathrm{X} \sim N(m, \sigma^2)$ ssi $(\mathrm{X} - m)/\sigma \sim N(0, 1)$.

6.16. Demuestre que si $\mathrm{X} \sim N(0, \sigma^2)$, entonces X^2 tiene distribución gama con parámetros $p = 1/2$ y $\lambda = 1/2\sigma^2$. (Vea Ejercicio 6.2(d).)

6.17. Si $\mathrm{X} \sim N(m, \sigma^2)$, calcule la densidad de $Y := e^{\mathrm{X}}$. La densidad de Y se conoce como **densidad lognormal** con parámetros m y σ^2.

7 La integral de Lebesgue

Contenido: La integral de Lebesgue, la integral de Lebesgue–Stieltjes (LS), la esperanza de una v.a.

En esta sección definimos las integrales de Lebesgue y de Lebesgue–Stieltjes (LS) y vemos su relación con la esperanza de una v.a.

Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida y $\mathrm{X} : \Omega \to \mathbb{R}$ una función $\mathcal{F}$–medible. Si $\mathrm{X} = \mathrm{I}_A$ es la función indicadora de un conjunto $A \in \mathcal{F}$ definimos la **integral de Lebesgue** (o simplemente la **integral**) de X con respecto a μ como

$$\int_\Omega \mathrm{X}\, d\mu := \mu(A). \tag{7.1}$$

Supóngase ahora que X es una función **simple**, es decir, X toma únicamente un número finito de valores (distintos) $x_1, \ldots, x_n$ en $\mathbb{R}$. Sea

$$A_i := \{\omega \in \Omega \mid \mathrm{X}(\omega) = x_i\} \quad \forall\, i = 1, \ldots, n,$$

de modo que podemos escribir X en la forma

$$\mathrm{X} = \sum_{i=1}^{n} x_i\, \mathrm{I}_{A_i}.$$

Entonces definimos la **integral** de X con respecto a μ como

$$\int_\Omega \mathrm{X}\, d\mu := \sum_{i=1}^{n} x_i\, \mu(A_i). \tag{7.2}$$

Si X es no negativa definimos su **integral** con respecto a μ como

$$\int_\Omega \mathrm{X}\, d\mu := \sup\{\int_\Omega h\, d\mu \mid h \text{ es simple y } 0 \le h \le \mathrm{X}\}. \tag{7.3}$$

Nótese que este supremo siempre existe pero puede ser $+\infty$; además, se puede mostrar que es único.

Para el caso de variables aleatorias considérese la siguiente definición basada en (7.3). Si X es una v.a. no negativa, se define la esperanza de X como

$$E(X) := \lim_{n\to\infty} E(X_n),$$

donde X_n son vv.aa. simples tales que $X_n \uparrow X$. En la siguiente proposicion vemos que (7.3) es en verdad único.

Proposición 7.1. Si $X_n \uparrow X$ y $Y_n \uparrow X$ para vv.aa. simples $X_n, Y_n, n = 1, 2 \ldots$, entonces

$$\lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} E(Y_n).$$

Demostración. Para m entero fijo y $\epsilon > 0$ arbitrario, sea

$$A_n := \{\omega : X_n \geq Y_m - \epsilon\}.$$

Como $X_n \uparrow X$ y $X \geq Y_m$, entonces $\bigcup_{n\geq 1} A_n = \Omega$. Además

$$X_n = X_n I_{A_n} + X_n I_{A_n^c} \geq X_n I_{A_n} \geq (Y_m - \epsilon) I_{A_n}.$$

Así

$$\begin{aligned}
&E(X_n) \geq E((Y_m - \epsilon)I_{A_n}) = E(Y_m I_{A_n}) - \epsilon P(A_n) \\
&= E(Y_m) - E(Y_m I_{A_n^c}) - \epsilon P(A_n) \geq E(Y_m) - \alpha P(A_n^c) - \epsilon P(A_n),
\end{aligned}$$

donde $\alpha := \max_{\omega\in\Omega} Y_m(\omega)$. Tomando $n \to \infty$ llegamos a

$$\lim_{n\to\infty} E(X_n) \geq E(Y_m) - \epsilon,$$

y como ϵ es arbitrario

$$\lim_{n\to\infty} E(X_n) \geq E(Y_m).$$

Tomando ahora $m \to \infty$,

$$\lim_{n\to\infty} E(X_n) \geq \lim_{m\to\infty} E(Y_m).$$

Aplicando el mismo procedimiento intercambiando las variables llegamos a la desigualdad inversa, luego a la igualdad. □

Ahora sea X una función medible arbitraria y considérese su *parte positiva*

$$\mathrm{X}^+(\omega) := \max\{\mathrm{X}(\omega), 0\} \quad \forall\, \omega \in \Omega$$

y su *parte negativa*

$$\mathrm{X}^-(\omega) := -\min\{\mathrm{X}(\omega), 0\} = \max\{-\mathrm{X}(\omega), 0\} \quad \forall\, \omega \in \Omega.$$

Nótese que $\mathrm{X} = \mathrm{X}^+ - \mathrm{X}^-$ y $|\mathrm{X}| = \mathrm{X}^+ + \mathrm{X}^-$, y que *ambas funciones* X^+ y X^- *son no negativas*. Por lo tanto, por (7.3), sus integrales

$$\int_\Omega \mathrm{X}^+ d\mu \quad \text{y} \quad \int_\Omega \mathrm{X}^- d\mu$$

están bien definidas. Si, además, ambas integrales son *finitas*, entonces decimos que X es **integrable con respecto a** μ y su **integral** es

$$\int_\Omega \mathrm{X}\, d\mu := \int_\Omega \mathrm{X}^+\, d\mu - \int_\Omega \mathrm{X}^-\, d\mu. \tag{7.4}$$

Denotaremos por $L_1(\Omega, \mathcal{F}, \mu)$, o simplemente $L_1(\mu)$ ó L_1, la familia de funciones integrables con respecto a μ. Nótese que $\mathrm{X} \in L_1$ ssi $|\mathrm{X}| \in L_1$, es decir,

$$\int_\Omega |\mathrm{X}| d\mu = \int_\Omega \mathrm{X}^+ d\mu + \int_\Omega \mathrm{X}^- d\mu < \infty.$$

El espacio L_1 es un espacio vectorial (vea el Teorema 6.5), y la integral es un operador lineal y "positivo" sobre L_1 (es decir, $\mathrm{X} \geq 0$ implica que $\int \mathrm{X}\, d\mu \geq 0$).

Definición 7.2. Supóngase que $\mu \equiv \mathrm{P}$ es una m.p. y que X es una v.a. integrable con respecto a P. Entonces la integral de X con respecto a P se llama la **esperanza** de X y escribimos

$$E\mathrm{X} := \int_\Omega \mathrm{X}\, d\mathrm{P}. \tag{7.5}$$

Esta definición coincide con la definición de esperanza para vv.aa. discretas y continuas que vimos en la § 6.

Ejemplo 7.3. Sea X una v.a. discreta con valores $x_1, \ldots, x_n$ y función de densidad

$$f_\mathrm{X}(x_i) = \mathrm{P}\{\mathrm{X} = x_i\} \quad \text{para } \ i = 1, \ldots, n.$$

Sea $A_i := \{\omega \in \Omega | \mathrm{X}(\omega) = x_i\}$ para $i = 1, \ldots, n$. Entonces, de (7.5) y (7.2),

$$\begin{aligned} E\mathrm{X} &:= \int_\Omega \mathrm{X}\, d\mathrm{P} = \sum_{i=1}^n x_i \mathrm{P}(A_i) \\ &= \sum_{i=1}^n x_i\, \mathrm{P}\{\mathrm{X} = x_i\} \\ &= \sum_{i=1}^n x_i\, f_\mathrm{X}(x_i), \end{aligned}$$

que coincide con nuestra definición de $E\mathrm{X}$ en (6.2). De hecho, un argumento similar muestra que (7.5) y (6.2) coinciden para cualquier v.a. discreta X, aunque tenga un número infinito numerable de valores. La condición (6.1) asegura que ambas esperanzas $E\mathrm{X}^+ = \int \mathrm{X}^+ d\mathrm{P}$ y $E\mathrm{X}^- = \int \mathrm{X}^- d\mathrm{P}$ son finitas, porque (6.1) es equivalente a

$$E|\mathrm{X}| = \int_\Omega \mathrm{X}^+ d\mathrm{P} + \int_\Omega \mathrm{X}^- d\mathrm{P} < \infty.$$

Sea F una función de distribución de probabilidad (f.d.p.) y sea μ_F la correspondiente medida de LS; vea (4.9). Si $h : \mathbb{R} \to \mathbb{R}$ es una función de Borel integrable con respecto a μ_F, escribimos

$$\int_{\mathbb{R}} h(x)\ dF(x) := \int_{\mathbb{R}} h\ d\mu_F \tag{7.6}$$

y decimos que (7.6) es la **integral de LS** de h con respecto a μ_F (o con respecto a F). En particular, si $F \equiv F_\mathrm{X}$ es la f.d. de una v.a. X, se puede demostrar que la esperanza de $h(\mathrm{X})$ existe ssi

$$\int_{\mathbb{R}} |h(x)| dF_\mathrm{X}(x) < \infty,$$

en cuyo caso

$$Eh(\mathrm{X}) = \int_\Omega h(\mathrm{X}) d\mathrm{P} = \int_{\mathbb{R}} h(x)\ dF_\mathrm{X}(x). \tag{7.7}$$

De aquí se puede ver que, por ejemplo, si X es (absolutamente) continua con densidad f_X, entonces la esperanza de $h(\mathrm{X})$ coincide con la definición en la Sección 6, i.e.

$$Eh(\mathrm{X}) = \int_{-\infty}^{\infty} h(x) f_\mathrm{X}(x) dx.$$

Más precisamente, tenemos el siguiente resultado.

Teorema 7.4. *Sea* X *una v.a. sobre* $(\Omega, \mathcal{F}, \mathrm{P})$ *y* $h : \mathbb{R} \to \mathbb{R}$ *una función de Borel. Sea* F_X *la f.d. de* X, *y* $\mathrm{P}_\mathrm{X}(B) := \mathrm{P}[\mathrm{X}^{-1}(B)]$, *para* $B \in \mathcal{B}(\mathbb{R})$, *la m.p. inducida por* X *sobre* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Entonces*

(a) $h(\mathrm{X}) \in L_1(\Omega, \mathcal{F}, \mathrm{P})$ *ssi* $h \in L_1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{P}_\mathrm{X})$. *Además,*

(b) *Si se satisface una de las dos condiciones en (a), entonces*

$$\int_\Omega h(\mathrm{X}(\omega)) \mathrm{P}(d\omega) = \int_{\mathbb{R}} h(x) \mathrm{P}_\mathrm{X}(dx). \tag{7.8}$$

(Recuerde que el lado izquierdo de (7.8) coincide con $Eh(\mathrm{X})$; vea (7.7).)

Demostración. Supóngase primero que $h = I_B$, la función indicadora de un conjunto de Borel B. Entonces, como

$$\mathrm{I}_B(\mathrm{X}(\omega)) = 1 \quad \text{ssi} \quad \omega \in \mathrm{X}^{-1}(B),$$

el lado izquierdo de (7.8) resulta

$$\int_\Omega I_B(\mathrm{X}(\omega))\mathrm{P}(d\omega) = \mathrm{P}(\mathrm{X}^{-1}(B)) = \mathrm{P}_\mathrm{X}(B) = \int_{\mathbb{R}} I_B(x)\mathrm{P}_\mathrm{X}(dx). \qquad (7.9)$$

Es decir, (8) se cumple para funciones indicadoras. Luego, por linealidad, también se cumple para funciones *simples*. Ahora supóngase que h es una función de Borel arbitraria pero no negativa. Entonces existe una sucesión no–decreciente de funciones simples h_n tal que $h_n \uparrow h$. Por lo tanto,

$$\begin{aligned} Eh(\mathrm{X}) &= \lim_n Eh_n(\mathrm{X}) \\ &= \lim_n \int h_n(x)\mathrm{P}_\mathrm{X}(dx) \\ &= \int h(x)\mathrm{P}_\mathrm{X}(dx). \end{aligned}$$

De aquí se siguen (a) y (b) para $h \geq 0$. Para h arbitraria, aplicamos el argumento anterior a h^+ y h^-. □

Si X es una v.a. absolutamente continua con densidad f_X, entonces para cualquier intervalo $(a, b]$ tenemos

$$\mathrm{P}_\mathrm{X}(a, b] = \mathrm{P}\{a < \mathrm{X} \leq b\} = F_\mathrm{X}(b) - F_\mathrm{X}(a) = \int_a^b f_\mathrm{X}(x)dx.$$

En general, para cualquier conjunto de Borel $B \subset \mathbb{R}$ tenemos

$$\mathrm{P}_\mathrm{X}(B) = \int_B f_\mathrm{X}(x)dx = \int f_\mathrm{X}(x)I_B(x)dx.$$

Comparando esta igualdad con (7.9) vemos que el mismo argumento de la demostración anterior da lo siguiente.

Corolario 7.5. *Sea* X *una v.a. continua con densidad* f_X*, y sea* h *una función de Borel. Si* $E|h(\mathrm{X})| < \infty$*, entonces*

$$Eh(\mathrm{X}) = \int_{-\infty}^{\infty} h(x)f_\mathrm{X}(x)dx.$$

Observación 7.6. Si $(\Omega, \mathcal{F}, \mu)$ es un espacio de medida arbitraria y $k \geq 1$, la familia de funciones $\mathcal{F}$–medibles $\mathrm{X} : \Omega \to \mathbb{R}$ tales que

$$\int_{\mathbb{R}} |\mathrm{X}|^k d\mu < \infty$$

se denota por $L_k(\Omega, \mathcal{F}, \mu)$. Como en el Teorema 6.5, se puede demostrar que dicha familia es un *espacio vectorial*. De aquí se sigue, en particular, que si $\mathrm{X}, Y \in L_1(\Omega, \mathcal{F}, \mu) \equiv L_1$ y $a, b \in \mathbb{R}$, entonces $a\mathrm{X} + bY$ está en L_1 y

$$\int (a\mathrm{X} + bY) d\mu = a \int \mathrm{X}\, d\mu + b \int Y d\mu.$$

Ejercicios § 7

7.1. Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de probabilidad en donde $\mu \equiv \delta_x$ es la medida de Dirac concentrada en el punto $x \in \Omega$. Demuestre que si $\mathrm{X} : \Omega \to \mathbb{R}$ es una v.a. no negativa, entonces $\int_\Omega \mathrm{X}\, d\mu = \mathrm{X}(x)$.

7.2. Sea $\Omega = \{1, 2, \ldots\}$, $\mathcal{F} = 2^\Omega$ el conjunto potencia de Ω, y $\mathrm{X} : \Omega \to \mathbb{R}$ una función no negativa. Sea μ la medida de conteo (definida en el Ejemplo 1.8) sobre $\mathcal{F}$. Demuestre que $\int_\Omega \mathrm{X}\, d\mu = \sum_{n=1}^\infty \mathrm{X}(n)$.

En los ejercicios siguientes $(\Omega, \mathcal{F}, \mu)$ es un espacio de medida arbitraria y X es una función $\mathcal{F}$–medible.

7.3. Demuestre:

(a) Si $\mathrm{X} = 0$ μ–c.d.q., entonces $\int_\Omega \mathrm{X}\, d\mu = 0$.

(b) Si X y Y son vv.aa. tales que $\mathrm{X} = Y$ c.s., entonces $E\mathrm{X} = EY$.

(c) Si $\mathrm{X} \geq 0$ y $\int_\Omega \mathrm{X}\, d\mu = 0$, entonces $\mathrm{X} = 0$ μ–c.d.q.

7.4. Si $\int_\Omega |\mathrm{X}| d\mu < \infty$, entonces X es *finita* μ–c.d.q.; es decir, el conjunto $A := \{\omega \in \Omega \mid |\mathrm{X}(\omega)| = \infty\}$ tiene medida $\mu(A) = 0$.

7.5. Si X está en $L_1(\Omega, \mathcal{F}, \mu)$ y $\mathrm{X} \geq 0$, demuestre que entonces

$$\int_\Omega \mathrm{X}\, d\mu \geq 0.$$

Por lo tanto, si X y Y son ambas funciones integrables y $\mathrm{X} \geq Y$, entonces

$$\int_\Omega \mathrm{X}\, d\mu \geq \int_\Omega Y\, d\mu.$$

Además, definiendo

$$\int_A \mathrm{X}\, d\mu := \int_\Omega \mathrm{X} \cdot I_A\, d\mu \quad \forall\, A \in \mathcal{F},$$

vemos que si $\mathrm{X} \geq 0$ y $A, B \in \mathcal{F}$ son tales que $A \subset B$, entonces

$$\int_A \mathrm{X}\, d\mu \leq \int_B \mathrm{X}\, d\mu.$$

7.6. Demuestre: si X es integrable, entonces $|\int_\Omega \mathrm{X}\, d\mu| \leq \int_\Omega |\mathrm{X}| d\mu$.

7.7. Sea $\mathrm{X} \geq 0$ una v.a. sobre $(\Omega, \mathcal{F}, \mathrm{P})$ con $E\mathrm{X} = 1$. Defina

$$Q(A) := E(\mathrm{X} I_A) \quad \forall\, A \in \mathcal{F}.$$

Demuestre:

(a) Q es una m.p. sobre $(\Omega, \mathcal{F})$.

(b) Si $\mathrm{P}(A) = 0$ entonces $Q(A) = 0$. Dé un ejemplo mostrando que el recíproco es falso en general, es decir, $Q(A) = 0$ *no* implica $\mathrm{P}(A) = 0$.

(c) Si $\mathrm{P}(\mathrm{X} > 0) = 1$, entonces la esperanza con respecto a Q, que denotamos por $E_Q(\cdot)$, satisface que $E_Q(Y) = E_{\mathrm{P}}(Y\mathrm{X})$, i.e.

$$\int Y\, dQ = \int Y\, \mathrm{X}\, d\mathrm{P}.$$

Además, la m.p. R definida como $R(A) := E_Q(I_A/\mathrm{X})$ para $A \in \mathcal{F}$ coincide con P, i.e. $R(\cdot) = \mathrm{P}(\cdot)$.

7.8. Demuestre el **Teorema de Convergencia Monótona:** Sea $\{\mathrm{X}_n\}$ una sucesión creciente de funciones de Borel, no–negativas, tales que $\mathrm{X}_n \uparrow \mathrm{X}$. Entonces

$$\int_\Omega \mathrm{X}_n\, d\mu \uparrow \int_\Omega \mathrm{X}\, d\mu.$$

7.9. Sean μ y λ dos medidas sobre un espacio medible $(\Omega, \mathcal{F})$, y sea $f : \Omega \to \mathbb{R}$ una función de Borel, no–negativa, tal que

$$\mu(B) = \int_B f(\omega) \lambda(d\omega) \quad \forall\, B \in \mathcal{F}.$$

(En este caso se dice que f es la *densidad* (o *derivada de Radon–Nikodym*) de μ con respecto a λ.) Demuestre que para cualquier función de Borel $g : \Omega \to \mathbb{R}$

$$\int_\Omega g(\omega)\mu(d\omega) = \int_\Omega g(\omega) f(\omega) \lambda(d\omega),$$

en el sentido de que si una de las dos integrales existe, entonces también existe la otra y, además, sus valores coinciden. (*Sugerencia:* Use el mismo argumento que se usó en la demostración de (7.8).)

8 Covarianza e independencia

Contenido: Covarianza, varianza de una suma, leyes débiles de grandes números.

En la Sección 5 estudiamos el concepto de independencia de vv.aa. En la siguiente proposición se muestra que la esperanza del producto de vv.aa. independientes tiene una forma particularmente simple.

Proposición 8.1. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$ vv.aa. independientes.

(a) Si además las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$ están en L_1, entonces su producto $\mathrm{X}_1 \cdots \mathrm{X}_n$ también está en L_1 y

$$E(\mathrm{X}_1 \cdots \mathrm{X}_n) = (E\mathrm{X}_1) \cdots (E\mathrm{X}_n). \tag{8.1}$$

(b) Asimismo, si para cada $k = 1, \ldots, n$, se tiene que $h_k : \mathbb{R} \to \mathbb{R}$ es una función de Borel tal que $h_k(\mathrm{X}_k)$ está en L_1, entonces el producto $h_1(\mathrm{X}_1) \cdots h_k(\mathrm{X}_n)$ está en L_1 y

$$E[h_1(\mathrm{X}_1) \cdots h_n(\mathrm{X}_n)] = Eh_1(\mathrm{X}_1) \cdots Eh_n(\mathrm{X}_n).$$

Demostración. Es claro que (a) y (b) son equivalentes. En efecto, (a) es un caso especial de (b) y, recíprocamente, (b) se sigue de (a) y del Teorema 5.2 (d). Por lo tanto, basta demostrar (por ejemplo) (b) en el caso $n = 2$ y después usar inducción para el caso general. (Explique.) Además, para simplicar la notación escribiremos $h_1(\mathrm{X}_1) \equiv f(\mathrm{X}), h_2(\mathrm{X}_2) \equiv g(Y)$.

Con esta convención, se obtiene (b) si f y g son funciones indicadoras y también, por linealidad, para funciones simples. Supongamos ahora que f y g son funciones medibles no negativas. Entonces existen funciones simples $f_n \geq 0$ y $g_n \geq 0$ tales que $f_n \uparrow f$ y $g_n \uparrow g$. Por lo tanto, $f_n(\mathrm{X})g_n(Y) \uparrow f(\mathrm{X})g(Y)$ y se sigue que,

$$\begin{aligned} E[f(\mathrm{X})g(Y)] &= \lim_n E[f_n(\mathrm{X})g_n(Y)] \\ &= \lim_n E[f_n(\mathrm{X})]E[g_n(Y_n)] \\ &= E[f(\mathrm{X})]E[g(Y)]. \end{aligned}$$

Finalmente, en el caso general tomamos $f = f^+ - f^-$ y $g = g^+ - g^-$ y (b) se obtiene por linealidad. $\square$

Observación 8.2. La conclusión en 8.1(a) **no** es válida si las vv.aa. $\mathrm{X}_1,\ldots,\mathrm{X}_n$ no son independientes. En otras palabras, si X y Y están en L_1, en general **no** se cumple que el producto XY está en L_1. Por ejemplo, sea $\mathrm{X} = Y$ una v.a. discreta con densidad $f(k) := c/k^3$ para $k = 1, 2, \ldots$, en donde c es una constante para la cual $\sum f(k) = 1$. (Recuerde la Nota al final del Ejemplo 6.4(b).) Entonces $\mathrm{X} = Y$ está en L_1 porque

$$E\mathrm{X} = \sum_{k=1}^{\infty} kf(k) = c \sum_{k=1}^{\infty} k^{-2} < \infty,$$

pero el producto $\mathrm{X}Y = \mathrm{X}^2$ **no** está en L_1, pues

$$E(\mathrm{X}^2) = \sum_{k=1}^{\infty} k^2 f(k) = c \sum_{k=1}^{\infty} k^{-1} \not< \infty.$$

Si X y Y son vv.aa. en L_2, entonces $\mathrm{X}Y \in L_1$ porque $|\mathrm{X}Y| \leq \frac{1}{2}(\mathrm{X}^2 + Y^2)$. En este caso, definimos la **covarianza** de X y Y como

$$\mathrm{Cov}(\mathrm{X}, Y) := E[(\mathrm{X} - m_{\mathrm{X}})(Y - m_Y)], \tag{8.2}$$

en donde $m_{\mathrm{X}} := E\mathrm{X}$ y $m_Y := EY$. Nótese que $\mathrm{Cov}(\mathrm{X}, \mathrm{X}) = \mathrm{Var}(\mathrm{X})$ y, por otra parte,

$$\mathrm{Cov}(\mathrm{X}, Y) = E(\mathrm{X}Y) - m_{\mathrm{X}} m_Y. \tag{8.3}$$

Si $\mathrm{Cov}(\mathrm{X}, Y) = 0$ se dice que X y Y **no están correlacionadas.** De (8.3) y de la Proposición 8.1(a) se deduce lo siguiente.

Proposición 8.3. Si $\mathrm{X}, Y \in L_2$ son independientes, entonces no están correlacionadas.

Observación 8.4. (a) El recíproco de 8.3 es **falso**; es decir, hay vv.aa. que no están correlacionadas y que, sin embargo, **no** son independientes. Como ejemplo de lo anterior vea el Ejercicio 8.1. Otro ejemplo es el siguiente: sean X y Y vv.aa. que tienen densidad conjunta $f(x, y)$ uniforme sobre el disco unitario $D := \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$, es decir

$$f(x, y) := 1/\pi \quad \text{si} \quad (x, y) \in D,$$

y $f(x,y) := 0$ para $(x,y) \notin D$. Entonces la densidad marginal de X y su esperanza son

$$f_{\mathrm{X}}(x) = \frac{2}{\pi}\sqrt{1-x^2} \quad \text{si} \quad -1 \leq x \leq 1, \quad E\mathrm{X} = 0$$

y similarmente para Y. En particular, $f(x,y) \neq f_{\mathrm{X}}(x) \cdot f_Y(y)$ de modo que X y Y **no** son independientes. Sin embargo, un cálculo directo demuestra que X y Y no están correlacionadas porque $E(\mathrm{X}Y) = 0 = E\mathrm{X} \cdot EY$.

(b) Una **excepción** al inciso (a) es cuando la distribución conjunta de X y Y es la **distribución normal bivariada** del Ejemplo 5.4. En tal caso se puede ver que la covariancia de X y Y es el parámetro ρ que aparece en (5.11), es decir $\mathrm{Cov}(\mathrm{X}, Y) = \rho$. Por lo tanto, del inciso (b) de dicho ejercicio concluimos que X y Y son independientes ssi no están correlacionadas (i.e. $\rho = 0$).

Si $\{\mathrm{X}_n\}$ es una colección de vv.aa. que tienen la misma distribución, diremos que las vv.aa. son **identicamente distribuidas**. Asimismo, la abreviación **i.i.d.** significa que son *independientes e identicamente distribuidas*. Además, diremos que las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *no están correlacionadas* si $\mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_j) = 0$ para $i \neq j$.

Proposición 8.5. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$ vv.aa. en L_2. Entonces

$$\mathrm{Var}(\mathrm{X}_1 + \cdots + \mathrm{X}_n) = \sum_{k=1}^{n} \mathrm{Var}(\mathrm{X}_k) + 2\sum_{k<j} \mathrm{Cov}(\mathrm{X}_k, \mathrm{X}_j). \tag{8.4}$$

Por lo tanto, si las vv.aa. no están correlacionadas (en particular, si son independientes)

$$\mathrm{Var}(\mathrm{X}_1 + \cdots + \mathrm{X}_n) = \sum_{k=1}^{n} \mathrm{Var}(\mathrm{X}_k). \tag{8.5}$$

Si además las vv.aa. tienen la misma varianza, digamos $\mathrm{Var}(\mathrm{X}_k) = \sigma^2$, entonces

$$\mathrm{Var}(\mathrm{X}_1 + \cdots + \mathrm{X}_n) = n\sigma^2. \tag{8.6}$$

Nótese que (8.6) se cumple, en particular, si $\mathrm{X}_1, \ldots, \mathrm{X}_n$ son i.i.d. con varianza común σ^2.

Demostración. Sea $S := \mathrm{X}_1 + \cdots + \mathrm{X}_n$ y $m_k := E\mathrm{X}_k$ la media de X_k $(k = 1, \dots, n)$. Entonces

$$ES = m_1 + \cdots + m_n \tag{8.7}$$

y

$$\mathrm{Var}(S) = E(S - ES)^2 = E[\sum_{k=1}^{n} (\mathrm{X}_k - m_k)]^2.$$

De aquí se sigue que (8.4) es consecuencia de la fórmula general

$$(x_1 + \cdots + x_n)^2 = \sum_{k=1}^{n} x_k^2 + 2 \sum_{k<j} x_k x_j.$$

para números reales $x_1, \dots, x_n$. $\square$

En el Ejercicio 8.2 se pide calcular una fórmula un poco más general que (8.4).

Definición 8.6. Si las vv.aa. $\mathrm{X}_1, \dots, \mathrm{X}_n$ son i.i.d. decimos que forman una **muestra aleatoria** de tamaño n. En este caso se dice que $S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n$ es una **suma muestral**.

Ejemplo 8.7. Sea $\mathrm{X}_1, \dots, \mathrm{X}_n$ una muestra aleatoria de vv.aa. en L_2 cada una con media μ y varianza σ^2. Sea $\overline{S}_n := S_n/n$ el **promedio muestral**, i.e.

$$\overline{S}_n := \frac{1}{n}(\mathrm{X}_1 + \cdots + \mathrm{X}_n).$$

Demuestre que cuando $n \to \infty$

$$\mathrm{Var}(\overline{S}_n) \to 0 \tag{8.8}$$

y, además, para cualquier $\varepsilon > 0$

$$\mathrm{P}\{|\overline{S}_n - \mu| \geq \varepsilon\} \to 0. \tag{8.9}$$

Demostración. Por (8.7) y (8.6), $ES_n = n\mu$ y $\mathrm{Var}(S_n) = n\sigma^2$. Por lo tanto,

$$E\overline{S}_n = \frac{1}{n} ES_n = \mu \quad \text{y} \quad \mathrm{Var}(\overline{S}_n) = \frac{1}{n^2}\,\mathrm{Var}(S_n) = \frac{1}{n}\sigma^2.$$

Esta última relación implica (8.8). Por otra parte, de la desigualdad de Chebyshev (6.9) vemos que

$$\mathrm{P}\{|\overline{S}_n - \mu| \geq \varepsilon\} \leq \mathrm{Var}(\overline{S}_n)/\varepsilon^2 = \sigma^2/n\varepsilon^2, \tag{8.10}$$

de lo cual se sigue (8.9). □

Debido a (8.9), en una sección posterior diremos que, cuando $n \to \infty$,

$$\overline{S}_n \to \mu \quad \text{“en probabilidad”}. \tag{8.11}$$

Asimismo, como $\mathrm{Var}(\overline{S}_n) = E(\overline{S}_n - \mu)^2 \to 0$, por (8), diremos que

$$\overline{S}_n \to \mu \quad \text{”en } L_2\text{” (o “en la media de orden 2”)}. \tag{8.12}$$

Los resultados en (8.11) y (8.12) se conocen como *leyes débiles de los grandes números* en probabilidad y en L_2, respectivamente. Nótese que, por (8.10), *la convergencia en* (8.12) *implica* (8.11).

Ejercicios § 8

8.1. Sea $\Omega = \{1, 2, 3\}$ un espacio equiprobable, i.e. $\mathrm{P}(\omega) = 1/3$ para $\omega = 1, 2, 3$. Sean X y Y las vv.aa.

$$\begin{aligned} \mathrm{X}(\omega) &= 1 \quad \text{si} \quad \omega = 1, \\ &= 0 \quad \text{si} \quad \omega = 2, \\ &= -1 \quad \text{si} \quad \omega = 3, \end{aligned} \qquad \begin{aligned} Y(\omega) &= 0 \quad \text{si} \quad \omega = 1 \text{ ó } 3, \\ &= 1 \quad \text{si} \quad \omega = 2. \end{aligned}$$

Demuestre que X y Y no están correlacionadas y que **no** son independientes.

8.2. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$ vv.aa. en L_2, y sean $b, a_1, \ldots, a_n$ números reales. Demuestre que

$$\operatorname{Var}(a_1\mathrm{X}_1 + \cdots + a_n\mathrm{X}_n + b) = \sum_{k=1}^{n} a_k^2 \operatorname{Var}(\mathrm{X}_k) + 2\sum_{k<j} a_k a_j \operatorname{Cov}(\mathrm{X}_k, \mathrm{X}_j).$$

8.3. Sean X y Y vv.aa. en L_2 con desviaciones estándar σ_{X} y σ_Y, respectivamente. Definimos el **coeficiente de correlación** de X y Y como

$$\rho(\mathrm{X}, Y) := \frac{\operatorname{Cov}(\mathrm{X}, Y)}{\sigma_{\mathrm{X}} \cdot \sigma_Y}.$$

(a) Sea a un número real y verifique que

$$0 \leq E(\mathrm{X} + aY)^2 = E(\mathrm{X}^2) + a^2 E(Y^2) + 2aE(\mathrm{X}Y).$$

(b) Tomando $a := -E(\mathrm{X}Y)/E(Y^2)$ deduzca la **desigualdad de Cauchy–Schwarz**:

$$|E(\mathrm{X}Y)| \leq \sqrt{E(\mathrm{X}^2)}\,\sqrt{E(Y^2)}. \qquad (*)$$

(c) Sustituyendo X y Y en (*) por $\mathrm{X} - E\mathrm{X}$ y $Y - EY$, respectivamente, concluya que

$$-1 \leq \rho(\mathrm{X}, Y) \leq 1.$$

(d) Demuestre que en (*) se cumple la *igualdad* ssi se satisface alguna de las siguientes dos condiciones:

(d$_1$) $\mathrm{P}\{\mathrm{X} = 0\} = 1$ ó $\mathrm{P}\{Y = 0\} = 1$;

(d$_2$) X y Y son "linealmente dependientes" con probabilidad 1, en el sentido de que $\mathrm{P}\{\mathrm{X} = cY\} = 1$ para alguna constante c.

8.4. Calcule $\mathrm{Var}(\mathrm{X}^2 Y)$, en donde X y Y son vv.aa. independientes con $E(\mathrm{X}^4) = 2,\ E(\mathrm{X}^2) = 1,\ E(Y^2) = 1$ y $E(Y) = 0$.

8.5. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$ i.i.d. Calcule la media y la varianza de $S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n$ en cada uno de los siguientes casos.

(a) $\mathrm{X}_1 \sim \mathrm{Geo}(p)$.

(b) $\mathrm{X}_1 \sim \mathrm{Poi}(\lambda)$.

(c) $\mathrm{X}_1 \sim \mathrm{Exp}(\lambda)$.

(d) $\mathrm{X}_1 \sim N(m, \sigma^2)$.

8.6. Sea $\mathrm{X}_1, \ldots, \mathrm{X}_n$ una muestra aleatoria de vv.aa. Bernoulli con distribución

$$\mathrm{P}\{\mathrm{X}_1 = 1\} = 1 - \mathrm{P}\{\mathrm{X}_1 = 0\} = 1/2,$$

y sea $\overline{S}_n$ el correspondiente promedio muestral. Demuestre que si $n = 100$, entonces

$$\mathrm{P}\{|\overline{S}_n - 0.5| \geq 0.1\} \leq 0.25.$$

(*Sugerencia:* use (8.10).)

8.7. Sean X y Y i.i.d. con distribución $\mathrm{P}(\mathrm{X} = i) = \mathrm{P}(Y = i) = 1/2$ para $i = 1, -1$. Sea $Z := \mathrm{X}Y$. Demuestre que X, Y y Z *no* son independientes, pero *sí* son independientes de dos en dos.

8.8. Sea $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n) \in \mathbb{R}^n$ un vector aleatorio. La **matriz de covarianza** de X es la matriz C_X $n \times n$ que tiene componentes

$$c_{ij} := \mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_j) \quad \forall\, i, j \in \{1, \ldots, n\}.$$

Demuestre que C_X es **semidefinida positiva**, es decir, es simétrica ($c_{ij} = c_{ji}$ para todo i, j) y, además, $\sum_{ij} a_i a_j c_{ij} \geq 0$ para todo $(a_1, \ldots, a_n) \in \mathbb{R}^n$.

8.9. Sea $\mathrm{X} \in \mathbb{R}^n$ un vector aleatorio con matriz de covarianza C, y sea A una matriz $m \times n$. Demuestre que $Y := A\mathrm{X} \in \mathbb{R}^m$ tiene matriz de covarianza ACA^*, en donde A^* es la *transpuesta* de A.

9 Convergencia de vv.aa.

Contenido: Convergencia puntual, c.d.q., en medida, en L_k, en distribución, teoremas de convergencia monótona, convergencia dominada, leyes de grandes números, convergencia débil de medidas.

Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida y $\mathrm{X}, \mathrm{X}_n : \Omega \to \mathbb{R}$ (para $n = 1, 2, \ldots$) funciones $\mathcal{F}$–medibles.

Definición 9.1. Decimos que la sucesión $\{\mathrm{X}_n\}$

(a) **converge puntualmente** a X si $\mathrm{X}_n(\omega) \to \mathrm{X}(\omega)$ para todo $\omega \in \Omega$;

(b) **converge μ–casi donde quiera** (c.d.q.) si $\mathrm{X}_n \to \mathrm{X}$ excepto en un conjunto de medida cero; es decir, si

$$A := \{\omega \in \Omega \mid \mathrm{X}_n(\omega) \not\to \mathrm{X}(\omega)\}, \tag{9.1}$$

entonces $\mu(A) = 0$. En particular, si $\mu \equiv \mathrm{P}$ entonces el conjunto en donde X_n sí converge a X tiene probabilidad 1, porque $\mathrm{P}(A^c) = 1 - \mathrm{P}(A) = 1$. Por lo tanto, en lugar de $\mathrm{X}_n \to \mathrm{X}$ μ–c.d.q. se dice que $\mathrm{X}_n \to \mathrm{X}$ **con probabilidad 1** (c.p. 1) o **casi seguramente** (c.s.).

Es evidente que en 9.1, (a) $\Rightarrow$ (b). Por otra parte, los tipos de convergencia en 9.1 están relacionados, en particular, con la siguiente pregunta: ¿Cuándo se cumple que $\lim \int \mathrm{X}_n \, d\mu = \int (\lim \mathrm{X}_n) d\mu$? Un resultado de este tipo es el Teorema de Convergencia Monótona que aparece en el Ejercicio 7.8 y que repetimos aquí:

Proposición 9.2. Si $\mathrm{X}_n \geq 0$ para todo n y $\mathrm{X}_n \uparrow \mathrm{X}$, entonces

$$\int \mathrm{X}_n \, d\mu \uparrow \int \mathrm{X} \, d\mu. \tag{9.2}$$

Teorema 9.3. ***Convergencia monótona "extendida".***

(a) *Si $\mathrm{X}_n \geq Y$ para todo n, en donde $\int Y \, d\mu > -\infty$, y $\mathrm{X}_n \uparrow \mathrm{X}$, entonces se cumple (9.2).*

(b) *Si $\mathrm{X}_n \leq Y$ para todo n, en donde $\int Y \, d\mu < \infty$, y $\mathrm{X}_n \downarrow \mathrm{X}$, entonces*

$$\int \mathrm{X}_n \, d\mu \downarrow \int \mathrm{X} \, d\mu.$$

Demostración. (a) Si $\int Y d\mu = +\infty$, entonces $\int \mathrm{X}_n d\mu = +\infty$ para todo n y por lo tanto (9.2) se cumple trivialmente. Supongamos ahora que $\int Y d\mu < \infty$, en cuyo caso $|Y| < \infty$ μ–c.d.q. Redefinimos Y como $Y(\omega) := 0$ si $Y(\omega) = \pm\infty$. Entonces

$$0 \leq \mathrm{X}_n - Y \uparrow \mathrm{X} - Y$$

y se sigue del Ejercicio 7.8 (=Proposición 9.2) que $\int (\mathrm{X}_n - Y) \uparrow \int (\mathrm{X} - Y)$, lo cual implica (9.2).

El inciso (b) se demuestra aplicando (a) a la sucesión $-\mathrm{X}_n \geq -Y$. $\square$

Como aplicaciones de 9.2 y 9.3 tenemos lo siguiente.

Ejemplo 9.4. Demuestre que si $\mathrm{X}_n \geq 0$ para todo n, entonces

$$\int \left(\sum_n^{\infty} \mathrm{X}_n \right) d\mu = \sum_n^{\infty} \int \mathrm{X}_n d\mu. \tag{9.3}$$

En particular, si $\mu \equiv \mathrm{P}$ es una m.p.,

$$E\left(\sum_n \mathrm{X}_n \right) = \sum_n E\mathrm{X}_n \quad \text{si} \quad \mathrm{X}_n \geq 0 \quad \forall\, n.$$

Solución. Para cada $k = 1, 2, \ldots$, sea $Y_k := \sum_{n=1}^{k} \mathrm{X}_n$. Luego, como las funciones X_n son no–negativas, la sucesión $\{Y_k\}$ es no–decreciente y $Y_k \uparrow Y := \sum_{n=1}^{\infty} \mathrm{X}_n$. Por lo tanto (9.3) se sigue de (9.2). (Explique.)

Ejemplo 9.5. Sea $\mathrm{X} \in L_1(\Omega, \mathcal{F}, \mu)$ una función no–negativa y defínase

$$\nu(A) := \int_A \mathrm{X}\, d\mu = \int_\Omega \mathrm{X} \cdot I_A\, d\mu \quad \forall\, A \in \mathcal{F}. \tag{9.4}$$

Demuestre que ν es una medida finita. En particular, si $\mu \equiv \mathrm{P}$ es una m.p. y $0 < E\mathrm{X} < \infty$, entonces $\nu(A) := E(\mathrm{X} \cdot I_A)/E\mathrm{X}$ define una m.p. (Supóngase que $\mathrm{X} \in L_1$, pero **no** satisface la condición $\mathrm{X} \geq 0$. Entonces la función $\nu(\cdot)$ en (9.4) es una **medida con signo**.)

Solución. Es evidente que ν satisface las condiciones (a) y (b) de la Definición 1.7, i.e., $\nu(\phi) = 0$ y $\nu(A) \geq 0$ para todo $A \in \mathcal{F}$. Para demostrar la condición de σ–aditividad en 1.7(c), considere una sucesión $\{A_n\} \subset \mathcal{F}$ de conjuntos *ajenos*, y sea $A := \bigcup_{n=1}^{\infty} A_n$. Deseamos probar que

$$\nu(A) = \sum_{n=1}^{\infty} \nu(A_n). \tag{9.5}$$

Para probar (9.5), sea I_{A_n} la función indicadora de A_n $(n = 1, 2, \ldots)$ y nótese que

$$I_A = \sum_{n=1}^{\infty} I_{A_n} \quad \text{(explique).}$$

Por lo tanto

$$\mathrm{X} \cdot I_A = \sum_{n=1}^{\infty} \mathrm{X} \cdot I_{A_n}$$

y de (9.4) y (9.3) se sigue que

$$\nu(A) = \int \mathrm{X} \cdot I_A \, d\mu = \sum_{n=1}^{\infty} \int \mathrm{X} \cdot I_{A_n} \, d\mu = \sum_{n=1}^{\infty} \nu(A_n).$$

Por último, nótese que ν es finita porque $\nu(\Omega) = \int_\Omega \mathrm{X} \, d\mu < \infty$.

La demostración del siguiente resultado usa el Lema de Fatou (caso general del Ejercicio 9.10).

Teorema 9.6. ***Convergencia dominada.*** *Si* $|\mathrm{X}_n| \leq Y$ *para todo* n*, con* $\int |Y| d\mu < \infty$*, y además* $\mathrm{X}_n \to \mathrm{X}$ μ*–c.d.q., entonces*

$$\int |\mathrm{X}| d\mu < \infty \quad y \quad \int \mathrm{X}_n \, d\mu \to \int \mathrm{X} \, d\mu. \tag{9.6}$$

Demostración. Por el Lema de Fatou,

$$\int (\liminf \mathrm{X}_n) \leq \liminf \int \mathrm{X}_n \leq \limsup \int \mathrm{X}_n \leq \int (\limsup \mathrm{X}_n).$$

Esto implica (9.6) porque $\mathrm{X}_n \to \mathrm{X}$, es decir, $\liminf \mathrm{X}_n = \limsup \mathrm{X}_n = \mathrm{X}$. $\square$

Caso especial: Teorema de convergencia acotada. Supóngase que $\mu \equiv \mathrm{P}$ es una m.p. y que existe una constante M tal que $|\mathrm{X}_n| \leq M$ para todo n. Si además $\mathrm{X}_n \to \mathrm{X}$ c.s., entonces se cumple (9.6), i.e. $E|\mathrm{X}| < \infty$ y $E\mathrm{X}_n \to E\mathrm{X}$.

Ejemplo 9.7. Sea $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ en donde λ es la medida de Lebesgue. Para cada $n = 1, 2, \ldots$ sea X_n la función indicadora del intervalo $[n, \infty)$, y $\mathrm{X} \equiv 0$. Demuestre:

(a) $\mathrm{X}_n \downarrow \mathrm{X}$, y

(b) $0 \leq \mathrm{X}_n \leq 1$ para todo n, pero

(c) $\int \mathrm{X}_n d\lambda \not\to \int \mathrm{X}\, d\lambda$. ¿Es ésto una contradicción al Teorema 9.3 o al 9.6?

Definición 9.8. Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida y sean $\mathrm{X}, \mathrm{X}_n : \Omega \to \mathbb{R}$ $(n = 1, 2, \ldots)$ funciones $\mathcal{F}$–medibles. Decimos que la sucesión $\{\mathrm{X}_n\}$

(a) **converge en medida** a X si

$$\lim_{n\to\infty} \mu\{|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon\} = 0 \quad \forall\, \varepsilon > 0; \tag{9.7}$$

(b) **converge en** L_k a X $(1 \leq k < \infty)$ si X, X_n están en $L_k(\Omega, \mathcal{F}, \mu)$ y

$$\lim_{n\to\infty} \int |\mathrm{X}_n - \mathrm{X}|^k d\mu = 0. \tag{9.8}$$

Si $\mu \equiv \mathrm{P}$ es una m.p. y se cumple (9.7), i.e.

$$\mathrm{P}\{|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon\} \to 0 \quad \forall\, \varepsilon > 0, \tag{9.9}$$

se dice que X_n converge a X **en probabilidad**. Asimismo, si se cumple (9.8), i.e.

$$E|\mathrm{X}_n - \mathrm{X}|^k \to 0, \tag{9.10}$$

se dice que X_n converge a X en L_k o **en la media de orden** k.

Para vv.aa. tenemos un tipo más de convergencia.

Definición 9.9. Sea $\{\mathrm{X}_n\}$ una sucesión de vv.aa. Se dice que $\mathrm{X}_n \to \mathrm{X}$ **en distribución** (o que $\mathrm{X}_n \to \mathrm{X}$ **débilmente**) si

$$F_{\mathrm{X_n}}(x) \to F_{\mathrm{X}}(x) \quad \forall\, x \in C(F_{\mathrm{X}}),$$

en donde $C(F_{\mathrm{X}}) := \{x \in \mathbb{R} | F_{\mathrm{X}} \text{ es continua en } x\}$.

Los distintos tipos de convergencia de vv.aa. están relacionados como sigue

Proposición 9.10. Para vv.aa.:

$$\begin{array}{ccccc} \text{c.s.} & \Rightarrow & \text{Prob.} & \Rightarrow & \text{Distribución} \\ & & \Uparrow & & \\ & & L_k & & \end{array}$$

La demostración de las implicaciones c.s. $\Rightarrow$ Prob. $\Rightarrow$ Distribución se puede ver en el Corolario 9.17 y la Proposición 9.18. La demostración de "$L_k \Rightarrow$ Prob." se sigue de la desigualdad de Chebyshev (6.9) sustituyendo X y $g(x)$ por $|\mathrm{X}_n - \mathrm{X}|$ y $g(x) = x^k$, lo cual da

$$\mathrm{P}\{|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon\} \leq E|\mathrm{X}_n - \mathrm{X}|^k/\varepsilon^k.$$

De esta desigualdad se ve que (9.10) $\Rightarrow$ (9.9).

Un hecho importante es que, en general, los recíprocos de las implicaciones en 9.10 **no** se cumplen, como se muestra en el siguiente ejemplo.

Ejemplo 9.11. Sea $(\Omega, \mathcal{F}, \mathrm{P}) = ([0,1], \mathcal{B}[0,1], \lambda)$, con λ restringida al intervalo $[0,1]$.

(a) Sea $\mathrm{X}_n := n\ I_{[1/n,2/n]}$ para $n = 1, 2, \ldots,$ y $\mathrm{X} \equiv 0$. Entonces $\mathrm{X}_n \to \mathrm{X}$ *en probabilidad* porque, digamos para $0 < \varepsilon \leq 1$,

$$\mathrm{P}\{|\mathrm{X}_n| \geq \varepsilon\} = \lambda[1/n, 2/n] = 1/n \to 0.$$

Sin embargo,

$$E|\mathrm{X}_n|^k = n^k \lambda[1/n, 2/n] = n^{k-1} \not\to 0 \quad \forall\ k \geq 1$$

de modo que X_n **no** converge en L_k. Luego, convergencia en probabilidad **no** implica convergencia en L_k.

Nótese, además, que $\mathrm{X}_n \to \mathrm{X}$ c.s., así que convergencia c.s. **no** implica convergencia en L_k.

(b) Sea $\mathrm{X} \equiv 0$ y sea $\mathrm{X}_n = I_{A_n}$ en donde $A_1 := [0, 1/2]$ y $A_2 := (1/2, 1]$; $A_3 := [0, 1/4]$, $A_4 := (1/4, 1/2]$, $A_5 := (1/2, 3/4]$, $A_6 := (3/4, 1]$; etc. Entonces

$$\mathrm{X}_n(\omega) \not\to \mathrm{X}(\omega) \quad \forall\ \omega,$$

pero $\mathrm{X}_n \to \mathrm{X}$ en probabilidad porque $\mathrm{P}\{|\mathrm{X}_n| \geq \varepsilon\} \to 0$ para cualquier $0 < \varepsilon < 1$. Es decir, convergencia en probabilidad **no** implica convergencia c.s. También se tiene convergencia en L_k.

(c) Primero definiremos dos vv.aa. X y Y que *no coinciden en ningún punto* pero que tienen *la misma distribución*, i.e.

$$\mathrm{X}(\omega) \neq Y(\omega) \quad \forall\, \omega \in \Omega, \quad \text{pero} \quad F_{\mathrm{X}}(x) = F_Y(x) \quad \forall\, x \in \mathbb{R}. \tag{9.11}$$

En efecto, tomando (por ejemplo) $\mathrm{X} := I_{[0,1/2]}$ y $Y := I_{(1/2,1]}$, es evidente que se cumplen las condiciones en (9.11).

Ahora considere la sucesión $\{Y_n\}$ con $Y_n \equiv \mathrm{X}$ para todo n. Entonces se tiene $F_{Y_n} = F_{\mathrm{X}} = F_Y$ para todo n y, por lo tanto,

$$Y_n \to Y \qquad \text{en distribución.}$$

Sin embargo, Y_n **no** converge a Y en probabilidad porque, para cualquier $0 < \varepsilon < 1$,

$$\mathrm{P}\{|Y_n - Y| \geq \varepsilon\} = \mathrm{P}\{|\mathrm{X} - Y| \geq \varepsilon\} = 1 \not\to 0.$$

Sea $\{\mathrm{X}_n\}$ una sucesión de vv.aa. i.i.d. Para cada $n = 1, 2, \ldots$, sean S_n y $\overline{S}_n$ la suma muestral y el promedio muestral, respectivamente, i.e.

$$S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n \quad \text{y} \quad \overline{S}_n := \frac{1}{n} S_n.$$

En (8.11) y (8.12) vimos dos leyes **débiles** de grandes números para el caso en el que las X_n están en L_2, a saber

$$\overline{S}_n \to \mu \quad \text{en} \quad L_2 \quad \text{y (por lo tanto) en probabilidad,}$$

donde $\mu := E(\mathrm{X}_1)$. La ley **fuerte** de los grandes números se refiere a convergencia c.s. y se enuncia como sigue,

Teorema 9.12. ***Ley fuerte de los grandes números.***
Si $\mathrm{X}_1, \mathrm{X}_2, \ldots$ *son vv.aa. i.i.d. con media* μ *finita, entonces*

$$\lim_{n\to\infty} \overline{S}_n = \mu \quad \text{c.s.}$$

Si $E|\mathrm{X}_1| = \infty$*, entonces* $\limsup |S_n/n| = \infty$ *c.s.*

Más adelante se da una demostración de 9.12.

Los dos ejemplos siguientes ilustran el *método Monte Carlo*, que en esencia es una aplicación de la ley fuerte de los grandes números 9.12.

Ejemplo 9.13. Se desea calcular o estimar el área de una región S contenida en el cuadrado unitario $C := [0,1] \times [0,1]$. Selecciónense al azar n puntos en C y sea n' el número de puntos que están en S. Afirmamos que

$$n'/n \approx \alpha := \text{ área de } S \tag{9.12}$$

y que la estimación es mejor cuando n crece.

En efecto, sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad que soporta vv.aa. i.i.d. $p_k, k = 1, 2. \ldots$ uniformes sobre C. Sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ vv.aa. i.i.d. definidas como

$$\mathrm{X}_k(\omega) := \begin{cases} 1 & \text{si } \; p_k(\omega) \in S, \\ 0 & \text{en c.c.} \end{cases}$$

Nótese que las X_k son vv.aa. Bernoulli con parámetro α, pues

$$\mathrm{P}\{\mathrm{X}_k = 1\} = \mathrm{P}\{\omega | p_k(\omega) \in S\} = \lambda(S) = \alpha,$$

donde λ es la medida de Lebesgue. Luego, $E\mathrm{X}_k = \alpha$ y $\sigma^2 :=\mathrm{Var}(\mathrm{X}_k) = \alpha(1-\alpha)$; además, del Teorema 9.12:

$$\frac{1}{n}(\mathrm{X}_1 + \cdots + \mathrm{X}_n) \to \alpha \quad \text{c.s.},$$

que es el enunciado preciso de (9.12). Note que cada $\omega \in \Omega$ hace que todas las vv.aa. p_k tomen un valor en C, por lo que $X_1(\omega), X_2(\omega), \ldots$ es una sucesión de ceros y unos. Además, el error de la estimación se puede precisar usando la desigualdad de Chebyshev:

$$\mathrm{P}\left\{ \left| \frac{1}{n} \sum_{k=1}^{n} \mathrm{X}_k - \alpha \right| \geq \varepsilon \right\} \leq \frac{\sigma^2}{n\varepsilon^2} \quad \forall\, \varepsilon > 0,\ n = 1, 2, \ldots.$$

Ejemplo 9.14. Sea $g : [a,b] \to \mathbb{R}$ una función medible. Se desea calcular la integral

$$I = \int_a^b g(x)dx.$$

Para tal fin, considérese una v.a. X arbitraria pero con densidad $f_{\mathrm{X}}(\cdot) > 0$ sobre $[a,b]$ y sea $Y := g(\mathrm{X})/f_{\mathrm{X}}(\mathrm{X})$. Luego,

$$EY = \int_a^b \left[\frac{g(x)}{f_{\mathrm{X}}(x)} \right] \cdot f_{\mathrm{X}}(x)dx = \int_a^b g(x)dx = I.$$

Por lo tanto, para calcular I consideremos vv.aa. $\mathrm{X}_1, \mathrm{X}_2, \ldots$ i.i.d. con densidad común f_{X}, de modo que las vv.aa. $Y_k := g(\mathrm{X}_k)/f_{\mathrm{X}}(\mathrm{X}_k)$ son i.i.d. con media finita $EY = I$. Luego, por el Teorema 9.12,

$$\frac{1}{n}\sum_{k=1}^{n} Y_k \to I \quad \text{c.p.1.} \tag{9.13}$$

Nótese que el error de la estimación en (9.13), usando la desigualdad de Chebyshev, es

$$\mathrm{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - I\right| \geq \varepsilon\right\} \leq \sigma^2/n\varepsilon^2$$

depende de la variancia $\sigma^2 = \mathrm{Var}(Y) = E(Y^2) - I^2$, i.e.

$$\sigma^2 = \int_a^b \frac{g^2(x)}{f_{\mathrm{X}}(x)} dx - I^2, \tag{9.14}$$

la cual es *mínima* si $f_{\mathrm{X}}(\cdot)$ es proporcional a $|g(\cdot)|$. En efecto, considérese la desigualdad de Cauchy–Shwartz

$$\left(\int_a^b |u(x)v(x)| dx\right)^2 \leq \int_a^b u^2(x)dx \cdot \int_a^b v^2(x)dx \tag{9.15}$$

con $u(x) := g(x)/\sqrt{f_{\mathrm{X}}(x)}$ y $v(x) := \sqrt{f_{\mathrm{X}}(x)}$. Entonces (9.15) resulta

$$\begin{aligned}\left(\int_a^b |g(x)| dx\right)^2 &\leq \int_a^b \frac{g^2(x)}{f_{\mathrm{X}}(x)} dx \cdot \int_a^b f_{\mathrm{X}}(x)dx \\ &= \sigma^2 + I^2 \quad [\text{por (9.14)}],\end{aligned}$$

y por lo tanto

$$\sigma^2 \geq \left(\int_a^b |g(x)| dx\right)^2 - I^2. \tag{9.16}$$

Finalmente, si tomamos $f_{\mathrm{X}}(x) := |g(x)|/C$, con $C := \int_a^b |g(x)|dx$, vemos que

$$\int_a^b \frac{g^2(x)}{f_{\mathrm{X}}(x)} dx = \left(\int_a^b |g(x)| dx\right)^2$$

y se sigue de (9.14) que $\sigma^2 = \mathrm{Var}(Y)$ coincide con el lado derecho de (9.16).

El siguiente resultado se demuestra usando el hecho de que la función exponencial e^x satisface que

$$\lim_{n\to\infty}\left(1+\frac{x_n}{n}\right)^n = e^x \quad \text{si} \quad x_n \to x. \tag{9.17}$$

En algunos textos de cálculo el límite (17) aparece en la forma

$$\lim_{n\to\infty}\left[1+\frac{x}{n}+o\left(\frac{1}{n}\right)\right]^n = e^x, \tag{9.18}$$

en donde $o(1/n)$ es cualquier sucesión tal que $n\cdot o(1/n) \to 0$ cuando $n \to \infty$.

A continuación tenemos una aproximación de Poisson a la distribución binomial.

Teorema 9.15. ***Teorema límite de Poisson.***
Sea $b(k;n,p)$ la densidad binomial con parámetros n y p, i.e.

$$b(k;n,p) := \binom{n}{k} p^k(1-p)^{n-k} \quad \forall\, k = 0,1,\ldots,n. \tag{9.19}$$

Supóngase que $p = p(n)$ es una función de n y que $p(n) \to 0$ en forma tal que

$$n\, p(n) \to \lambda \quad \textit{cuando} \quad n \to \infty, \tag{9.20}$$

donde $\lambda > 0$. Entonces, cuando $n \to \infty$,

$$b(k;n,p(n)) \to e^{-\lambda}\lambda^k/k! \quad \forall\, k = 0,1,\ldots. \tag{9.21}$$

Demostración. Primero observe que

$$\binom{n}{k} = n\cdot(n-1)\cdots(n-k+1)/k!, \quad p^k = n^k p^k/n^k,$$

$$(1-p)^{n-k} = (1-p)^n/(1-p)^k.$$

Por lo tanto, podemos expresar (9.19) como

$$b(k;n,p) = A(n)\cdot B(n)\cdot C(n),$$

en donde

$$A(n) := n\cdot(n-1)\cdots(n-k+1)/n^k \to 1$$

$$B(n) := (np)^k/k! \to \lambda^k/k! \qquad (\text{por } (9.20))$$

$$C(n) := (1-p)^n/(1-p)^k \to e^{-\lambda}$$

porque $(1-p)^n = (1-np/n)^n \to e^{-\lambda}$ (por (9.20) y (9.17)) y $(1-p)^k \to 1$ (pues $p = p(n) \to 0$). Combinando estos resultados se obtiene (9.21). $\square$

El límite en (9.21) se usa para aproximar $b(k; n, p)$ cuando n es "grande" y el parámetro p es "pequeño", tomando $\lambda \approx np$, i.e.

$$b(k; n, p) \approx e^{-np}(np)^k/k! \tag{9.22}$$

Por ejemplo, suponga que la v.a. $\mathrm{X} \sim \mathrm{Bin}(n, p)$, con $n = 10^3$ y $p = 10^{-4}$, representa el número de accidentes automovilísticos en una cierta intersección de calles, durante algún período dado de tiempo (e.g. entre 4 y 6 p.m.) Se desea calcular la probabilidad de que ocurran dos o más accidentes. Usando directamente la distribución binomial obtenemos

$$\begin{aligned} \mathrm{P}\{\mathrm{X} \geq 2\} &= 1 - \mathrm{P}\{\mathrm{X} < 2\} \\ &= 1 - [\mathrm{P}\{\mathrm{X} = 0\} + \mathrm{P}\{\mathrm{X} = 1\}] \\ &= 1 - (q^n + npq^{n-1}) \end{aligned}$$

con $n = 10^3$, $p = 10^{-4}$ y $q = 1 - p = 0.9999$. Por otra parte, si usamos (9.22) con $\lambda = np = 10^3 \cdot 10^{-4} = 0.1$ tenemos

$$\mathrm{P}\{\mathrm{X} = k\} \approx e^{-\lambda}\lambda^k/k! = e^{-0.1}(0.1)^k/k! \quad \forall\, k = 0, 1, \ldots.$$

Luego,

$$\begin{aligned} \mathrm{P}\{\mathrm{X} \geq 2\} &= 1 - [\mathrm{P}\{\mathrm{X} = 0\} + \mathrm{P}\{\mathrm{X} = 1\}] \\ &\approx 1 - e^{-0.1}(1 + 0.1) = 0.0045 \end{aligned}$$

El siguiente resultado es muy útil para verificar convergencia c.s.; vea, por ejemplo el Ejercicio 9.5.

Proposición 9.16. $\mathrm{X}_n \to \mathrm{X}$ c.s. ssi para cada $\varepsilon > 0$

$$\mathrm{P}(\limsup\{|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon\}) = \lim_{n\to\infty} \mathrm{P}\left[\bigcup_{k=n}^{\infty}\{|\mathrm{X}_k - \mathrm{X}| \geq \varepsilon\}\right] = 0.$$

Demostración. Sea $B(\varepsilon) := \limsup B_n(\varepsilon)$, en donde

$$B_n(\varepsilon) := \{|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon\} \equiv \{\omega |\, |\mathrm{X}_n(\omega) - \mathrm{X}(\omega)| \geq \varepsilon\}.$$

Nótese que

$$\begin{aligned}
\{\omega | \mathrm{X}_n(\omega) \not\to \mathrm{X}(\omega)\} &= \bigcup_{\varepsilon>0} B(\varepsilon) \\
&= \bigcup_{m=1}^{\infty} B(1/m)
\end{aligned}$$

porque $B(\varepsilon) \subset B(\varepsilon')$ para $\varepsilon > \varepsilon'$. Por lo tanto,

$$\mathrm{X}_n \to \mathrm{X} \quad \text{c.s.} \quad \text{ssi} \quad \mathrm{P}(B(\varepsilon)) = 0 \quad \forall\, \varepsilon > 0.$$

□

De la Proposición 9.16 trivialmente se obtiene lo siguiente.

Corolario 9.17. *Convergencia c.s. implica convergencia en probabilidad.*

En la siguiente demostración usaremos el Ejercicio 1.9:

$$|\mathrm{P}(A) - \mathrm{P}(B)| \le \mathrm{P}(A\Delta B).$$

para cualquiera dos eventos A, B.

Proposición 9.18. Convergencia en probabilidad implica convergencia en distribución.

Demostración. Supóngase que $\mathrm{X}_n \to \mathrm{X}$ en probabilidad, es decir, para cada $\varepsilon > 0$,

$$\mathrm{P}(|\mathrm{X}_n - \mathrm{X}| \ge \varepsilon) \to 0.$$

Sea $A_n := \{\mathrm{X}_n \le x\}$ y $B := \{\mathrm{X} \le x\}$. Entonces, usando F y F_n para denotar la distribución de X y de X_n, respectivamente,

$$\begin{aligned}
|F_n(x) - F(x)| &= |\mathrm{P}(A_n) - \mathrm{P}(B)| \\
&\le \mathrm{P}(A_n \Delta B) \\
&= \mathrm{P}(A_n \cap B^c) + \mathrm{P}(B \cap A_n^c).
\end{aligned} \tag{9.23}$$

Supóngase que x está en $C(F)$. Deseamos demostrar que cada término en (9.23) converge a cero. Tómese $\varepsilon > 0$ arbitrario. Entonces

$$\begin{aligned}
\mathrm{P}(A_n \cap B^c) &= \mathrm{P}(\mathrm{X}_n \le x, \mathrm{X} > x, |\mathrm{X}_n - \mathrm{X}| < \varepsilon) \\
&+ \mathrm{P}(\mathrm{X}_n \le x, \mathrm{X} > x, |\mathrm{X}_n - \mathrm{X}| \ge \varepsilon) \\
&\le \mathrm{P}(x < \mathrm{X} \le x + \varepsilon) + \mathrm{P}(|\mathrm{X}_n - \mathrm{X}| \ge \varepsilon) \\
&= F(x+\varepsilon) - F(x) + \mathrm{P}(|\mathrm{X}_n - \mathrm{X}| \ge \varepsilon) \\
&\to F(x+\varepsilon) - F(x) \quad \text{cuando} \quad n \to \infty.
\end{aligned}$$

Finalmente, tomando $\varepsilon \downarrow 0$ se obtiene que el primer término en (9.23) tiende a 0. Analogamente se demuestra que también el segundo término tiende a 0. $\square$

Demostración. Ahora deseamos demostrar la ley fuerte de los grandes numeros, Teorema 9.12. Primero veremos un caso sencillo bajo la **hipotesis adicional** de que existe una constante $c > 0$ tal que

$$E(\mathrm{X}_i - \mu)^4 \leq c. \tag{9.24}$$

Sea $S_n := n\,\overline{S}_n = \mathrm{X}_1 + \cdots \mathrm{X}_n$. Por la hipótesis de independencia junto con $\mu = 0$, todos los términos de la esperanza $E(S_n^4)$ se anulan, excepto los n términos de la forma $E(\mathrm{X}_i^4)$, y los

$$\binom{n}{2}\binom{4}{2} = 3n(n-1)$$

términos de la forma $E(\mathrm{X}_i^2\mathrm{X}_j^2) = E(\mathrm{X}_i^2)E(\mathrm{X}_j^2) = (\sigma^2)^2 = \sigma^4$. Luego, por (9.24),

$$E(S_n^4) \leq nc + 3n(n-1)\sigma^4 \leq (c + 3\sigma^4)n^2 \quad \forall\, n,$$

y por la desigualdad de Chebyshev, para cualquier $\varepsilon > 0$,

$$\begin{aligned} \mathrm{P}\{|\overline{S}_n| \geq \varepsilon\} &\leq \varepsilon^{-4}n^{-4}E(S_n^4) \\ &\leq (c + 3\sigma^4)\varepsilon^{-4}n^{-2}. \end{aligned}$$

Esta desigualdad y la Proposición 9.16 dan que $\overline{S}_n \to 0$ c.s. $\square$

Para demostrar el Teorema 9.12 en el caso general requerimos algunos resultados preliminares que son importantes es sí mismos. Empezaremos con el siguiente.

Teorema 9.19. ***Desigualdad de Kolmogorov.*** *Sean $X_1, X_2, \ldots, X_n$ vv.aa. independientes con media cero y varianza finita. Entonces*

$$(\forall \epsilon > 0)\ P(\max_{k=1,\ldots,n} |S_k| \geq \epsilon) \leq \frac{E(S_n^2)}{\epsilon^2}, \tag{9.25}$$

donde $S_k := \sum_{i=1}^k X_i$.

Demostración. Sea $A := \{\max_{k=1,\ldots,n} |S_k| \geq \epsilon\}$,

$$A_1 := \{|S_1| \geq \epsilon\} \text{ y } A_k := \{|S_1| < \epsilon, \ldots, |S_{k-1}| < \epsilon, |S_k| \geq \epsilon\},\ k = 2, \ldots, n.$$

Note que los conjuntos A_k son ajenos y la union de todos es A, por lo que

$$E(S_n^2) \geq E(S_n^2 I_A) = \sum_{k=1}^{n} E(S_n^2 I_{A_k}).$$

Pero como $E\left(S_k(\sum_{i=k+1}^n X_i) I_{A_k}\right) = 0$ (diga porqué),

$$\begin{aligned} E(S_n^2 I_{A_k}) &= E\left((S_k + X_{k+1} + \ldots + X_n)^2 I_{A_k}\right) \\ &= E\left(S_k^2 I_{A_k}\right) + 2E\left(S_k(\sum_{i=k+1}^{n} X_i) I_{A_k}\right) + E\left((\sum_{i=k+1}^{n} X_i)^2 I_{A_k}\right) \geq E(S_k^2 I_{A_k}). \end{aligned}$$

Entonces

$$E(S_n^2) \geq \sum_{k=1}^{n} E(S_k^2 I_{A_k}) \geq \epsilon^2 \sum_{k=1}^{n} P(A_k) = \epsilon^2 P(A).$$

De donde surge (9.25). □

El siguiente resultado se conoce como **Teorema de Kolmogorov-Kinchin**.

Teorema 9.20. *Sean $X_1, X_2, \ldots$ vv.aa. independientes con media cero y varianza finita, y sea $S_n := \sum_{k=1}^n X_k$.*
Si $\sum_{n=1}^\infty E(X_n^2) < \infty$, entonces S_n converge c.s. cuando $n \to \infty$.

Demostración. Sea $\epsilon_0 > 0$ arbitrario. Vemos que

$$\{\omega : \{S_k(\omega)\}_{k=1}^\infty \text{ no es de Cauchy}\} = \bigcup_{0<\epsilon<\epsilon_0} B(\epsilon),$$

donde

$$B(\epsilon) := \left\{\omega : \lim_{n\to\infty} \sup_{k\geq n} |S_k(\omega) - S_n(\omega)| > \epsilon \right\}.$$

Además, si $\epsilon_2 > \epsilon_1$, $B(\epsilon_2) \subset B(\epsilon_1)$. Entonces $\bigcup_{0<\epsilon<\epsilon_0} B(\epsilon) = \bigcup_{m\geq N} B(1/m)$ con $\epsilon_0 < 1/N$. Veamos que $P(B(\epsilon)) = 0$.

Notemos primero que

$$B(\epsilon) = \bigcap_{n\geq 1} \bigcup_{k>n} \left\{\omega : \left|\sum_{i=n+1}^{k} X_i(\omega)\right| > \epsilon \right\}.$$

Luego, usando la desigualdad de Kolmogorov tenemos que para cualquier n

$$\begin{aligned}
P(B(\epsilon)) &\leq P\left(\bigcup_{k>n}\left\{\left|\sum_{i=n+1}^{k} X_i\right| > \epsilon\right\}\right) \\
&= \lim_{N\to\infty} P\left(\bigcup_{k=n+1}^{N}\left\{\left|\sum_{i=n+1}^{k} X_i\right| > \epsilon\right\}\right) \\
&= \lim_{N\to\infty} P\left(\max_{k=n+1,\ldots,N} |X_{n+1}+\ldots+X_k| > \epsilon\right) \\
&\leq \lim_{N\to\infty} \frac{1}{\epsilon^2} E\left(|X_{n+1}+\ldots+X_N|^2\right) \\
&= \lim_{N\to\infty} \frac{1}{\epsilon^2} \sum_{i=n+1}^{N} E\left(X_i^2\right) = \frac{1}{\epsilon^2}\sum_{i=n+1}^{\infty} E\left(X_i^2\right) = 0.
\end{aligned}$$

Tomando el límite cuando $n \to \infty$ obtenemos el resultado. □

Los resultados (i) y (ii) en el siguiente lema se conocen como **Lema de Toeplitz** y **Lema de Kronecker**, respectivamente, y dejaremos su demostracion al lector (ver Ejercicio 9.20).

Lema 9.21. i) Sean $\{a_n\}_{n=1}^{\infty} \subset [0,\infty)$ y $b_n := \sum_{i=1}^{n} a_i$. Suponga que $b_n \to \infty,\ n \to \infty$. Sea también $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$ tal que $x_n \to x,\ n \to \infty$. Entonces

$$\frac{1}{b_n}\sum_{i=1}^{n} a_i x_i \to x,\ n \to \infty.$$

ii) Sea $\{b_k\}_{k=1}^{\infty} \subset (0,\infty)$ creciente y tal que $b_n \to \infty,\ n \to \infty$. Sea también $\{x_n\}_{n=1}^{\infty}$ tal que $\sum_{i=1}^{n} x_i$ converge cuando $n \to \infty$. Entonces

$$\frac{1}{b_n}\sum_{i=1}^{n} b_i x_i \to 0,\ n \to \infty.$$

Lema 9.22. Sea $X \geq 0$ v.a. Entonces

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

Demostración. Sabemos que

$$E(X) = \int_0^\infty xF(dx) = \sum_{n=1}^\infty \int_{[n-1,n)} xF(dx),$$

donde F es la f.d.p. de X. Podemos pues acontar $\int_{[n-1,n)} xF(dx)$ por debajo o por arriba, con $n-1$ o n, respectivamente. $\square$

Podemos ahora demostrar el Teorema 9.12.

Demostración. Sin pérdida de generalidad supongamos que $E(X_1) = 0$. Tenemos que

$$\begin{aligned}
&P(|X_n| \geq n \text{ para una infinidad de valores } n = 1, 2, \ldots) \\
&= P(\limsup\{|X_n| \geq n\}) = \lim_{k\to\infty} P(\bigcup_{n=k}^\infty \{|X_n| \geq n\}) \\
&\leq \lim_{k\to\infty} \sum_{n=k}^\infty P(|X_n| \geq n) = \lim_{k\to\infty} \sum_{n=k}^\infty P(|X_1| \geq n) = 0,
\end{aligned}$$

pues del lema anterior sabemos que $\sum_{n=1}^\infty P(|X_1| \geq n) \leq E(|X_1|) < \infty$. Entonces se tiene que

$$\frac{1}{n}\sum_{k=1}^n X_k I_{\{|X_k|\geq k\}} \to 0 \ c.s.$$

Considere la v.a. truncada $\tilde{X}_n := X_n I_{\{|X_n|<n\}}$. Si $\frac{1}{n}\sum_{k=1}^n \tilde{X}_k \stackrel{c.s.}{\to} 0$ entonces $\frac{1}{n}\sum_{k=1}^n X_k \stackrel{c.s.}{\to} 0$ pues

$$\frac{1}{n}\sum_{k=1}^n X_k = \frac{1}{n}\sum_{k=1}^n \tilde{X}_k + \frac{1}{n}\sum_{k=1}^n X_k I_{\{|X_k|\geq k\}}.$$

Como $E(\tilde{X}_n) \to 0$, por el Lema de Toeplitz $\frac{1}{n}\sum_{k=1}^n E(\tilde{X}_k) \to 0$; entonces para mostrar $\frac{1}{n}\sum_{k=1}^n \tilde{X}_k \stackrel{c.s.}{\to} 0$ basta mostrar que

$$\frac{1}{n}\sum_{k=1}^n (\tilde{X}_k - E(\tilde{X}_k)) \to 0 \ c.s.$$

A su vez, por el Lema de Kronecker, esto queda demostrado si se prueba que

$$\sum_{k=1}^{n} \frac{\tilde{X}_k - E(\tilde{X}_k)}{k} \text{ converge c.s.}$$

Así, para demostrar esto último, por el Teorema de Kolmogorov-Khinchin, verificamos que

$$\sum_{k=1}^{\infty} E\left[\left(\frac{\tilde{X}_k - E(\tilde{X}_k)}{k}\right)^2\right] < \infty.$$

En efecto, primero vemos que

$$\sum_{k=1}^{\infty} E\left[\left(\frac{\tilde{X}_k - E(\tilde{X}_k)}{k}\right)^2\right] \leq \sum_{k=1}^{\infty} \frac{E(\tilde{X}_k^2)}{k^2}.$$

De donde obtenemos que

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{E(\tilde{X}_k^2)}{k^2} &= \sum_{k=1}^{\infty} \frac{E\left(X_k^2 I_{\{|X_k| \leq k\}}\right)}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{i=1}^{k} E\left(X_1^2 I_{\{i-1<|X_1|\leq i\}}\right) \\
&= \sum_{i=1}^{\infty} E\left(X_1^2 I_{\{i-1<X_1\leq i\}}\right) \sum_{k=i}^{\infty} \frac{1}{k^2} \leq 2 \sum_{i=1}^{\infty} \frac{1}{i} E\left(X_1^2 I_{\{i-1<|X_1|\leq i\}}\right) \\
&\leq 2 \sum_{i=1}^{\infty} E\left(|X_1| I_{\{i-1<|X_1|\leq i\}}\right) = 2E(|X_1|) < \infty.
\end{aligned}$$

Donde usamos que $\sum_{k=i}^{\infty} 1/k^2 = 1/i^2 + \sum_{k=i+1}^{\infty} 1/k^2 \leq 1/i + 1/i$. □

Por último presentamos un ejemplo de convergencia c.s. usando el Lema de Borel-Cantelli.

Ejemplo 9.23. Sean $X_1, X_2, \ldots$ vv.aa i.i.d. $\sim exp(1)$. Entonces

$$\lim_{n\to\infty} \sup_{k\geq n} \frac{X_k}{\log k} = 1 \text{ c.p.1.}$$

Demostración. Sea $\epsilon > 0$ arbitrario. Vemos que

$$p := P\left(\begin{array}{c}\text{Para una infinidad de } k \\ \frac{X_k}{\log k} \geq 1+\epsilon\end{array}\right) = P\left(\limsup \frac{X_k}{\log k} \geq 1+\epsilon\right).$$

Si definimos $A_n = \left\{ \frac{X_n}{\log n} \geq 1 + \epsilon \right\}$, entonces

$$P(A_n) = P(X_n \geq (1+\epsilon)\log n) = e^{-(1+\epsilon)\log n} = \frac{1}{n^{1+\epsilon}}.$$

De aqui se sigue que $\sum_{n=1}^{\infty} P(A_n) < \infty$. Luego por el Lemma 3.11, $P(\limsup A_n) = 1$, lo cual implica que $p = 0$. Tomando $\epsilon \to 0$, tenemos que

$$\limsup \frac{X_n}{\log n} \leq 1 \text{ c.p.1.}$$

Por otro lado, sea

$$q := P \left(\begin{array}{c} \text{Para una infinidad de } k \\ \frac{X_k}{\log k} \geq 1 - \epsilon \end{array} \right) = P \left(\limsup \frac{X_k}{\log k} \geq 1 - \epsilon \right).$$

Ahora definimos $B_n = \left\{ \frac{X_n}{\log n} \geq 1 - \epsilon \right\}$. Por lo tanto $P(B_n) = \frac{1}{n^{1-\epsilon}}$, lo que implica $\sum_{n=1}^{\infty} P(B_n) = \infty$, lo cual también se cumple para $\epsilon = 0$. Luego, de la independencia de las vv.aa. resulta que los B_n son eventos independientes (vea Definición 5.1). Asi, de nuevo por el Lema 3.11, se obtiene $q = 1$. Tomando $\epsilon \to 0$, tenemos que

$$\limsup \frac{X_n}{\log n} \geq 1 \text{ c.p.1.}$$

De aquí concluimos el resultado. □

Convergencia débil

Sean μ y μ_n $(n = 1, 2, \ldots)$ medidas de probabilidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, y sea $C_b(\mathbb{R}) \equiv C_b$ el conjunto de las funciones continuas y acotadas de $\mathbb{R}$ en sí mismo.

Definición 9.24. La sucesión $\{\mu_n\}$ **converge débilmente** a μ si

$$\int_{\mathbb{R}} h \, d\mu_n \to \int_{\mathbb{R}} h \, d\mu \quad \forall\, h \in C_b.$$

Supóngase que μ y μ_n es la m.p. inducida por X y X$_n$, respectivamente, $n = 1, 2, \ldots$. A continuación demostraremos, en particular, lo siguiente:

Teorema 9.25. *Las siguientes condiciones son equivalentes:*

(a) $\mu_n \to \mu$ *débilmente.*

(b) $\mathrm{X}_n \to \mathrm{X}$ *en distribución.*

(c) $Eh(\mathrm{X}_n) \to Eh(\mathrm{X})$ *para toda función* $h \in C_b$.

La equivalencia de (a) y (c) se obtiene observando que

$$Eh(\mathrm{X}_n) = \int_\Omega h(\mathrm{X}_n) d\mathrm{P} = \int_{\mathbb{R}} h(x)\mu_n(dx). \tag{9.26}$$

Por otra parte, la equivalencia de (a) y (b) se obtendrá de los siguientes dos resultados. (Compare el siguiente lema con la Proposición 4.10.)

Lema 9.26. Sea F una f.d.p. Sea Y una v.a. con distribución Uni$[0,1]$ y defínase la función $\varphi(y) := \inf\{x|F(x) \geq y\}$ para $0 < y < 1$. Entonces F es la f.d.p. de $\varphi(Y)$, es decir,

$$\mathrm{P}\{\varphi(Y) \leq x\} = F(x) \quad \forall\, x \in \mathbb{R}.$$

Demostración. Puesto que F es continua por la derecha, se tiene que

$$\inf\{x|F(x) \geq y\} = \min\{x|F(x) \geq y\},$$

es decir, el ínfimo se alcanza en algún punto. Por lo tanto, $F(x) \geq y$ si y sólo si $\varphi(y) \leq x$. Luego, como $0 \leq F(x) \leq 1$ y $Y \sim$ Uni$[0,1]$,

$$\mathrm{P}\{\varphi(Y) \leq x\} = \mathrm{P}\{Y \leq F(x)\} = F(x).$$

□

Denotaremos por ∂A la **frontera** de un conjunto $A \subset \mathbb{R}$, i.e.,

$$\partial A := \{x \in \mathbb{R} | \forall\, \varepsilon > 0,\ A \cap (x-\varepsilon,\ x+\varepsilon) \neq \phi \ \text{ y } \ A^c \cap (x-\varepsilon, x+\varepsilon) \neq \phi\}.$$

Lema 9.27. Las siguientes proposiciones son equivalentes para una sucesión de m.p.'s μ_n sobre $\mathcal{B}(\mathbb{R})$:

(a) $\mu_n \to \mu$ débilmente.

(b) $\mu_n(A) \to \mu(A)$ para todo conjunto de Borel $A \subset \mathbb{R}$ con $\mu(\partial A) = 0$.

(c) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ para todo $x \in \mathbb{R}$ tal que $\mu(\{x\}) = 0$.

(d) Existen vv.aa. $Y, Y_1, Y_2, \ldots$ definidas sobre algún espacio de probabilidad, con $Y \sim \mu$ y $Y_n \sim \mu_n$ para todo n, y tales que $Y_n \to Y$ c.s.

La equivalencia de los incisos (a) y (d) se conoce como **Teorema de Skorokhod**.

Demostración. (b) $\Rightarrow$ (c): Esto es obvio porque la frontera de $(-\infty, x]$ es $\{x\}$.

(a) $\Rightarrow$ (c): Fíjese $x \in \mathbb{R}$ y $\varepsilon > 0$ arbitrarios, y sea f la función in C_b definida como

$$f(t) := \begin{cases} 1 & \text{si} \quad t < x, \\ 0 & \text{si} \quad t > x + \varepsilon, \\ 1 - (t - x)/\varepsilon & \text{si} \quad x \leq t \leq x + \varepsilon. \end{cases}$$

Entonces, como $I_{(-\infty,x]} \leq f \leq I_{(-\infty,x+\varepsilon]}$,

$$\limsup \mu_n((-\infty, x]) \leq \limsup \int f \, d\mu_n = \int f \, d\mu \leq \mu((-\infty, x + \varepsilon]);$$

es decir, $\limsup \mu_n((-\infty, x]) \leq \mu((-\infty, x])$ porque $\varepsilon > 0$ era arbitrario.

Análogamente, sea $g \in C_b$ la función definida como

$$g(t) := \begin{cases} 1 & \text{si} \quad t < x - \varepsilon, \\ 0 & \text{si} \quad t > x, \\ 1 - (t - x + \varepsilon)/\varepsilon & \text{si} \quad x - \varepsilon \leq t \leq x. \end{cases}$$

Entonces, como $I_{(-\infty,x-\varepsilon]} \leq g \leq I_{(-\infty,x]}$, se sigue que

$$\liminf \mu_n((-\infty, x]) \geq \liminf \int g \, d\mu_n = \int g \, d\mu \geq \mu((-\infty, x - \varepsilon]);$$

es decir, $\liminf \mu_n((-\infty, x]) \geq \mu((-\infty, x))$.

Finalmente, si $\mu(\{x\}) = 0$, entonces $\mu((-\infty, x]) = \mu((-\infty, x))$, de modo que

$$\liminf \mu_n((-\infty, x]) = \limsup((-\infty, x]) = \mu((-\infty, x]),$$

como se deseaba demostrar.

(c) $\Rightarrow$ (d): Sean F y F_n ($n = 1, 2, \ldots$) funciones de distribución asociadas a μ y μ_n, respectivamente, es decir, $F(x) := \mu((-\infty, x])$ y $F_n(x) :=$

$\mu_n((-\infty, x])$ para todo $x \in \mathbb{R}$. Sea $(\Omega, \mathcal{F}, \mathrm{P})$ el espacio de probabilidad unitario $([0,1], \mathcal{B}[0,1], \lambda)$, con $\lambda =$ medida de Lebesgue, y sean

$$Y(\omega) := \inf\{x | F(x) \geq \omega\}, \quad Y_n(\omega) := \inf\{x | F_n(x) \geq \omega\}$$

vv.aa. definidas como en el Lema 9.26 con $\omega \in (0,1)$. Luego, $Y \sim \mu$ y $Y_n \sim \mu_n$ para todo n.

Nótese que las gráficas de las funciones Y_n y Y son las inversas de las gráficas de F_n y F. Además, por hipótesis, $F_n(x) \to F(x)$ en cada punto x en el que F es continua. Por lo tanto, $Y_n(\omega) \to Y(\omega)$ en cada punto $\omega \in (0,1)$ correspondiente a un punto en el que F es estrictamente creciente, es decir, para cada ω en donde Y es continua. De aquí se sigue que $Y_n \to Y$ c.s. porque, como Y es no–decreciente, el conjunto en donde Y es discontinua es a lo más numerable.

(d) $\Rightarrow$ (a), (b): Sea h una función en C_b. En particular, h es continua, y como $Y_n \to Y$ c.s., se sigue que $h(Y_n) \to h(Y)$ c.s. Además, h es acotada así que $Eh(Y_n) \to Eh(Y)$ y se obtiene (a).

Para demostrar (b), supóngase que h es una función medible y acotada con $\mu(D_h) = 0$, donde $D_h := \{\text{puntos de discontinuidad de } h\}$. Entonces c.p.1 $Y \notin D_h$ y por lo tanto, $h(Y_n) \to h(Y)$ c.p.1 y, de nuevo, por convergencia acotada se obtiene que $Eh(Y_n) \to Eh(Y)$. Finalmente, para obtener (b), sea A un conjunto de Borel con $\mu(\partial A) = 0$ y tómese $h = I_A$. Luego, como $D_h = \partial A$, se obtiene (b). $\square$

Demostración del Teorema 9.25. Este teorema se obtiene de (9.26) y de la equivalencia de (a) y (c) en el Lema 9.27. $\square$

Para concluir, observe que el Teorema 9.25(b),(c), combinado con los Ejercicios 9.13 y 9.14, da una nueva demostración de la Proposición 9.18, es decir, convergencia en probabilidad implica convergencia en distribución.

Ejercicios § 9

9.1. Sea c una constante. Demuestre que si $X_n \to c$ en distribución, entonces $X_n \to c$ en probabilidad.

9.2. Sea $\{\mathrm{X}_n\}$ una sucesión de vv.aa. que converge uniformemente a X, i.e. $\sup_\omega |\mathrm{X}_n(\omega) - \mathrm{X}(\omega)| \to 0$ cuando $n \to \infty$. Demuestre que $E\mathrm{X}_n \to E\mathrm{X}$.

9.3. Sea $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ en donde λ es la medida de Lebesgue. Además, para cada $n = 1, 2, \ldots$, sea $\mathrm{X}_n := \frac{1}{n} I_{[0,n]}$ y $\mathrm{X} \equiv 0$. Demuestre que

(a) $\mathrm{X}_n \to \mathrm{X}$ uniformemente y

(b) $0 \leq \mathrm{X}_n \leq 1$ para todo n, pero

(c) $\int \mathrm{X}_n \, d\lambda \not\to \int \mathrm{X} \, d\lambda$. ¿Es ésto una contradicción al Teorema 9.6 ó al Ejercicio 9.1?

9.4. Sea $(\Omega, \mathcal{F}, \mu)$ como en el Ejercicio 9.2, y sean $\mathrm{X}_n := I_{[n,\infty)}$ y $\mathrm{X} \equiv 0$. Demuestre:

(a) $\mathrm{X}_n \downarrow \mathrm{X}$, y

(b) $0 \leq \mathrm{X}_n \leq 1$ para todo n, pero

(c) $\int \mathrm{X}_n d\lambda \not\to \int \mathrm{X} d\lambda$. ¿Es ésto una contradicción a la Proposición 9.2 o al Teorema 9.6?

9.5. Sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ vv.aa. independientes en L_2, con $E(\mathrm{X}_j) = m_j$ y $\mathrm{Var}(\mathrm{X}_j) = \sigma_j^2$. Supóngase que existe una constante M tal que $\sigma_j^2 \leq M$ para todo j. Sea

$$Y_n := \sum_{j=1}^{n} (\mathrm{X}_j - m_j).$$

Demuestre que

(a) $\mathrm{Var}(Y_n) \leq nM$,

(b) $\frac{1}{n} Y_n \to 0$ en L_2 y, por lo tanto, en probabilidad.

Al resultado en (b) se le conoce como *ley débil de los grandes números de Chebyshev.*

9.6. Demuestre que si $\sum_{n=1}^{\infty} \mathrm{P}\{|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon\} < \infty$ para cada $\varepsilon > 0$, entonces $\mathrm{X}_n \to \mathrm{X}$ c.s. (*Sugerencia:* use la Proposición 9.16.)

9.7. Demuestre:

(a) Si X, X_n $(n = 1, 2, \ldots)$ son vv.aa. en L_2 tales que $\sum_{n=1}^{\infty} E(\mathrm{X}_n - \mathrm{X})^2 < \infty$, entonces $\mathrm{X}_n \to \mathrm{X}$ c.s. (*Sugerencia:* use el Ejercicio 9.6.)

(b) Si $\{\mathrm{X}_n\}$ es una sucesión de vv.aa. discretas con

$$\mathrm{P}\{\mathrm{X}_n = 1/n\} = \mathrm{P}\{\mathrm{X}_n = -1/n\} = 1/2 \quad \forall\, n = 1, 2, \ldots,$$

entonces $\mathrm{X}_n \to 0$ c.s.

9.8. Demuestre que si $\mathrm{X}_n \to \mathrm{X}$ c.s., entonces $\mathrm{X}_n \to \mathrm{X}$ en probabilidad. (*Sugerencia:* use la Proposición 9.16.)

9.9. Supóngase que el 0.005% de la población muere anualmente debido a un cierto tipo de accidente de trabajo, y que una compañía de seguros tiene entre sus clientes 10,000 que están asegurados contra ese tipo de accidente.

(a) Calcule la probabilidad de que la compañía deba pagar más de 3 pólizas en un año dado.

(b) Repita (a) usando la aproximación de Poisson (9.22).

9.10. Demuestre el **Lema de Fatou**: sean $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}$ vv.aa.

(a) Si $\mathrm{X}_n \geq \mathrm{X}$ para todo n, donde $E(\mathrm{X}) > -\infty$, entonces

$$\liminf_{n\to\infty} E(\mathrm{X}_n) \geq E(\liminf_{n\to\infty} \mathrm{X}_n).$$

(b) Si $\mathrm{X}_n \leq \mathrm{X}$ para todo n, donde $E(\mathrm{X}) < \infty$, entonces

$$\limsup_{n\to\infty} E(\mathrm{X}_n) \leq E(\limsup_{n\to\infty} \mathrm{X}_n).$$

(Sugerencia: Sea $Y_n := \inf_{k\geq n} \mathrm{X}_k$ y $Y := \liminf \mathrm{X}_n = \lim Y_n$, y obsérvese que $Y_n \uparrow Y$. Ahora, para demostrar (a) use el Teorema de convergencia monótona. Para demostrar (b) use el hecho de que $\limsup \mathrm{X}_n = -\liminf(-\mathrm{X}_n)$ y aplique el inciso (a).).

9.11. Sea $f : [0,\infty) \to [0,1]$ la función definida como $f(x) := \frac{x}{1+x}$ para $x \geq 0$. Demuestre que $\mathrm{X}_n \to \mathrm{X}$ en probabilidad ssi

$$\lim_{n\to\infty} E[f(|\mathrm{X}_n - \mathrm{X}|)] = 0.$$

9.12. Demuestre que si $X_n \to X$ en probabilidad y $X_n \in [-c, c]$ para todo n, entonces $X_n \to X$ en L_r para todo $r \geq 1$.

9.13. Si $\mathrm{X}_n \to \mathrm{X}$ en L_p o en probabilidad, entonces existe una subsucesión n_k tal que $\mathrm{X}_{n_k} \to \mathrm{X}$ c.s. cuando $k \to \infty$.

9.14. (Extensión del Teorema de Convergencia Dominada.) Supóngase que $\mathrm{X}_n \to \mathrm{X}$ en probabilidad y que existe $Y \in L_p$ tal que $|\mathrm{X}_n| \leq Y$ para todo n. Entonces X está en L_p y $\mathrm{X}_n \to \mathrm{X}$ en L_p.

9.15. Sea f una función continua. Si $\mathrm{X}_n \to \mathrm{X}$ c.s. ó en probabilidad, entonces $f(\mathrm{X}_n) \to f(\mathrm{X})$ c.s. o en probabilidad, respectivamente.

9.16. Sea μ la medida de Lebesgue sobre el intervalo unitario, y sea μ_n la medida uniforme sobre $\{0, 1, \ldots, n\}$, es decir $\mu_n(\{i/n\}) = 1/(n+1)$ para $i \in \{0, 1, \ldots, n\}$. Demuestre que μ_n converge débilmente a μ.

9.17. Supóngase que $\mathrm{X}_n \to \mathrm{X}$ en distribución, donde $\mathrm{X}_n \geq 0$. Demuestre que

$$E\mathrm{X} \leq \liminf E\mathrm{X}_n.$$

9.18. Sea $\mathrm{X} \equiv 0$ y X_n tal que $\mathrm{P}\{\mathrm{X}_n = n\} = 1/n$ y $\mathrm{P}\{\mathrm{X}_n = 0\} = 1 - 1/n$. Decir en que sentido X_n converge a X.

9.19. Sea μ_n la distribución normal $N(0, 1/n)$. Diga si $\{\mu_n\}$ converge débilmente. En caso afirmativo, diga a qué medida converge $\{\mu_n\}$.

9.20. Demuestre el Lema 9.21.
Sugerencia: Para i), se usa que

$$(\forall\epsilon > 0)(\exists N)(n > N)|x_n - x| < \epsilon,$$

para después acotar

$$\left|\frac{1}{b_n}\sum_{i=1}^{n} a_i x_i - x\right|$$

usando la desigualdad del triangulo. Para ii) se puede usar i) reescribiendo

$$\sum_{i=1}^{n} b_i x_i = \sum_{i=1}^{n} b_i (S_i - S_{i-1}),$$

donde $S_i := \sum_{k=1}^{i} x_k$.

9.21. Demuestre que $E(X^2) < \infty$ si y solo si $\sum_{n=1}^{\infty} nP(|X| > n) < \infty$.

9.22. Sean $X_1, X_2, \ldots$ vv.aa. i.i.d. tales que $P(X_n = -1) = P(X_n = 1) = 1/2$. Demuestre que

$$\frac{X_1 + \ldots + X_n}{\sqrt{n}\log(n)} \to 0 \ c.s.$$

Sugerencia: Vea la técnica de la ley fuerte de los grandes números usando el Lema de Kronecker y el Teorema de Kolmogorov-Khinchin. También use que

$$\sum_{k=1}^{\infty} \frac{1}{k \log^2(k)} < \infty.$$

10 Funciones características y el TLC

Contenido. Función generadora de momentos, función característica, teorema de continuidad, teorema límite central, función generadora de probabilidad.

En este capítulo introducimos dos herramientas muy útiles en la teoría de la probabilidad y sus aplicaciones: la función generadora de momentos (f.g.m.) y la función característica (f.c.) de una variable aleatoria.

La función generadora de momentos

Sea X una v.a. sobre un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. La **función generadora de momentos** (f.g.m.) de X se define como

$$M_X(t) := E(e^{tX}) \tag{10.1}$$

para todo $t \in \mathbb{R}$ para el cual la esperanza en (10.1) es finita. Para tales valores de t, por (7.7) podemos escribir

$$M_X(t) = \int_\Omega e^{tX} dP = \int_\mathbb{R} e^{tx} dF_X(x), \tag{10.2}$$

donde F_X es la f.d. de X. Debido al lado derecho de (10.2), la f.g.m. de X también se conoce como la **transformada de Laplace** de F_X.

Nótese que si X es una v.a. discreta con valores x_k y función de densidad f_X, entonces

$$M_X(t) = \sum_k e^{tx_k} f_X(x_k). \tag{10.3}$$

Por otra parte, si X es (absolutamente) continua con densidad f_X, entonces

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx, \tag{10.4}$$

Como ejemplo, supóngase que X tiene densidad binomial $B(n, p)$, es decir (como en el Ejemplo 2.2(b)),

$$f_X(k) = c(n, k) p^k (1-p)^{n-k}$$

para $k = 0, 1, ..., n$, donde $c(n, k)$ es el coeficiente binomial en (2.7). Luego, por (10.3), la f.g.m. de X es

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{n} c(n,k)(e^t p)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n. \end{aligned}$$

Analogamente, si X tiene densidad exponencial con parámetro $\lambda > 0$ como en el Ejemplo 2.11(b), entonces de (10.4) tenemos que

$$\begin{aligned} M_X(t) &= \lambda \int_0^\infty e^{tx} e^{-\lambda x} dx \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

para $t < \lambda$.

El nombre *función generadora de momentos de* X se debe a que, bajo ciertas condiciones, los momentos $m_k := E(X^k)$ para $k = 0, 1, ...$ se obtienen derivando la f.g.m. y evaluando las derivadas en $t = 0$, es decir,

$$\frac{d}{dt^k} E(e^{tX})|_{t=0} = E(X^k) = m_k, \quad k = 1, 2, ... \tag{10.5}$$

Esto se debe a la definicion de la funcion exponencial,

$$e^x := \sum_{k=0}^{\infty} x^k / k! \quad \forall x \in \mathbb{R}, \tag{10.6}$$

y al siguiente teorema.

Teorema 10.1. *Sea X una v.a. tal que, para algun $a > 0$,*

$$|M_X(t)| < \infty \quad \forall |t| < a. \tag{10.7}$$

Then

$$M_X(t) = \sum_{k=0}^{\infty} E(X^k) t^k / k! \quad \forall |t| < a.$$

Demostración. Para cualquier $n = 0, 1, ...,$

$$\begin{aligned} |\sum_{k=0}^{n}(tX)^k/k!| &\leq \sum_{k=0}^{n}|tX|^k/k! \qquad (10.8)\\ &\leq e^{|tX|} \qquad \text{(por (10.6))}\\ &\leq e^{tX} + e^{-tX}. \end{aligned}$$

Por otra parte, para $|t| < a$,

$$E(e^{tX} + e^{-tX}) = M_X(t) + M_X(-t) < \infty.$$

Luego, por (10.6), (10.8) y el Teorema de Convergencia Dominada,

$$M_X(t) = E(e^{tX}) = E\left(\sum_{k=0}^{\infty}(tX)^k/k!\right) = \sum_{k=0}^{\infty}\frac{t^k}{k!}E(X^k).$$

$\square$

La función característica

La función característica (f.c.) de X se define como

$$C_X(t) := E(e^{itX}) \ \forall t \in \mathbb{R}, \qquad (10.9)$$

donde $i = \sqrt{-1}$ es la unidad imaginaria. Observe que la f.c. se puede escribir en terminos de la f.g.m. como $C_X(t) = M_X(it)$ y que, en lugar de (10.2), tenemos

$$C_X(t) = \int_{\Omega} e^{itX}dP = \int_{\mathbb{R}} e^{itx}dF_X(x). \qquad (10.10)$$

En vista del lado derecho de esta igualdad, la f.c. de X es la **transformada de Fourier** de F_X. Por otra parte, si X es discreta o continua, entonces su f.c. se puede obtener de (10.3) y (10.4), respectivamente, sustituyendo t por it.

Nota 10.2. Para cualquier $x \in \mathbb{R}$ se tiene que

$$e^{ix} = \cos x + i \operatorname{sen} x$$

y, por lo tanto, $|e^{ix}| = (\cos^2 x + \operatorname{sen}^2 x)^{1/2} = 1$. De aquí se sigue que

$$|C_X(t)| \leq \int_{\mathbb{R}} |e^{itx}|dF_X(x) = 1 \quad \textit{para todo} \quad t \in \mathbb{R}.$$

Esto significa que la f.c. $C_{\mathrm{X}}(t)$ está definida y es uniformemente acotada para todo $t \in \mathbb{R}$. Observese que la f.g.m. $M_{\mathrm{X}}(t)$ puede **no** estar definida para algunos valores de t. (Vea Ejercicios 10.1 (b) y 10.2 (a).) También se puede ver que la f.c. es *uniformemente continua*. En efecto, para todo t y h en $\mathbb{R}$,

$$\begin{aligned}|C_X(t+h) - C_X(t)| &= |E[e^{itX}(e^{ihX} - 1)]| \\ &\leq E|e^{ihX} - 1| \to 0,\end{aligned}$$

cuando $h \to 0$, por el Teorema de Convergencia Acotada.

Nota 10.3. En el siguiente Teorema 10.4 se obtiene para la f.c. el análogo de (10.5), i.e.,

$$\frac{d}{dt^k} C_X(t)|_{t=0} = i^k E(X^k). \tag{10.11}$$

En la demostración del teorema se usa la desigualdad

$$|e^{ix(t+h)} - e^{ixt}| \leq |xh| \quad \forall x, t, h \in \mathbb{R}.$$

En particular $|e^{ix} - 1| \leq |x|$, y que se obtiene como sigue. Si $f(x) = e^{ix}$, entonces para todo a y b,

$$\begin{aligned}|f(b) - f(a)| &= |\int_a^b f'(x)dx \\ &\leq |b-a| \sup_{a \leq x \leq b} |f'(x)| = |b-a|.\end{aligned}$$

Teorema 10.4. *Sea n un entero positivo y X una v.a. en L_n (i.e. $E|X|^n < \infty$). Entonces la f.c. de X es n veces diferenciable con derivadas*

$$C_X^{(k)}(t) = E[(iX)^k e^{itX}] \text{ para } k = 0, 1, ..., n. \tag{10.12}$$

En particular, con $t = 0$ se tiene (10.11) para $k = 0, 1, ..., n$.

Demostración. Para $k = 0$ (10.12) se cumple trivialmente. Para $k = 1$, se sigue de la nota anterior y del Teorema de Convergencia Dominada que la f.c. C_X es diferenciable bajo el signo de integral y asi se obtiene (10.12) para $k = 1$. Iterando este argumento se obtiene (10.12) para $k = 2, ..., n$. □

Nota 10.5. Una pregunta importante es: dada una función $C : \mathbb{R} \to \mathbb{C}$, ¿cómo sabemos si C es la función característica de una v.a.? La respuesta se conoce como *Teorema de Bochner* (o de Bochner-Khinchin) que dice lo siguiente: La funcion C es una función característica si y sólo si C es continua, $C(0) = 1$ y C es positiva semi-definida, es decir, para cualquier entero $n \geq 1$ y cualesquiera dos conjuntos $\{t_1, ..., t_n\} \subset \mathbb{R}$ y $\{z_1, ..., z_n\} \subset \mathbb{C}$ se tiene que

$$\sum_{i,j=1}^{n} C(t_i - t_j) z_i z_j \geq 0.$$

La demostración de este teorema se puede ver, por ejemplo, en el libro de Laha & Rohatgi [**?**], Teorema 3.6.1.

El siguiente teorema establece una propiedad fundamental de las funciones características. Sin embargo, como la demostración es muy técnica (basada en la fórmula de inversión de la transformada de Fourier), la pospondremos para más adelante (ver Lema 10.16).

Teorema 10.6. *(***Propiedad de unicidad de la f.c.***)*
Dos vv. aa. tienen la misma f.d. ssi tienen la misma f.c. En otras palabras, $X \sim Y$ *ssi* $C_X(t) = C_Y(t)$ *para todo* $t \in \mathbb{R}$.

Las siguientes propiedades son fáciles de demostrar.

Proposición 10.7. (a) Si $Y = a\mathrm{X} + b$, entonces

$$C_Y(t) = e^{ibt}\, C_{\mathrm{X}}(at). \tag{10.13}$$

(b) Si $\mathrm{X}_1, \ldots, \mathrm{X}_n$ son vv.aa. independientes y $S := \mathrm{X}_1 + \cdots + \mathrm{X}_n$, entonces

$$C_S(t) = C_{\mathrm{X}_1}(t) \cdots C_{\mathrm{X}_n}(t) \quad \forall\, t \in \mathbb{R}. \tag{10.14}$$

En particular, si $\mathrm{X}_1, \ldots, \mathrm{X}_n$ son i.i.d., entonces

$$C_S(t) = [C_{\mathrm{X}_1}(t)]^n. \tag{10.15}$$

Ejemplo 10.8. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_r$ vv.aa. *independientes* y sea $S := \mathrm{X}_1 + \cdots + \mathrm{X}_r$. Demuestre:

(a) Si $\mathrm{X}_k \sim \mathrm{Poi}(\lambda_k)$ para $k = 1, \ldots, r$, entonces $S \sim \mathrm{Poi}(\lambda_1 + \cdots + \lambda_r)$.

(b) Si $\mathrm{X}_k \sim N(\mu_k, \sigma_k^2)$ para $k = 1, \ldots, r$, entonces $S \sim N(\mu_1 + \cdots + \mu_r,\ \sigma_1^2 + \cdots + \sigma_r^2)$.

Solución. (a) Por la Proposición 10.7(b),

$$C_S(t) = \prod_{k=1}^{r} C_{\mathrm{X}_k}(t) = \prod_{k=1}^{r} e^{\lambda_k(e^{it}-1)}$$

de modo que

$$C_S(t) = e^{(\lambda_1+\cdots+\lambda_r)(e^{it}-1)}.$$

Es decir, la f.c. C_S coincide con la de una variable Poisson con parámetro $\lambda = \lambda_1 + \cdots + \lambda_r$. Por lo tanto, el resultado se sigue de la propiedad de unicidad en el Teorema 10.6. La demostración de (b) es similar. $\square$

En el resto de esta sección veremos varios resultados relacionados con el concepto de convergencia en distribución de vv.aa. Recuerde que según la Definición 9.9, $\mathrm{X}_n \to \mathrm{X}$ en distribución si $F_{\mathrm{X}_n}(x) \to F_{\mathrm{X}}(x)$ para todo $x \in \mathbb{R}$ en el que la f.d. F_{X} es continua. Un criterio relativamente sencillo de usar para este tipo de convergencia es el siguiente, que ya vimos en el Teorema 9.25.

Proposición 10.9. $\mathrm{X}_n \to \mathrm{X}$ en distribución ssi $Eh(\mathrm{X}_n) \to Eh(\mathrm{X})$ para toda función $h : \mathbb{R} \to \mathbb{R}$ continua y acotada.

De hecho, aquí usaremos el siguiente criterio de convergencia en distribución basado en funciones características, llamado el **Teorema de continuidad**.

Teorema 10.10. *Sean* X *y* X_n $(n = 1, 2, \ldots)$ *vv.aa. con funciones características* C_{X} *y* C_n*, respectivamente. Entonces* $\mathrm{X}_n \to \mathrm{X}$ *en distribución ssi* $C_n(t) \to C_{\mathrm{X}}(t)$ *para todo* t *en* $\mathbb{R}$*.*

De hecho, una parte del Teorema 10.10 se sigue directamente de la Proposición 10.9. En efecto, tomando $h_1(x) = cosx$ y $h_2(x) = senx$, de la Proposicion 10.9 se sigue que si $X_n \to X$ en distribucion, entonces $Eh_i(X_n) \to Eh_i(X)$ para $i = 1, 2$. Esto implica que $C_n(\cdot)$ converge a $C_X(\cdot)$. El reciproco de esta proposicion se obtiene de la formula de inversion que veremos en el Lema 10.16.

Ejemplo 10.11. Supóngase que $\mathrm{X}_n \to \mathrm{X}$ en L_2. Demuestre que:

(a) $E\mathrm{X}_n \to E\mathrm{X}$, $E(\mathrm{X}_n^2) \to E(\mathrm{X}^2)$ y $\mathrm{Var}(\mathrm{X}_n) \to \mathrm{Var}(\mathrm{X})$;

(b) Si además $X_n \sim N(\mu_n, \sigma_n^2)$ para cada n, con $\mu_n \to \mu$ y $\sigma_n^2 \to \sigma^2 > 0$, entonces X tiene distribución $N(\mu, \sigma^2)$ y $X_n \to X$ en distribución.

Solución. (a) Por definición de convergencia en L_2, $E(X_n - X)^2 \to 0$ cuando $n \to \infty$. Por la desigualdad de Jensen o la desigualdad de Cauchy–Schwarz (Ejercicio 8.3)

$$(EX_n - EX)^2 = [E(X_n - X)]^2 \leq E(X_n - X)^2 \to 0,$$

esto implica que $EX_n \to EX$. Otra vez usando la desigualdad de Cauchy–Schwarz, $E[X \cdot (X_n - X)] \to 0$ pues

$$(E[X \cdot (X_n - X)])^2 \leq E(X^2) \cdot E(X_n - X)^2 \to 0.$$

Por lo tanto, como

$$X_n^2 - X^2 = (X_n - X)^2 + 2X \cdot (X_n - X),$$

se sigue $E(X_n^2) \to E(X^2)$. Finalmente,

$$\mathrm{Var}(X_n) = E(X_n^2) - (EX_n)^2 \to E(X^2) - (EX)^2 = \mathrm{Var}(X).$$

(b) La f.c. de X_n converge a la f.c. de X pues

$$C_{X_n}(t) = \exp(i\mu_n t - \sigma_n^2 t^2/2) \to \exp(i\mu t - \sigma^2 t^2/2) = C_X(t)$$

para todo $t \in \mathbb{R}$. Es decir, $X \sim N(\mu, \sigma^2)$ y del Teorema 10.10 concluimos que $X_n \to X$ en distribución. (Observe que esta última conclusión también se podría haber obtenido de la Proposición 9.10.) $\square$

En la Nota 10.2(b) se vió que si $E|X|^n < \infty$, entonces las derivadas de $C_X(t)$ en $t = 0$ satisfacen que $C_X^{(k)}(0) = i^k E(X^k)$ para $k = 0, 1, \ldots, n$. Además, aplicando la fórmula de Taylor a $C_X(t)$ podemos ver que

$$C_X(t) = \sum_{k=0}^{n} (it)^k E(X^k)/k! + o(t^n) \quad \forall t \in \mathbb{R}, \quad n = 1, 2, ...,$$

donde $o(t^n)$ es una función tal que $\lim_{t \to 0} \frac{o(t^n)}{t^n} = 0$.

En efecto, se sabe que

$$e^{ix} = \cos(x) + i \sin(x) = \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!}(\cos(\theta_1 x) + i \sin(\theta_2 x))$$

con $x \in \mathbb{R}$ y $\theta_1, \theta_2 \in [-1, 1]$. Así que, sustituyendo tX por x y tomando esperanzas, se obtiene

$$E\left[e^{itX}\right] = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} E(X^k) + \frac{(it)^n}{n!} \left\{E(X^n) + \epsilon_n(t)\right\},$$

donde

$$\epsilon_n(t) := E\left[X^n(\cos(\theta_1 tX) + i\sin(\theta_2 tX) - 1)\right].$$

Sin embargo, como $|\epsilon_n(t)| \leq 3E[|X|^n]$, por el teorema de convergencia dominada $\epsilon_n(t) \to 0$ cuando $t \to 0$.

En particular, para $n = 2$ tenemos

$$C_{\mathrm{X}}(t) = 1 + (it)E\mathrm{X} - t^2 E(\mathrm{X}^2)/2 + o(t^2). \tag{10.16}$$

Usaremos esta expresión para demostrar a continuación el célebre **Teorema del Límite Central** (TLC).

Teorema 10.12. *Sea $\{\mathrm{X}_n\}$ una sucesión de vv.aa. i.i.d. con media μ y variancia $\sigma^2 > 0$ finita. Para cada $n = 1, 2, \ldots$, sean S_n y $\overline{S}_n$ la suma y el promedio muestral, respectivamente, i.e.*

$$S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n \quad y \quad \overline{S}_n := \frac{1}{n} S_n.$$

Asimismo, sea

$$Y_n := \frac{S_n - ES_n}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}; \tag{10.17}$$

nótese que

$$Y_n = \frac{\overline{S}_n - \mu}{\sigma/\sqrt{n}}.$$

Entonces

$$Y_n \to Z \quad \text{en distribución}, \tag{10.18}$$

en donde $Z \sim N(0, 1)$. Equivalentemente, como

$$F_Z(x) := \mathrm{P}\{Z \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

es continua en todo $x \in \mathbb{R}$, se tiene que

$$\mathrm{P}\{Y_n \leq x\} \to F_Z(x) \quad \forall\, x \in \mathbb{R}. \tag{10.19}$$

Demostración. Sea $C_n(t)$ la f.c. de Y_n. Por el Teorema de Continuidad 10.10, para demostrar (10.19) basta verificar que

$$C_n(t) \to C_Z(t) = e^{-t^2/2} \quad \forall\, t \in \mathbb{R}. \tag{10.20}$$

Con este fin, nótese primero que

$$Y_n = \sum_{k=1}^{n} \overline{\mathrm{X}}_k/\sqrt{n}, \tag{10.21}$$

en donde $\overline{\mathrm{X}}_k := (\mathrm{X}_k - \mu)/\sigma$ es una v.a. con media 0 y varianza 1; además, por (10.16), la f.c. de $\overline{\mathrm{X}}_k$ satisface que

$$C_{\overline{\mathrm{X}}_k}(t) = 1 - t^2/2 + o(t^2) =: h(t) \quad \forall\, k = 1, 2, \ldots$$

Por lo tanto, como las $\overline{\mathrm{X}}_k$ son i.i.d., se sigue que

$$\begin{aligned} C_n(t) &= [h(t/\sqrt{n})]^n \\ &= [1 - t^2/2n + o(t^2/n)]^n. \end{aligned}$$

Finalmente, tomando el límite cuando $n \to \infty$ y usando (9.18) obtenemos (10.20). □

Nota 10.13. (a) La **desigualdad de Berry–Essen** da una estimación de la rapidez de convergencia en (10.19) bajo las siguientes condiciones. Si las vv.aa. $\mathrm{X}_1, \mathrm{X}_2, \ldots$ son i.i.d. con media cero, varianza $\sigma^2 > 0$ y tercer momento $\rho := E|\mathrm{X}_k|^3 < \infty$, entonces

$$|\mathrm{P}\{Y_n \leq x\} - F_Z(x)| \leq 3\rho/\sigma^3\sqrt{n} \quad \forall\, x \in \mathbb{R}.$$

(b) Reescríbase (10.17) como

$$Y_n = \sum_{k=1}^{n} Z_{n,k}$$

en donde $Z_{n,k} := (\mathrm{X}_k - \mu)/\sigma\sqrt{n}$, para $k = 1, \ldots, n$, es una v.a. con media cero y varianza $1/n \to 0$ cuando $n \to \infty$. En este caso se dice que las vv.aa. $Z_{n,k}$ son "asintóticamente despreciables".

(c) Se llama **Teorema de De Moivre–Laplace** al caso especial del TLC en el que las vv.aa. X_k son Bernoulli con parámetro p, i.e.

$$\mathrm{P}\{\mathrm{X}_k = 1\} = p, \quad \mathrm{P}\{\mathrm{X}_k = 0\} = 1 - p =: q.$$

En tal caso, la suma muestral $S_n \sim \mathrm{Bin}(n, p)$ de modo que (10.17) resulta

$$Y_n = (S_n - np)/\sqrt{npq}. \tag{10.22}$$

(d) Una pregunta natural es si en el TLC se puede tener convergencia en un sentido más fuerte que convergencia en distribución. La respuesta es **no**. En efecto, sea Y_n como en (10.17) pero con $\mu = 0$ (para simplificar la notación), es decir $Y_n := S_n/\sigma\sqrt{n}$, y supóngase que

$$Y_n \to \mathrm{X} \quad \text{en probabilidad.} \tag{$*$}$$

Esto implica que $Y_n \to \mathrm{X}$ en distribución y, por el TLC, $\mathrm{X} = Z \sim N(0, 1)$. Ahora, de nuevo por $(*)$,

$$Y_{2n} := \frac{S_{2n}}{\sigma\sqrt{2n}} \to \mathrm{X} \quad \text{en probabilidad,}$$

y, por lo tanto,

$$\begin{aligned} U_n \quad &:= \quad \frac{\mathrm{X}_{n+1} + \cdots + \mathrm{X}_{2n}}{\sigma\sqrt{2n}} = \frac{S_{2n}}{\sigma\sqrt{2n}} - \frac{S_n}{\sigma\sqrt{2n}} \\ &= \quad Y_{2n} - \frac{1}{\sqrt{2}} Y_n \to \left(1 - \frac{1}{\sqrt{2}}\right) \mathrm{X} \quad \text{en probabilidad.} \end{aligned}$$

Esto último se obtiene viendo que

$$P\left(\left| \left(\frac{S_{2n}}{\sigma\sqrt{2n}} - X \right) - \frac{1}{\sqrt{2}} \left(\frac{S_n}{\sigma\sqrt{n}} - X \right) \right| > \epsilon \right)$$

$$\leq P\left(\left| \frac{S_{2n}}{\sigma\sqrt{2n}} - X \right| > \epsilon/2 \right) + P\left(\left| \frac{1}{\sqrt{2}} \left(\frac{S_n}{\sigma\sqrt{n}} - X \right) \right| > \epsilon/2 \right).$$

De aquí se sigue que la sucesión

$$V_n := \frac{\mathrm{X}_{n+1} + \cdots + \mathrm{X}_{2n}}{\sigma\sqrt{n}} = \sqrt{2} U_n \to (\sqrt{2} - 1)\mathrm{X} \quad \text{en probabilidad.} \tag{$+$}$$

Para concluir, nótese que $V_n \sim Y_n$ para todo n, de modo que, por el TLC,

$$V_n \to N(0, 1) \quad \text{en distribución.}$$

Por otra parte, por (+),

$$V_n \to (\sqrt{2}-1)N(0,1) \quad \text{en distribución},$$

lo cual es una contradicción. Luego, $(*)$ **no** puede ocurrir.

Ejemplo 10.14. En una encuesta preelectoral se encuentra que el 44% de la población está a favor de un cierto candidato. Calcule la probabilidad de que en una muestra de 400 personas escogidas al azar más de la mitad estén a favor de dicho candidato.

Solución. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$, con $n = 400$, vv.aa. i.i.d. de Bernoulli con parámetro $p = 0.44$. Para cada $k = 1, \ldots, n$, $\mathrm{X}_k = 1$ ("éxito") si la k–ésima persona está a favor del candidato. Se pide calcular la probabilidad de que

$$S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n > 200.$$

Si usamos directamente la distribución de $S_n \sim \mathrm{Bin}(n,p)$, debemos calcular

$$\mathrm{P}\{S_n > 200\} = 1 - \mathrm{P}\{S_n \leq 200\} = 1 - \sum_{k=0}^{200} \binom{n}{k} p^k\, q^{n-k},$$

con $q = 1-p = 0.56$, que obviamente es un cálculo complicado. Sin embargo, usando el TLC con Y_n como en (10.22), en donde

$$E(S_n) = np = (400)(0.44) = 176 \quad \text{y} \quad \mathrm{Var}(S_n) = npq = 98.56,$$

vemos que

$$\begin{aligned}
\mathrm{P}\{S_n \leq 200\} &= \mathrm{P}\{Y_n \leq (200-176)/\sqrt{98.56}\} \\
&\approx F_Z(2.42) \qquad \text{[por (10.18)]} \\
&= 0.9922 \qquad \text{[de la "tabla normal"]}.
\end{aligned}$$

Por lo tanto $\mathrm{P}\{S_n > 200\} \approx 1 - 0.9922 = 0.0078 = 0.78\%$.

Ejemplo 10.15. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_n$, con $n = 50$, vv.aa. i.i.d. de Poisson con parámetro $\lambda = 0.03$. Calcule $\mathrm{P}\{S_n \geq 3\}$.

Solución. Nótese que $S_n \sim \mathrm{Poi}(n\lambda)$ de modo que

$$E(S_n) = \mathrm{Var}(S_n) = n\lambda = (50)(0.03) = 1.5.$$

Por lo tanto, el cálculo exacto sería $\mathrm{P}\{S_n \geq 3\} = 1 - \mathrm{P}\{S_n \leq 2\}$ con

$$\mathrm{P}\{S_n \leq 2\} = \sum_{k=0}^{2} \mathrm{P}\{S_n = k\} = e^{-1.5}[1 + 1.5 + (1.5)^2/2].$$

Por otra parte, usando el TLC con $Y_n = (S_n - n\lambda)/\sqrt{n\lambda}$ obtenemos

$$\begin{aligned} \mathrm{P}\{S_n \leq 2\} &= \mathrm{P}\{Y_n \leq (2 - 1.5)/\sqrt{1.5}\} \\ &\approx F_Z(0.5/\sqrt{1.5}) \\ &= 0.6591. \end{aligned}$$

Luego, $\mathrm{P}\{S_n \geq 3\} \approx 0.3409$.

Lema 10.16. (Fórmula de inversión de Fourier) Sea F una f.d.p. y $F(a, b] := F(b) - F(a)$ para $a < b$. Si h es la función característica de F, es decir, $h(t) := \int_{-\infty}^{\infty} e^{itx} dF(x)$, entonces

$$F(a, b] = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot h(t) dt \tag{10.23}$$

para todo a y b que sean puntos de continuidad de F. Si además h es integrable sobre $\mathbb{R}$ con respecto a la medida de Lebesgue, entonces la función

$$f(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} h(t) dt$$

es una densidad para F, es decir, f es no–negativa y

$$F(x) = \int_{-\infty}^{x} f(u) du \quad \forall\, x \in \mathbb{R}.$$

Demostración. Como $|h(t)| \leq 1$,

$$\begin{aligned} \left| \frac{e^{-ita} - e^{-itb}}{it} \cdot h(t) \right| &\leq \left| \frac{e^{-ita} - e^{-itb}}{it} \right| \\ &= \left| \int_a^b e^{-itx} dx \right| \\ &\leq b - a, \end{aligned}$$

de modo que

$$\int_{-T}^{T} \left| \frac{e^{-ita} - e^{-itb}}{it} \cdot h(t) \right| dt \le 2T(b-a) < \infty.$$

Por lo tanto, si J_T es la integral en (10.23), i.e.

$$J_T := \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot h(t) dt,$$

usando el Teorema de Fubini vemos que

$$\begin{aligned} J_T &= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot \int_{-\infty}^{\infty} e^{itx} dF(x) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt\, dF(x) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{\operatorname{sen} t(x-a) - \operatorname{sen} t(x-b)}{t} dt\, dF(x) \qquad (10.24) \end{aligned}$$

porque la función $t^{-1}\cos(ct)$ es impar así que su integral sobre $[-T, T]$ es cero.

Sea $\operatorname{sgn}(r)$ la *función signo*, definida como

$$\operatorname{sgn}(r) := \begin{cases} 1 & \text{si} \quad r > 0, \\ 0 & \text{si} \quad r = 0, \\ -1 & \text{si} \quad r < 0. \end{cases}$$

Entonces, usando la fórmula

$$\int_0^t \frac{\operatorname{sen} x}{x} dx = \int_0^t \operatorname{sen} x \left(\int_0^\infty e^{-rx} dr \right) dx$$

e intercambiando el orden de integración, se obtiene que

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{\operatorname{sen}(rt)}{t} dt = \pi \operatorname{sgn}(r);$$

dicho intercambio surge de las fórmulas

$$\int e^{-ux} \sin(rx) dx = -\frac{e^{-ux}}{u^2 + r^2} (u \sin(rx) + r \cos(rx))$$

y

$$\int \frac{du}{r^2+u^2} = \frac{1}{r}\tan^{-1}(x/r).$$

Por lo tanto, de (10.24), usando el Teorema de Convergencia Acotada tenemos:

$$\begin{aligned}
\lim_{T\to\infty} J_T &= \frac{1}{2}\int_{-\infty}^{\infty}[\operatorname{sgn}(x-a)-\operatorname{sgn}(x-b)]dF(x) \\
&= \frac{1}{2}\left[-1\times(F(a-)-F(-\infty))+1\times(F(\infty)-F(a+))\right] \\
&- \frac{1}{2}\left[-1\times(F(b-)-F(-\infty))+1\times(F(\infty)-F(b+))\right],
\end{aligned}$$

lo cual da $F(b)-F(a)$ si F es continua en a y b. Esto completa la demostración de (10.23).

Supóngase ahora que h es integrable y sea

$$f(x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}h(t)dt \quad \forall\, x\in\mathbb{R}.$$

Como h es integrable, la función f está bien definida, es continua (por el teorema de convergencia dominada) y acotada. Además, por el teorema de Fubini,

$$\begin{aligned}
\int_a^b f(x)dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t)\left[\int_a^b e^{-itx}dx\right]dt \\
&= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} h(t)\cdot\frac{e^{-ita}-e^{-itb}}{it}dt,
\end{aligned}$$

i.e.,

$$\int_a^b f(x)dx = F(a,b] = F(b)-F(a)$$

por (10.23), si F es continua en a y b, y por lo tanto para todo a,b en $\mathbb{R}$ (porque F tiene a lo más un conjunto numerable de discontinuidades, y cualquier punto en $\mathbb{R}$ es el límite por la derecha de puntos de continuidad de F). $\square$

A continuación se tiene la prueba de la Proposición 10.7(c): Unicidad de la f.c.

Demostración. Por la fórmula de inversión (10.23), la f.c. h determina F en todos los puntos de continuidad. Pero, además, cada punto en $\mathbb{R}$ es límite por la derecha de puntos de continuidad, así que h determina F en todo punto. $\square$

La distribución normal multivariada

Notación 10.17. Un vector siempre se interpretará como matriz *columna*, aunque ocasionalmente en el texto lo escribiremos como fila. Luego (de acuerdo con la definición de producto de matrices) el *producto escalar* de dos n–vectores $u = (u_1, \ldots, u_n)$ y $v = (v_1, \ldots, v_n)$ es

$$u'v = \sum_{j=1}^{n} u_j v_j = u_1 v_1 + \cdots + u_n v_n.$$

donde u' es la *transpuesta* de u. Si $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ es un n–vector aleatorio denotamos por $E\mathrm{X}$ y $\mathrm{Cov}(\mathrm{X}) := E[(\mathrm{X} - E\mathrm{X})(\mathrm{X} - E\mathrm{X})']$ su *vector medio* y su *matriz de covarianza*, respectivamente. Es decir, $E\mathrm{X}$ es el n–vector con coordenadas $E\mathrm{X}_j$ $(j = 1, \ldots, n)$ y $\mathrm{Cov}(\mathrm{X})$ es la matriz $n \times n$ con componentes.

$$\mathrm{Cov}(\mathrm{X})_{jk} = \mathrm{Cov}(\mathrm{X}_j, \mathrm{X}_k) \quad \text{para} \quad j, k = 1, \ldots, n.$$

$\square$

De acuerdo con la Definición 5.6, se dice que un n-vector aleatorio $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ es *gaussiano* (o que tiene *distribución normal multivariada* o que las vv.aa. $\mathrm{X}_1, \ldots, \mathrm{X}_n$ son *conjuntamente gaussianas*) si

$$a'\mathrm{X} = a_1 \mathrm{X}_1 + \cdots + a_n \mathrm{X}_n \tag{10.25}$$

tiene distribución normal (univariada) para todo vector $a = (a_1, \ldots, a_n)$ en $\mathbb{R}^n$. Ahora tenemos la siguiente caracterización de la distribución normal multivariada en base a la definición de **función característica** de un n–vector X:

$$C_{\mathrm{X}}(t) := E(e^{it'\mathrm{X}}) \quad \forall\, t = (t_1, \ldots, t_n) \in \mathbb{R}^n. \tag{10.26}$$

Teorema 10.18. *Un n–vector aleatorio* X *es gaussiano ssi su función característica es de la forma*

$$C_{\mathrm{X}}(t) = \exp(it'\mu - \frac{1}{2} t'Qt) \quad \forall\, t \in \mathbb{R}^n, \tag{10.27}$$

donde μ es un n–vector y Q es una matriz $n \times n$ simétrica y no–negativa definida. En este caso $\mu = E\mathrm{X}$ es el vector medio de X *y $Q =$Cov(X) es la matriz de covarianza.*

Demostración. Supóngamos que el vector aleatorio X tiene f.c. dada por (10.27). Sea $a \in \mathbb{R}^n$ un n–vector arbitrario y sea $Y := a'\mathrm{X} = a_1\mathrm{X}_1 + \cdots + a_n\mathrm{X}_n$. Para demostrar que X es gaussiano basta verificar que la f.c. de Y corresponde a la f.c. de una variable normal, vea el Ejercicio 10.2. Para este fin, tómese $u \in \mathbb{R}$ y observe que

$$C_Y(u) := E(e^{iuY}) = E(e^{i(ua)'\mathrm{X}}) = C_{\mathrm{X}}(ua),$$

es decir, por (10.27),

$$C_Y(u) = \exp[iu(a'\mu) - \frac{1}{2}u^2(a'Qa)].$$

Esto significa de $Y = a'\mathrm{X}$ es una v.a. normal $N(a'\mu, a'Qa)$.

Recíprocamente, supóngase que X es un vector gaussiano con vector medio $\mu := E\mathrm{X}$ y matriz de covariancia $Q := \mathrm{Cov}(\mathrm{X})$. Luego (por la Definición 5.6), para cada n–vector a, la v.a. $Y := a'\mathrm{X}$ es una v.a. normal con media y varianza

$$EY = a'E\mathrm{X} = a'\mu \quad \text{y} \quad \mathrm{Var}(Y) = a'Qa,$$

respectivamente. De aquí se sigue que (por (10.28)) la función característica de Y es

$$C_Y(u) = \exp[iu(a'\mu) - \frac{1}{2}u^2(a'Qa)] \quad \forall\, u \in \mathbb{R}.$$

En particular, tomando $u = 1$ vemos que

$$C_Y(1) = C_{a'\mathrm{X}}(1) = E(e^{ia'\mathrm{X}}) = C_{\mathrm{X}}(a)$$

y se tiene (10.27) con $t = a$. □

Como consecuencia inmediata del Teorema 10.18 se obtiene lo siguiente (que se pide demostrar en el Ejercicio 10.10.)

Corolario 10.19. *Si* $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *son vv.aa.* $N(\mu_j, \sigma_j^2)(j = 1, \ldots, n)$ *independientes, entonces el vector* $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *es gaussiano con vector medio* $\mu = (\mu_1, \ldots, \mu_n)$ *y matriz de covarianza* $Cov(\mathrm{X}) = diag(\sigma_1^2, \ldots, \sigma_n^2)$. *Recíprocamente, si* X *es un vector gaussiano cuya matriz de covarianza es diagonal, entonces las componentes de* X *son variables normales independientes.*

Ejercicios § 10

10.1. En cada uno de los casos siguientes verifique que la f.g.m. y la f.c. de X tienen el valor que se indica.

(a) $\mathrm{X} \sim \mathrm{Bin}(n, p)$, $q := 1 - p$. Para todo $t \in \mathbb{R}$:

$$M_{\mathrm{X}}(t) = (pe^t + q)^n, \quad C_{\mathrm{X}}(t) = (pe^{it} + q)^n.$$

(b) $\mathrm{X} \sim \mathrm{Geo}(p)$, $q := 1 - p$.

$$\begin{aligned} M_{\mathrm{X}}(t) &= p/(1 - qe^t) \quad \text{si} \quad qe^t < 1, \\ C_{\mathrm{X}}(t) &= p/(1 - qe^{it}) \quad \forall\, t \in \mathbb{R}. \end{aligned}$$

(c) $\mathrm{X} \sim \mathrm{Poi}(\lambda)$. Para todo $t \in \mathbb{R}$:

$$M_{\mathrm{X}}(t) = e^{\lambda(e^t - 1)}, \quad C_{\mathrm{X}}(t) = e^{\lambda(e^{it} - 1)}.$$

10.2. Repita el problema 10.1 para los casos siguientes:

(a) $\mathrm{X} \sim \mathrm{Exp}(\lambda)$.

$$\begin{aligned} M_{\mathrm{X}}(t) &= \lambda/(\lambda - t) \quad \text{si } \ t < \lambda, \\ C_{\mathrm{X}}(t) &= \lambda/(\lambda - it) \quad \forall\, t \in \mathbb{R}. \end{aligned}$$

(b) $\mathrm{X} \sim N(\mu, \sigma^2)$. Para todo $t \in \mathbb{R}$:

$$M_{\mathrm{X}}(t) = \exp(\mu t + \sigma^2 t^2/2), \quad C_{\mathrm{X}}(t) = \exp(i\mu t - \sigma^2 t^2/2). \tag{10.28}$$

En particular, si $\mathrm{X} \sim N(0, 1)$ entonces

$$M_{\mathrm{X}}(t) = e^{t^2/2} \quad \text{y} \quad C_{\mathrm{X}}(t) = e^{-t^2/2} \quad \forall\, t \in \mathbb{R}. \tag{10.29}$$

10.3. Sea X una v.a. discreta con valores en un subconjunto de los enteros no negativos, $\{0, 1, \ldots\}$, y densidad $f(k) := \mathrm{P}(\mathrm{X} = k)$. La **función generadora de probabilidad** (f.g.p.) de X es la función

$$G_{\mathrm{X}}(t) := E(t^{\mathrm{X}}) = \sum_k t^k\, f(k) \quad \text{para} \quad |t| \leq 1.$$

En cada uno de los casos siguientes demuestre que la f.g.p. tiene el valor indicado:

(a) $\mathrm{X} \sim \mathrm{Bin}(n,p)$. $G_{\mathrm{X}}(t) = (pt+q)^n$ con $q := 1-p$.

(b) $\mathrm{X} \sim \mathrm{Geo}(p)$. $G_{\mathrm{X}}(t) = p/(1-tq)$ con $q := 1-p$

(c) $\mathrm{X} \sim \mathrm{Poi}(\lambda)$. $G_{\mathrm{X}}(t) = e^{\lambda(t-1)}$.

Observe que $M_{\mathrm{X}}(t) = G_{\mathrm{X}}(e^t)$ y $C_{\mathrm{X}}(t) = M_{\mathrm{X}}(it) = G_{\mathrm{X}}(e^{it})$.

10.4. Sean X y Y dos vv.aa. discretas como el Ejercicio 10.3. Demuestre que X y Y tienen la misma distribución ssi $G_{\mathrm{X}}(\cdot) = G_Y(\cdot)$.

10.5. Sean $\mathrm{X}_1, \ldots, \mathrm{X}_r$ vv.aa. discretas como en el Ejercicio 10.3 y, además, son independientes. Sea $S := \mathrm{X}_1 + \cdots \mathrm{X}_n$. Demuestre:

(a) $G_S(t) = G_{\mathrm{X}_1}(t) \cdots G_{\mathrm{X}_r}(t)$.

(b) Si $\mathrm{X}_k \sim \mathrm{Bin}(n_k, p)$ para $k = 1, \ldots, r$ entonces $S \sim \mathrm{Bin}(n_1 + \cdots + n_r, p)$. (*Sugerencia:* use el inciso (a) y los Ejercicios 10.3(b) y 10.4.)

10.6. Supóngase que $\mathrm{X} \sim N(0, \sigma^2)$. Use (10.28) para verificar que

$$\begin{aligned} E(\mathrm{X}^k) &= 0 \quad \text{si} \quad k \quad \text{es impar,} \\ &= k!\sigma^k / 2^{k/2}(k/2)! \quad \text{si} \quad k \quad \text{es par.} \end{aligned}$$

10.7. Sea X una v.a. continua con densidad $f(x) = \frac{1}{2}e^{-|x|}$ para todo $x \in \mathbb{R}$.

(a) Demuestre que la f.g.m. de X es $M(t) = 1/(1-t^2)$ para $-1 < t < 1$.

(b) Use $M(t)$ para demostrar que $E(\mathrm{X}^n) = 0$ si n es un entero positivo impar, y para encontrar una expresión de $E(\mathrm{X}^n)$ cuando n es par.

10.8. Para cada $n = 1, 2, \ldots$, sea X_n una v.a. con f.c. $C_n(t)$. Demuestre que los siguientes tres enunciados son equivalentes:

(a) $\mathrm{X}_n \to 0$ en probabilidad.

(b) $\mathrm{X}_n \to 0$ en distribución.

(c) $C_n(t) \to 1$ para todo $t \in \mathbb{R}$.

10.9. Se sabe que el 5% de las computadoras fabricadas por una cierta empresa son defectuosas. Si se seleccionan al azar 100 computadoras de dicha empresa, calcule la probabilidad de que a lo más una sea defectuosa usando:

(a) la distribución binomial,

(b) la aproximación de Poisson a la distribución binomial (ver 9.15), y

(c) el TLC.

10.10. Demuestre el Corolario 10.19.

10.11. Sea X un n–vector gaussiano y Y un m–vector gaussiano. Demuestre que si X y Y son independientes, entonces (X, Y) es un $(n+m)$–vector gaussiano.

10.12. (Ejemplo de una distribución bivariada que **no** es gaussiana bivariada pero cuyas marginales sí son gaussianas — vea también el Ejemplo 5.7.) Sea X una v.a. $N(0,1)$ y sea Y la v.a. definida, para algún $a > 0$, como

$$Y := \mathrm{X} \cdot I_{\{|\mathrm{X}| \leq a\}} - \mathrm{X} \cdot I_{\{|\mathrm{X}| > a\}}.$$

Demuestre que $Y \sim N(0,1)$, pero $\mathrm{X} + Y$ **no** es normal. Por lo tanto, cada v.a. X y Y es gaussiana, pero el vector (X, Y) no es gaussiano.

11 Esperanza condicional

Contenido. Densidad y distribución condicional, esperanza condicional dada una σ–álgebra, estimador en la media cuadrática, estimador lineal.

Sean X y Y vv.aa. discretas con densidad conjunta $f(x,y)$, i.e.,

$$f(x,y) := \mathrm{P}\{\mathrm{X}=x,\ Y=y\},$$

y sean

$$f_{\mathrm{X}}(x) := \mathrm{P}(\mathrm{X}=x) = \sum_y f(x,y) \quad \text{y} \quad f_Y(y) := \mathrm{P}(Y=y) = \sum_x f(x,y)$$

las densidades marginales de X y Y, respectivamente. Definimos la **densidad condicional de Y dado que** $\mathrm{X}=x$ como

$$\begin{aligned} f(y|x) &:= f(x,y)/f_{\mathrm{X}}(x) && \text{si} \quad f_{\mathrm{X}}(x)>0, \\ &:= 0 && \text{si} \quad f_{\mathrm{X}}(x)=0. \end{aligned} \tag{11.1}$$

Asimismo, la **distribución condicional** de Y dado que $\mathrm{X}=x$ es

$$F(y|x) := \mathrm{P}(Y \le y|\mathrm{X}=x) = \sum_{y' \le y} f(y'|x). \tag{11.2}$$

Si además Y está en L_1, definimos la **esperanza condicional** de Y dado que $\mathrm{X}=x$ como

$$E(Y|\mathrm{X}=x) := \sum_y y\, f(y|x). \tag{11.3}$$

Observamos que $h(x) := E(Y|\mathrm{X}=x)$ como función de x cumple dos cosas:
1) $h:\mathbb{R}\to\mathbb{R}$ es medible, y
2) para todo $A \in \sigma\{X\}$ se tiene que

$$\begin{aligned} E[YI_A] &= \sum_{y,x} yI_A(x)f(x,y) = \sum_{x\in A}\sum_y yf(y|x)f(x) \\ &= \sum_{x\in A} h(x)f(x) = E(h(X)I_A). \end{aligned}$$

Así, vemos entonces que se cumple que
(a) $E(Y|X) := h(X)$ es medible c.r.a $\sigma\{X\}$, y
(b) para todo $A \in \sigma\{X\}$

$$E[YI_A] = E[E(Y|X)I_A]. \tag{11.4}$$

La abreviación c.r.a se referiere a *con respecto a*.

Ejemplo 11.1. Sean $X_1, X_2, \ldots$ i.i.d. con valores en $\{0, 1, \ldots\}$. Sea $S_0 := 0$ y $S_n := X_1 + \cdots + X_n$ para $n \geq 1$. Sea N una v.a. con valores en $\{0, 1, \ldots\}$ e independiente de $\{X_j\}$. Considere la **suma aleatoria** $S_N := X_1 + \cdots + X_N$. Demuestre que:

(a) $P\{S_N = x\} = \sum_{n=0}^{\infty} P\{N = n\}\, P\{S_n = x\}$.

(b) Si además las X_j y N tienen medias finitas μ y EN, respectivamente, entonces $ES_N = \mu \cdot EN$.

Solución. (a) Nótese que

$$P\{S_N = x|N = n\} = P\{S_n = x|N = n\} = P\{S_n = x\}. \tag{11.5}$$

Además, por la ley de probabilidad total,

$$\begin{aligned} P\{S_N = x\} &= \sum_{n=0}^{\infty} P\{S_N = x,\ N = n\} \\ &= \sum_{n=0}^{\infty} P\{N = n\} P\{S_N = x|N = n\} \\ &= \sum_{n=0}^{\infty} P\{N = n\} P\{S_n = x\} \quad \text{[por (11.5)].} \end{aligned}$$

Esto completa la demostración de (a).

(b) Primero observe que

$$E(S_n) = \sum_x x\, P\{S_n = x\} = n\mu, \quad \text{y} \quad EN = \sum_n n\, P\{N = n\}. \tag{11.6}$$

Además, por definición de esperanza,

$$\begin{aligned} ES_N &:= \sum_x x\, P\{S_N = x\} \\ &= \sum_x x \sum_n P\{N = n\} P\{S_n = x\} \qquad \text{[por (a)]} \\ &= \sum_n P\{N = n\} \sum_x x\, P\{S_n = x\} \\ &= \mu \cdot \sum_n n\, P\{N = n\} \qquad \text{[por (11.6)]} \\ &= \mu \cdot EN \qquad \text{[por (11.6)].} \end{aligned}$$

Para vv.aa. continuas tenemos definiciones análogas. Es decir, sean X y Y vv.aa. continuas con densidad conjunta $f(x,y)$ y sea

$$f_{\mathrm{X}}(x) := \int_{-\infty}^{\infty} f(x,y)dy$$

la densidad marginal de X. Entonces definimos la **densidad condicional** de Y dado que $\mathrm{X} = x$ como

$$f(y|x) := f(x,y)/f_{\mathrm{X}}(x) \quad \text{si} \quad f_{\mathrm{X}}(x) > 0 \tag{11.7}$$

y $f(y|x) := 0$ si $f_{\mathrm{X}}(x) = 0$. Además, tenemos la **distribución condicional**

$$F(y|x) := \mathrm{P}(Y \leq y|\mathrm{X} = x) = \int_{-\infty}^{y} f(y'|x)dy' \tag{11.8}$$

y, para Y en L_1, la **esperanza condicional**

$$E(Y|\mathrm{X} = x) := \int_{-\infty}^{\infty} y\ f(y|x)dy. \tag{11.9}$$

Así como en el caso discreto, podemos también verificar que $h(x) := E(Y|\mathrm{X} = x)$ es medible, en cuyo caso, tomando $A \in \sigma\{X\}$, se llega a la identidad

$$E(YI_A) = E(h(X)I_A).$$

Así, definiendo $E(Y|X) := h(X)$ se cumpliría que
(a) $E(Y|X)$ es medible c.r.a $\sigma\{X\}$, y
(b) para todo $A \in \sigma\{X\}$

$$E[YI_A] = E[E(Y|X)I_A]. \tag{11.10}$$

En ambos casos, discreto y continuo, se puede ver que la densidad condicional $f(\cdot|x)$ es una densidad de probabilidad si $f_{\mathrm{X}}(x) > 0$. Por ejemplo, en el caso discreto (1) tenemos $f(\cdot|x) \geq 0$ y, además,

$$\sum_y f(y|x) = \frac{1}{f_{\mathrm{X}}(x)} \sum_y f(x,y) = \frac{1}{f_{\mathrm{X}}(x)} f_{\mathrm{X}}(x) = 1.$$

También tenemos el siguiente resultado cuyo inciso (a) se puede interpretar como una versión de la **ley de la probabilidad total** en el Teorema 3.5(a).

Proposición 11.2. (a) Si X y Y son vv.aa. discretas, entonces

$$f_Y(y) = \sum_x f(y|x) f_{\mathrm{X}}(x), \tag{11.11}$$

y si son continuas, entonces

$$f_Y(y) = \int_{-\infty}^{\infty} f(y|x)\ f_{\mathrm{X}}(x) dx. \tag{11.12}$$

(b) Si X y Y son *independientes*, entonces

$$(b_1)\ f(y|x) = f_Y(y), \quad \text{y} \quad (b_2)\ E(Y|\mathrm{X}=x) = EY \quad (\text{para } Y \in L_1). \tag{11.13}$$

Demostración. (a) Por (11.1), el lado derecho de (11.11) resulta

$$\sum_x f(y|x) f_{\mathrm{X}}(x) = \sum_x f(x,y) = f_Y(y).$$

Análogamente, (11.12) se obtiene de (11.7).

(b) Por el Teorema 5.2, si X y Y son independientes, entonces

$$f(x,y) = f_{\mathrm{X}}(x) f_Y(y).$$

Por lo tanto, de (11.1) se obtiene la condición (b_1) en (11.13). Asimismo (b_2) se sigue de (b_1) y de la definición de esperanza condicional. En efecto, en el caso discreto (11.3) obtenemos

$$E(Y|\mathrm{X}=x) := \sum_y y\ f(y|x) = \sum_y y\ f_Y(y) = EY.$$

En forma similar, en el caso continuo (b_2) se obtiene de (b_1) y (11.9). □

Ejemplo 11.3. Sean X y Y las vv.aa. en el Ejemplo 5.4; es decir, el vector (X, Y) tiene densidad conjunta (normal bivariada estándar)

$$f(x,y) = \frac{1}{2\pi r} e^{-(x^2 - 2\rho xy + y^2)/2r^2} \quad \forall\ (x,y) \in \mathbb{R}^2,$$

con $|\rho| < 1$ y $r := (1-\rho^2)^{1/2}$. Por el inciso (a) de dicho ejemplo, cada una de las vv.aa. X y Y tiene densidad normal estándar. En particular, la densidad marginal de X es

$$f_{\mathrm{X}}(x) = (2\pi)^{-1/2} e^{-x^2/2} \quad \forall\ x \in \mathbb{R}.$$

Por lo tanto, por (11.7), la densidad condicional de Y dado que $\mathrm{X} = x$ es

$$f(y|x) := f(x,y)/f_{\mathrm{X}}(x) = (2\pi r^2)^{-1/2} e^{-(y-\rho x)^2/2r^2} \tag{11.14}$$

la cual es una densidad normal $N(\rho x, r^2)$. En otras palabras, la densidad marginal $f_Y(y)$ es $N(0,1)$, pero la *densidad condicional de* Y *dado* $\mathrm{X} = x$ es $N(\rho x, r^2)$. De aquí se sigue que

$$E(Y|\mathrm{X} = x) = \rho x \quad \forall x \in \mathbb{R} \tag{11.15}$$

y, en consecuencia, la esperanza condicional $E(Y|\mathrm{X})$ es la v.a.

$$E(Y|\mathrm{X}) = \rho \mathrm{X}. \tag{11.16}$$

Del inciso (b) del Ejemplo 5.4 recuérdese que X y Y son independientes ssi $\rho = 0$, en cuyo caso $r = 1$. Por lo tanto, cuando X y Y son independientes (11.14) se reduce a la densidad normal estándar $N(0,1)$, tal como se indica en la Proposición 11.2(b), mientras que [como en (11.14) ó (11.11) (b_2)] (11.15) y (11.16) se reducen a

$$E(Y|\mathrm{X}) = EY = 0 \quad .$$

Por analogía con la ley de la probabilidad total en el Teorema 3.5(a), a la expresión (11.17) en el siguiente teorema se le conoce como **ley de la esperanza total** — en algunos textos se le llama la **propiedad de la esperanza iterada**.

Teorema 11.4. *Sean X y $Y \in L_1$ vv.aa. discreta o absolutamente continua. Entonces*

(a)

$$EY = E[E(Y|\mathrm{X})]. \tag{11.17}$$

En particular,

$$EY = \sum_x E(Y|\mathrm{X} = x) f_{\mathrm{X}}(x) \quad \textit{si} \quad \mathrm{X} \quad \textit{es discreta,} \tag{11.18}$$

$$EY = \int_{-\infty}^{\infty} E(Y|\mathrm{X} = x) f_{\mathrm{X}}(x) dx \quad \textit{si} \quad \mathrm{X} \quad \textit{es continua.} \tag{11.19}$$

(b) *Supóngase que, además,* X *es una v.a. y* $g: \mathbb{R} \to \mathbb{R}$ *una función de Borel tal que* Y *y* $Y \cdot g(\mathrm{X})$ *están en* L_1. *Entonces*

$$E[Yg(\mathrm{X})|\mathrm{X}=x] = g(x)E(Y|\mathrm{X}=x) \tag{11.20}$$

y, por lo tanto,

$$E[Yg(\mathrm{X})|\mathrm{X}] = g(\mathrm{X})E(Y|\mathrm{X}). \tag{11.21}$$

En particular, si $Y = 1$ *se sigue de (11.20) y (11.21) que*

$$E[g(\mathrm{X})|\mathrm{X}=x] = g(x) \quad y \quad E[g(\mathrm{X})|\mathrm{X}] = g(\mathrm{X}) \tag{11.22}$$

Demostración. (a) Supóngase que X es una v.a. discreta. Luego, usando (11.11) y la definición de esperanza obtenemos que

$$EY := \sum_y y\ f_Y(y) = \sum_y\ y\ \sum_x f(y|x)f_{\mathrm{X}}(x).$$

Luego, intercambiando las sumatorias (lo cual es válido porque Y está en L_1),

$$EY = \sum_x [\sum_y y\ f(y|x)]f_{\mathrm{X}}(x) = \sum_x E(Y|\mathrm{X}=x)f_{\mathrm{X}}(x).$$

Esto da (11.18) y, por lo tanto, (11.17) si X es discreta. Cuando X es continua, la demostración de (11.19) es similar, usando ahora (11.12).

(b) En el caso discreto, usando (11.3) el lado derecho de (11.20) resulta

$$\begin{aligned} g(x)E(Y|\mathrm{X}=x) &= g(x)\sum_y y\ f(y|x) \\ &= \sum_y y\ g(x)f(y|x) \\ &= E[Y\ g(x)|\mathrm{X}=x] \qquad \text{[por (11.3)]} \\ &= E[Y\ g(\mathrm{X})|\mathrm{X}=x], \end{aligned}$$

lo cual demuestra (11.20). El caso continuo se demuestra en forma similar. □

Ejemplo 11.5. Una máquina produce un número aleatorio N de artículos, en donde $N \sim \mathrm{Poi}(\lambda)$. Cada artículo puede ser defectuoso con probabilidad p $(0 < p < 1)$ independientemente de los otros artículos e independiente de

N. Si X es el número total de artículos defectuosos, calcule EX.
Solución. Por (11.17) ó (11.18).

$$EX = E[E(X|N)] = \sum_{n=0}^{\infty} E(X|N=n)P(N=n),$$

en donde $P(N=n) = e^{-\lambda}\lambda^n/n!$ para $n = 0, 1, \ldots$. Por otra parte, como

$$P(X=k|N=n) = \binom{n}{k} p^k q^{n-k} \quad \forall\, k = 0, 1, \ldots, n,$$

en donde $q := 1-p$, vemos que

$$E(X|N=n) = \sum_{k=0}^{n} k\, P(X=k|N=n) = n\,p \quad \text{(explique).}$$

Por lo tanto, $E(X|N) = Np$ de manera que $EX = p \cdot EN = p \cdot \lambda$. □

Ejemplo 11.6. Sean X y Y vv.aa. independientes que tienen distribución Geo(p_1) y Geo(p_2), respectivamente. Calcule $P\{X \geq Y\}$.
Solución. Por probabilidad total

$$P\{X \geq Y\} = \sum_{n=0}^{\infty} P\{X \geq Y|Y=n\}P\{Y=n\},$$

en donde $P\{Y=n\} = p_2 q_2^n$ para todo $n = 0, 1, \ldots$, con $q_2 := 1-p_2$. Además,

$$\begin{aligned} P\{X \geq Y|Y=n\} &= P\{X \geq n|Y=n\} \\ &= P\{X \geq n\} \qquad \text{(por independencia)} \\ &= \sum_{k=n}^{\infty} P\{X=k\} \\ &= q_1^n \qquad \text{con} \quad q_1 := 1-p_1. \end{aligned}$$

Por lo tanto

$$P\{X \geq Y\} = \sum_{n=0}^{\infty} q_1^n p_2 q_2^n = p_2/(1-q_1 q_2). \qquad \square$$

Tomando $A := \{\mathrm{X} \geq Y\}$ en el Ejemplo 11.6, vemos que se hizo un cálculo del siguiente tipo (ver (22)–(24) más adelante)

$$P(A) = E(E(I_A|X)).$$

Para establecer una definición general de esperanza condicional, tenemos el siguiente lema con respecto a la propiedad (b) mencionada en (11.4) y en (11.10); dejamos la demostración como ejercicio.

Lema 11.7. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $\mathcal{G}$ una $\sigma-$álgebra contenida en $\mathcal{F}$. Si X es medible c.r.a $\mathcal{G}$ y

$$(\forall A \in \mathcal{G}) \int_A X dP = 0,$$

entonces $X = 0$ c.s.

Nos gustaría pues definir la esperanza condicional en el caso genera, para lo cual el siguiente resultado es clave.

Teorema 11.8. ***(Esperanza condicional dada una $\boldsymbol{\sigma}$–álgebra (Radon–Nikodym))***
Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y $\mathcal{G} \subset \mathcal{F}$ una sub–σ–álgebra de $\mathcal{F}$. Si $\mathrm{X} : \Omega \to \mathbb{R}$ es una v.a. en L_1, entonces existe una única v.a. $\widehat{\mathrm{X}}$ tal que

(a) $\widehat{\mathrm{X}}$ *es medible con respecto a $\mathcal{G}$, y*

(b) $E(\widehat{\mathrm{X}} I_B) = E(\mathrm{X} I_B) \quad \forall B \in \mathcal{G}$.

Esta v.a. es única c.s., en el sentido de que si Z es otra v.a. que satisface (a) y (b), entonces $\widehat{\mathrm{X}} = Z$ c.s., es decir, $\mathrm{P}\{\widehat{\mathrm{X}} = Z\} = 1$.

Nota 11.9. El Teorema 11.8 se obtiene directamente del Teorema de Radon–Nikodym. Este último afima lo siguiente. Sean μ y ν dos medidas $\sigma-$finitas sobre un mismo espacio medible $(\Omega, \mathcal{F})$. Si ν es absolutamente continua con respecto a μ, es decir, $\nu(A) = 0$ cuando $\mu(A) = 0$. Entonces existe una única función medible f tal que para todo $A \in \mathcal{F}$

$$\nu(A) = \int_A f d\mu.$$

Así, construyendo la medida

$$\nu(B) := E(\mathrm{X} I_B) = \int_B \mathrm{X}\, d\mathrm{P}, \quad B \in \mathcal{G},$$

se llega al teorema anterior.

En vista de lo comentado para variables discretas y absolutamente continuas, tenemos la siguiente definición:

Definición 11.10. Sean Y una v.a. sobre $(\Omega, \mathcal{F}, \mathrm{P})$, con Y en L_1 y sea $\mathcal{G}$ $\boldsymbol{\sigma}$–álgebra contenida en $\mathcal{F}$. Definimos la **esperanza condicional de Y dada** $\mathcal{G}$ como la v.a. $E(Y|\mathcal{G})$

(a) $E(Y|\mathcal{G})$ es medible con respecto a $\mathcal{G}$, y

(b) $E[E(Y|\mathcal{G}) \cdot I_A] = E(Y \cdot I_A)$ para todo $A \in \mathcal{G}$.

A continuacion veremos algunos casos particulares de la Definicion 11.10.

Nota 11.11. Identificaremos vv.aa. que son iguales c.s. En otras palabras, si $\mathrm{X} = Y$ c.s. entonces escribiremos simplemente $\mathrm{X} = Y$.

Nota 11.12. Casos particulares.
(a) Si $\mathcal{G} = \{\phi, \Omega\}$ es la σ–álgebra trivial, entonces

$$E(\mathrm{X}|\mathcal{G}) = E\mathrm{X}.$$

(b) Sea $\sigma\{Y\} \equiv Y^{-1}(\mathcal{B}) := \{Y^{-1}(B) | B \in \mathcal{B}\}$ la σ–álgebra generada por la v.a. Y, en donde $\mathcal{B} \equiv \mathcal{B}(\mathbb{R})$; vea el Ejercicio 4.6. Entonces, si $\mathcal{G} = \sigma\{Y\}$, en lugar de escribir $E(\mathrm{X}|\mathcal{G})$ escribimos $E(\mathrm{X}|Y)$, i.e.

$$E(\mathrm{X}|\sigma\{Y\}) \equiv E(\mathrm{X}|Y). \tag{11.23}$$

(c) La σ–álgebra generada (o inducida) por una familia $\{Y_i,\ i \in I\}$ de vv.aa. se define como la mínima σ–álgebra que contiene a $\cup_{i\in I}\sigma\{Y_i\}$, es decir,

$$\sigma\{Y_i,\ i \in I\} := \sigma\{\bigcup_{i\in I} \sigma\{Y_i\}\}.$$

Si $\mathcal{G} = \sigma\{\mathrm{X}_i,\ i \in I\}$ entonces [como en (11.23)] escribimos

$$E(\mathrm{X}|\sigma\{Y_i,\ i \in i\}) \equiv E(\mathrm{X}|Y_i,\ i \in I).$$

Por ejemplo, $E(\mathrm{X}|\sigma\{Y_1, \ldots, Y_n\}) \equiv E(\mathrm{X}|Y_1, \ldots, Y_n)$.

(d) Dado un evento $A \in \mathcal{F}$ y una v.a. X, definimos la **probabilidad condicional de A dado que** $\mathrm{X} = x$ como

$$\mathrm{P}(A|\mathrm{X} = x) := E(I_A|\mathrm{X} = x), \tag{11.24}$$

en donde I_A es la función indicadora de A. Asimismo, la **probabilidad condicional de A dada la v.a.** X es la v.a.

$$\mathrm{P}(A|\mathrm{X}) := E(I_A|\mathrm{X}). \tag{11.25}$$

Por supuesto, los resultados para esperanzas condicionales también son válidos para probabilidades condicionales. Por ejemplo, como $\mathrm{P}(A) = EI_A$, tenemos que

$$\mathrm{P}(A) = E[\mathrm{P}(A|\mathrm{X})] \tag{11.26}$$

para cualquier v.a. X. También podemos definir $E(\mathrm{X}|A) := E(\mathrm{X}|I_A)$. Esto se usó en el Ejemplo 11.6.

Más generalmente, la **probabilidad condicional** de un evento $A \in \mathcal{F}$ dada la sub–σ–álgebra $\mathcal{G}$ de $\mathcal{F}$ es

$$\mathrm{P}(A|\mathcal{G}) := E(I_A|\mathcal{G}).$$

El siguiente teorema establece, en particular, que $E(\mathrm{X}|Y)$ es una función de Y.

Teorema 11.13. *Sean* X *y* Y *vv.aa.*

(a) *Si* $h : \mathbb{R} \to \mathbb{R}$ *es una función de Borel tal que* $\mathrm{X} = h(Y)$, *entonces* $\sigma\{\mathrm{X}\} \subset \sigma\{Y\}$.

(b) *Recíprocamente, si* $\sigma\{\mathrm{X}\} \subset \sigma\{Y\}$, *entonces existe una función de Borel* h *tal que* $\mathrm{X} = h(Y)$.

(c) *Si* $\mathrm{X} \in L_1$, *entonces existe una función de Borel* h *tal que* $E(\mathrm{X}|Y) = h(Y)$.

Demostración. El inciso (a) se vió en el Ejercicio 4.7.

(b) Supóngase que X es una función indicadora, digamos $\mathrm{X} = I_C$, entonces C pertenece a $\sigma\{Y\}$. Así, $C = Y^{-1}(A)$ para algún conjunto de Borel A. Sea $h := I_A$. Entonces $h \circ Y = I_{\{Y \in A\}} = I_C = \mathrm{X}$. Análogamente, si $\mathrm{X} = \sum_{k=1}^{n} x_k I_{C_k}$ con $C_k \in \sigma(Y)$ es una función simple y tomamos $I_{C_k} = h_k \circ Y$ como antes, entonces $\mathrm{X} = h \circ Y$ con $h = \sum_{k=1}^{n} x_k h_k$. En general, sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ funciones simples medibles c.r.a $\sigma\{X\}$ tales que $\mathrm{X}_n \to \mathrm{X}$, y tómese $\mathrm{X}_n = h_n \circ Y$. Sea $h := \lim h_n$ donde este límite existe y $h := 0$ en caso contrario. Entonces

$$\mathrm{X}(\omega) = \lim \mathrm{X}_n(\omega) = \lim h_n(Y(\omega)) = h(Y(\omega)).$$

Esto demuestra (b).

La parte (c) del teorema se sigue de (b) y del hecho de que, por el Teorema 11.8(a), la esperanza condicional $\widehat{\mathrm{X}} = E(\mathrm{X}|Y)$ es medible con respecto a $\mathcal{G} := \sigma\{Y\}$; luego, $\sigma\{\widehat{\mathrm{X}}\} \subset \sigma\{Y\}$. □

Teorema 11.14. *Sean* X *y* Y *vv.aa. en* L_1 *y* $\mathcal{G}$ *una sub–σ–álgebra de* $\mathcal{F}$. *Entonces:*

(a) $E(a\mathrm{X} + bY|\mathcal{G}) = aE(\mathrm{X}|\mathcal{G}) + bE(Y|\mathcal{G}) \quad \forall\, a, b \in \mathbb{R}$.

(b) **Propiedad de la esperanza iterada:** $E[E(\mathrm{X}|\mathcal{G})] = E\mathrm{X}$. *(Compare con (11.17).)*

(c) *Si* X *es* $\mathcal{G}$*–medible, entonces* $E(\mathrm{X}Y|\mathcal{G}) = \mathrm{X} \cdot E(Y|\mathcal{G})$. *(Compare con (11.21).)*

(d) *Si* X *es* $\mathcal{G}$*–medible, entonces* $E(\mathrm{X}|\mathcal{G}) = \mathrm{X}$*; tome* $Y \equiv 1$ *en (c). (Compare con (11.22).)*

(e) *Si* X *es independiente de* $\mathcal{G}$*, entonces* $E(\mathrm{X}|\mathcal{G}) = E\mathrm{X}$.

(f) *Si* $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$*, entonces*

$$E(\mathrm{X}|\mathcal{G}_1) = E[E(\mathrm{X}|\mathcal{G}_1)|\mathcal{G}_2] = E[E(\mathrm{X}|\mathcal{G}_2)|\mathcal{G}_1]. \tag{11.27}$$

(g) **Desigualdad de Jensen.** *Si* $h : \mathbb{R} \to \mathbb{R}$ *es convexa y* $h(\mathrm{X}) \in L_1$*, entonces* $h[E(\mathrm{X}|\mathcal{G})] \leq E[h(\mathrm{X})|\mathcal{G}]$*. En particular,* $[E(\mathrm{X}|\mathcal{G})]^2 \leq E(\mathrm{X}^2|\mathcal{G})$ *si* $\mathrm{X} \in L_2$.

La demostración de los incisos (a)–(f) del Teorema 11.14 se sigue directamente de la definición de esperanza condicional. La demostración de (g) se obtiene del *Teorema de la línea de soporte* que dice lo siguiente (vea el libro de Ash (1972), Teorema 7.3.4): si $h : \mathbb{R} \to \mathbb{R}$ es una función convexa, entonces existen sucesiones de números reales a_n, b_n tales que $h(x) = \sup_n(a_n x + b_n)$ para todo $x \in \mathbb{R}$. Esto significa que $h(\mathrm{X}) \geq a_n \mathrm{X} + b_n$ para todo n, de modo que

$$E[h(\mathrm{X})|\mathcal{G}] \geq a_n E(\mathrm{X}|\mathcal{G}) + b_n \quad \text{c.s.}$$

Tomando el sup sobre n se obtiene (g).

Por otra parte, usando el Teorema 11.14 es fácil dar una interpretación de $E(\mathrm{X}|\mathcal{G})$ como el "mejor estimador" de X en el siguiente sentido.

Definición 11.15. Considérese el espacio vectorial $L_2 := L_2(\Omega, \mathcal{F}, \mathrm{P})$ con la función de **distancia**

$$\| \mathrm{X} - Y \| := [E(\mathrm{X} - Y)^2]^{1/2} \quad \forall\, \mathrm{X},\ Y \in L_2.$$

(Vea el Ejercicio 11.6, y recuerde la *Nota* 11.11.) Sea L^o un subespacio de L_2 y $\mathrm{X} \in L_2$. Decimos que una v.a. $Z \in L^o$ es la **proyección** de X sobre L^o si

(a) $Z \in L^o$, y

(b) $\| \mathrm{X} - Z \| = \min\{\| \mathrm{X} - Y \| \,|Y \in L^o\}$.

En este caso también se dice que Z es el **mejor estimador** (o **predictor**) de X en L^o, o que Z es el **estimador en la media cuadrática** de X.

Proposición 11.16. (Interpretación de $E(\mathrm{X}|\mathcal{G})$)
Sea $\mathcal{G}$ una sub–σ–álgebra de $\mathcal{F}$ y sea $L^o \subset L_2$ el subespacio que consiste de las vv.aa. en L_2 que son $\mathcal{G}$–medible, es decir, $L^o := L_2(\Omega, \mathcal{G}, \mathrm{P})$. Si X está en L_2, entonces $\widehat{\mathrm{X}} := E(\mathrm{X}|\mathcal{G})$ es la proyección de X sobre L^o.

Demostración. Por el Teorema 11.8(a), la esperanza condicional $\widehat{\mathrm{X}}$ es $\mathcal{G}$–medible, mientras que por la desigualdad de Jensen en el Teorema 11.14(g), $\widehat{\mathrm{X}}$ está en L_2 (explique). Por lo tanto, $\widehat{\mathrm{X}}$ satisface la condición (a) en la Definición 11.15, es decir, $\widehat{\mathrm{X}}$ está en $L^o = L_2(\Omega, \mathcal{G}, \mathrm{P})$. Ahora demostraremos que

$$\| \mathrm{X} - \widehat{\mathrm{X}} \| = \min\{\| \mathrm{X} - Y \| \,|Y \in L^o.\} \tag{11.28}$$

En efecto, para cualquier $Y \in L^o$,

$$
\begin{aligned}
E(\mathrm{X}-Y)^2 &= E[(\mathrm{X}-\widehat{\mathrm{X}})+(\widehat{\mathrm{X}}-Y)]^2 \\
&= E(\mathrm{X}-\widehat{\mathrm{X}})^2 + 2\,E[(\mathrm{X}-\widehat{\mathrm{X}})(\widehat{\mathrm{X}}-Y)] + E(\widehat{\mathrm{X}}-Y)^2 \qquad (11.29)\\
&\geq E(\mathrm{X}-\widehat{\mathrm{X}})^2 + 2\,E[(\mathrm{X}-\widehat{\mathrm{X}})(\widehat{\mathrm{X}}-Y)].
\end{aligned}
$$

Pero $E[(\mathrm{X}-\widehat{\mathrm{X}})(\widehat{\mathrm{X}}-Y)] = 0$ porque

$$
\begin{aligned}
E[(\mathrm{X}-\widehat{\mathrm{X}})(\widehat{\mathrm{X}}-Y)] &= E\{E[(\mathrm{X}-\widehat{\mathrm{X}})(\widehat{\mathrm{X}}-Y)|\mathcal{G}]\} && \text{[por 11.14(b)]}\\
&= E\{(\widehat{\mathrm{X}}-Y)E(\mathrm{X}-\widehat{\mathrm{X}}|\mathcal{G})\} && \text{[por 11.14(c)]}\\
&= E\{(\widehat{\mathrm{X}}-Y)(\widehat{\mathrm{X}}-\widehat{\mathrm{X}})\} && \text{[por 11.14(d)]}\\
&= 0.
\end{aligned}
$$

Esto significa que

$$E(\mathrm{X}-Y)^2 \geq E(\mathrm{X}-\widehat{\mathrm{X}})^2 \quad \forall\, Y \in L^o,$$

es decir, $\| \mathrm{X}-Y \| \geq \| \mathrm{X}-\widehat{\mathrm{X}} \|$ para todo $Y \in L^o$. Esto demuestra (11.28). $\square$

Tomando $Y = E\mathrm{X}$ en (11.29) se obtiene lo siguiente.

Corolario 11.17. *Para cualquier v.a.* $\mathrm{X} \in L_2$

$$\mathrm{Var}(\mathrm{X}) \geq \mathrm{Var}[\mathrm{E}(\mathrm{X}|\mathcal{G})],$$

con igualdad ssi $\mathrm{X} = E(\mathrm{X}|\mathcal{G})$ *c.s.*

En lugar del "mejor estimador" de una v.a. X, en el sentido de la Definición 11.15, podemos considerar el mejor estimador **lineal** de X como en el siguiente ejemplo.

Ejemplo 11.18. Sean $\mathrm{X}, Y_1, \ldots, Y_n$ vv.aa. en L_2. Supóngase que $Y_1, \ldots, Y_n$ son independientes, con media cero y varianzas positivas σ_j^2. Sea $L^* \subset L_2$ el subespacio vectorial que consiste de todas las combinaciones lineales de $Y_1, \ldots, Y_n$, es decir,

$$Y = a_1 Y_1 + \cdots + a_n Y_n, \quad \text{con} \quad a_1, \ldots, a_n \in \mathbb{R}.$$

Demuestre que la proyección de X sobre L^* es la v.a.

$$Z = \sum_{j=1}^{n} \hat{a}_j Y_j \qquad (11.30)$$

con coeficientes $\hat{a}_j = E(\mathrm{X}Y_j)/\sigma_j^2$ para $j = 1, \ldots, n$.

Solución. Sea $Y = \sum_{j=1}^n a_j Y_j$ una v.a. en L^*. Entonces

$$E(\mathrm{X} - Y)^2 = E(\mathrm{X} - \sum_j a_j Y_j)^2 = E[\mathrm{X}^2 + (\sum_j a_j Y_j)^2 - 2\sum_j a_j \mathrm{X} Y_j]$$
$$= E(\mathrm{X}^2) + \sum_j a_j^2 E(Y_j^2) - 2\sum_j a_j E(\mathrm{X}Y_j).$$

Calculando las derivadas parciales de la expresión anterior con respecto a $a_j (j = 1, \ldots, n)$ vemos que

$$\frac{\partial}{\partial a_j} E(\mathrm{X} - Y)^2 = 2a_j E(Y_j^2) - 2E(\mathrm{X}Y_j) \quad \forall\, j = 1, \ldots, n.$$

Como estas derivadas son cero si $a_j = E(\mathrm{X}Y_j)/E(Y_j^2)$ y, además, las segundas derivadas son no-negativas, obtenemos (11.30).

A la v.a. Z en (11.30) se le llama también el mejor **estimador lineal** de X en términos de las vv.aa. $Y_1, \ldots, Y_n$.

Ejercicios § 11

11.1. Sean X y Y vv.aa. discretas o continuas, y sea $F(y|x)$ la distribución condicional de Y dado $\mathrm{X} = x$. Demuestre que si X y Y son independientes, entonces $F(y|x) = F_Y(y)$. (Compare con la Proposición 11.2(b).)

11.2. Si Y está en L_2, definimos la *varianza condicional de* Y *dado que* $\mathrm{X} = x$ como

$$\mathrm{Var}(Y|\mathrm{X} = x) := E[(Y - E(Y|\mathrm{X} = x))^2 | \mathrm{X} = x]$$

y la *varianza condicional de* Y *dada la v.a.* X como la v.a.

$$\mathrm{Var}(Y|\mathrm{X}) := E[(Y - E(Y|\mathrm{X}))^2 | \mathrm{X}]. \qquad (*)$$

Demuestre que

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y|\mathrm{X})] + \mathrm{Var}[E(Y|\mathrm{X})]. \qquad (**)$$

Sugerencia: Escriba $\mathrm{Var}(Y) := E(Y - EY)^2 = E[(Y - E(Y|\mathrm{X})) + (E(Y|\mathrm{X}) - EY)]^2$. Expanda el lado derecho de esta expresión y después use (*) y la propiedad de la esperanza iterada (11.17).

11.3. Sean $\mathrm{X}_1, \mathrm{X}_2, \ldots, S_n$ y N como en el Ejemplo 11.1. Suponga que $\mu := E\mathrm{X}_1$ y EN son finitas. Demuestre que

(a) $E(S_N|N) = \mu \cdot N$,

(b) $E(S_N) = \mu \cdot EN$.

Si además X_1 y N están en L_2, entonces

(c) $\mathrm{Var}[E(S_N|N)] = \mu^2 \ \mathrm{Var}(N)$,

(d) $\mathrm{Var}(S_N|N = n) = n\sigma^2$, en donde $\sigma^2 := \mathrm{Var}(\mathrm{X}_1)$,

(e) $E[\mathrm{Var}(S_N|N)] = \sigma^2 EN$, y

(f) $\mathrm{Var}(S_N) = \sigma^2 EN + \mu^2 \ \mathrm{Var}(N)$. (*Sugerencia:* use (**) del Ejercicio 11.2.)

11.4. Sea $\mathrm{X} \sim \mathrm{Poi}(\lambda_1)$ y $Y \sim \mathrm{Poi}(\lambda_2)$ vv.aa. independientes [de modo que $\mathrm{X} + Y \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$]. Dado un entero positivo n, demuestre que

$$\mathrm{P}(Y = k|\mathrm{X} + Y = n) \quad \forall\, k = 0, 1, \ldots, n$$

es una distribución $\mathrm{Bin}(n, p)$ con $p := \lambda_2/(\lambda_1 + \lambda_2)$.

11.5. Sean X y Y vv.aa. continuas con densidad conjunta

$$\begin{aligned} f(x, y) &:= 1/x \quad \text{si} \quad 0 \leq y < x \leq 1, \\ &:= 0 \qquad \text{en c.c.} \end{aligned}$$

Demuestre que dado el evento $\{\mathrm{X} = x\}$, con $x > 0$, la v.a. Y tiene densidad uniforme sobre el intervalo $[0, x]$.

11.6. Demuestre que $\| \mathrm{X} \| := \sqrt{E(\mathrm{X}^2)}$ define una **norma** sobre L_2, es decir, (a) $\| \mathrm{X} \| \geq 0$ y $\| \mathrm{X} \| = 0$ ssi $\mathrm{X} = 0$; (b) $\| a\mathrm{X} \| = |a| \ \| \mathrm{X} \| \ \ \forall\, a \in \mathbb{R}$; (c) $\| \mathrm{X} + Y \| \leq \| \mathrm{X} \| + \| Y \|$. (Nota: la demostración de "$\| \mathrm{X} \| = 0 \Rightarrow \mathrm{X} = 0$" requiere la *Nota* 11.11.)

11.7. Sean X y Y vv.aa. con densidad conjunta

(1) $f(x, y) := \lambda^2 e^{-\lambda y}$ para $0 \leq x \leq y$,

(2) $f(x, y) := xe^{-x(y+1)}$ para $x, y \geq 0$.

En cada caso calcule (1) las densidades marginales de X y Y, y (2) la densidad condicional y la esperanza condicional de Y dada X.

11.8. Sean X y Y vv.aa. discretas con densidades f_{X} y f_Y respectivamente. Definimos la **convolución** de f_{X} y f_Y como la función $f_{\mathrm{X}} * f_Y$ dada por

$$(f_{\mathrm{X}} * f_Y)(y) := \sum_x f_{\mathrm{X}}(x) f_Y(y-x).$$

(a) Demuestre que si X y Y son *independientes*, entonces la densidad de $\mathrm{X}+Y$ es la convolución de las densidades de X y Y, i.e.

$$f_{\mathrm{X}+Y}(y) := \mathrm{P}(\mathrm{X}+Y=y) = (f_{\mathrm{X}} * f_Y)(y).$$

(b) Suponga que X y Y son i.i.d. con densidad común

$$f(k) := p(1-p)^{k-1} \quad \forall\ k = 1, 2, \ldots.$$

Demuestre que $f_{\mathrm{X}+Y}(y) = (y-1)p^2(1-p)^{y-2}$ para $y = 2, 3 \ldots$.

11.9. Sean X y Y vv.aa. continuas con densidades f_{X} y f_Y respectivamente. La **convolución** de f_{X} y f_Y es la función $f_{\mathrm{X}} * f_Y$ definida como

$$(f_{\mathrm{X}} * f_Y)(z) := \int_{-\infty}^{\infty} f_{\mathrm{X}}(x) f_Y(z-x) dx.$$

Sea $f(x,y)$ la densidad conjunta de X y Y, y $f(y|x)$ la densidad condicional de Y dado que $\mathrm{X}=x$. Demuestre que

(a) $\mathrm{P}(\mathrm{X}+Y \leq z | \mathrm{X}=x) = \int\limits_{-\infty}^{z-x} f(y|x) dy$,

(b) $F_{\mathrm{X}+Y}(z) := \mathrm{P}\{\mathrm{X}+Y \leq z\} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{z-x} f(y|x) dy\ f_{\mathrm{X}}(x) dx$,

(c) La densidad de $\mathrm{X}+Y$ es $f_{\mathrm{X}+Y}(z) = \int\limits_{-\infty}^{\infty} f(x, z-x) dx$.

(d) Si además X y Y son *independientes*, entonces $f_{\mathrm{X}+Y} = f_{\mathrm{X}} * f_Y$.

11.10. Sean X y Y vv.aa. i.i.d. Calcule la densidad de $\mathrm{X}+Y$ en cada uno de los siguientes casos: (a) $\mathrm{X} \sim \mathrm{Exp}(\lambda)$, (b) $\mathrm{X} \sim N(0,1)$. (*Sugerencia:* use los Ejercicios 11.8 y 11.9.)

11.11. Suponga que los tiempos que requieren dos estudiantes para resolver un mismo problema son independientes y tienen distribución exponencial con parámetro λ. Calcule la probabilidad de que el primer estudiante tome al menos el doble del tiempo que requiere el segundo estudiante para resolver el problema.

11.12. Sean X y Y vv.aa. continuas con densidad conjunta de la forma

$$f(x,y) = c\, e^{-(x^2-xy+4y^2)/2} \quad \forall\, x,y \in \mathbb{R}.$$

(a) Calcule el valor de c para el que f es efectivamente una densidad de probabilidad.

(b) Calcule las densidades marginales de X y Y.

(c) Calcule $E(Y|\mathrm{X})$ y $E(\mathrm{X}|Y)$.

Parte 2: PROCESOS ESTOCÁSTICOS

12 Cadenas de Markov: conceptos básicos

Contenido: Clasificación de procesos estocásticos, procesos de Markov, cadenas de Markov, ecuación de Chapman–Kolmogorov, distribución invariante, cadenas de nacimiento y muerte.

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad, $(S, \mathcal{S})$ un espacio medible, y T un conjunto arbitrario de "parámetros" o "índices".

Definición 12.1. Una colección $\mathrm{X}_\bullet = \{\mathrm{X}_t, t \in T\}$ de vv.aa. X_t sobre $(\Omega, \mathcal{F}, \mathrm{P})$ con valores en S se dice que es un **proceso estocástico** (PE), también llamado **proceso aleatorio**, con **espacio de estados** S y **conjunto de índices** T.

Los PEs se pueden clasificar de varias maneras. En particular, en términos de T y S los casos típicos son los siguientes.

- Si $T \subset \mathbb{R}$ es un conjunto a lo más numerable, se dice que $\mathrm{X}_\bullet$ es un PE a **tiempo discreto**.
- Si $T \subset \mathbb{R}$ es un intervalo, $\mathrm{X}_\bullet$ es un PE a **tiempo continuo**.
- Si $T \subset \mathbb{R}^n$ con $n \geq 2$, $\mathrm{X}_\bullet$ es un **campo aleatorio**.
- Si S es un conjunto a lo más numerable, se dice que $\mathrm{X}_\bullet$ es un PE **discreto**.
- Si S es "continuo", $\mathrm{X}_\bullet$ es un PE **continuo**.

Por supuesto, estas clasificaciones se pueden combinar en varias formas posibles. Por ejemplo, $\mathrm{X}_\bullet$ puede ser un PE discreto a tiempo continuo o discreto a tiempo discreto.

Por otra parte, un PE $\mathrm{X}_\bullet = \{\mathrm{X}_t, t \in T\}$ se puede ver como una colección de funciones de dos variables $(t, \omega) \mapsto \mathrm{X}_t(\omega) \equiv \mathrm{X}(t, \omega)$ de $T \times \Omega \to S$ tales que:

- para cada $t \in T$ **fijo**, la función $\omega \mapsto \mathrm{X}_t(\omega)$ es una v.a. de Ω en S;

- para cada $\omega \in \Omega$ **fijo**, $t \mapsto \mathrm{X}_t(\omega)$ es una función de T en S que se llama una **trayectoria** del PE.

En el siguiente ejemplo se ve una clase especial de PEs continuos a tiempo continuo con el que se intenta motivar la condición o propiedad de Markov.

Ejemplo 12.2. Sistemas determinísticos. Considere un proceso $x(\cdot) \in \mathbb{R}^n$ definido por la ecuación diferencial

$$\dot{x}(t) = F(x(t)) \quad \forall\, t \geq 0, \quad \text{con} \quad x(0) = x_0. \tag{12.1}$$

Bajo ciertas hipótesis, la ecuación (1) tiene una solución única dada por

$$x(t) = x_0 + \int_0^t F(x(r))dr \quad \forall\, t \geq 0.$$

Además, para $t \geq s \geq 0$ tenemos

$$x(t) = x(s) + \int_s^t F(x(r))dr. \tag{12.2}$$

Interpretando a s como el "tiempo presente" y a $t > s$ como el "tiempo futuro", la ecuación (2) dice que el estado presente $x(s)$ "determina el futuro" $x(t)$, o bien que, dado el estado presente $x(s)$, el futuro $x(t)$ es "independiente del pasado" $x(r)$ para $r < s$. Por tal motivo se dice que (2) es una condición de **causalidad** (el presente determina el futuro); también se dice que el sistema determinístico $x(\cdot)$ **no tiene memoria** o que satisface la **condición de Markov** (también llamada **propiedad de Markov**). $\square$

Para PEs la condición de Markov se expresa de manera análoga a (2). Por ejemplo, si $X_\bullet = \{\mathrm{X}_t, t \geq 0\}$ es un PE a tiempo continuo, se dice que $\mathrm{X}_\bullet$ satisface la propiedad de Markov o que $\mathrm{X}_\bullet$ es un **proceso de Markov** si

$$\mathrm{P}(\mathrm{X}_t \in B | \mathrm{X}_r \,\forall\, 0 \leq r \leq s) = \mathrm{P}(\mathrm{X}_t \in B | \mathrm{X}_s) \tag{12.3}$$

para todo $B \in \mathcal{S}$ y todo $0 \leq s \leq t$. Análogamente, para un PE $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ a tiempo discreto la propiedad de Markov se satisface si

$$\mathrm{P}(\mathrm{X}_m \in B | \mathrm{X}_0, \ldots, \mathrm{X}_n) = \mathrm{P}(\mathrm{X}_m \in B | \mathrm{X}_n) \tag{12.4}$$

para todo $B \in \mathcal{S}$ y todo $m \geq n \geq 0$. En el caso a tiempo discreto se dice que $X_\bullet$ es una **cadena de Markov** (CM). Además, veremos que basta verificar (4) para $m = n+1$, es decir, la propiedad de Markov se puede expresar como

$$P(X_{n+1} \in B | X_0, \ldots, X_n) = P(X_{n+1} \in B | X_n) \quad \forall\, B \in \mathcal{S},\ n = 0, 1, \ldots. \tag{12.5}$$

En el resto de esta sección estudiaremos algunos conceptos básicos sobre CMs **discretas** y **continuas**, es decir, con espacio de estados S numerable y no–numerable, respectivamente.

En el caso continuo supondremos que S es un conjunto de Borel en $\mathbb{R}$ y que la σ–álgebra $\mathcal{S} = \mathcal{B}(S)$. Entonces podemos escribir (5) como

$$P(X_{n+1} \in B | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} \in B | X_n = x_n) \tag{12.6}$$

para todo $B \in \mathcal{B}(S)$, $x_0, \ldots, x_n \in S$, y $n = 0, 1, \ldots$. Tomando $x_n = x$, la probabilidad condicional en el lado derecho de (6) se llama la **probabilidad de transición de x a B en un paso** y escribimos

$$P(x, B) := P(X_{n+1} \in B | X_n = x). \tag{12.7}$$

En el caso discreto (S numerable), en (6) y (7) podemos sustituir el conjunto $B \in \mathcal{B}(S)$ por un estado arbitrario $y \in S$, y obtenemos

$$P(X_{n+1} = y | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = y | X_n = x_n), \tag{12.8}$$

$$P_n(x, y) := P(X_{n+1} = y | X_n = x). \tag{12.9}$$

Algunas veces escribiremos $P_n(x, y)$ como $P^{(n)}_{xy}$.

CMs discretas. Considérese el caso discreto (8), (9). A menos que se diga lo contrario, siempre supondremos que la CM es **homogénea** (en el tiempo), lo cual significa que la probabilidad de transición (9) **no** depende de n, es decir,

$$P(x, y) = P(X_{n+1} = y | X_n = x) = P(X_1 = y | X_0 = x) \quad \forall\, n = 0, 1, \ldots. \tag{12.10}$$

La matriz $P = [P(x, y)] \equiv [P_{xy}]$ se llama la **matriz de transición** (en un paso) de la CM $X_\bullet$. Es evidente que P es una **matriz estocástica**, i.e.

$$(a)\ P(x, y) \geq 0\ \ \forall\, x, y \in S, \quad (b) \sum_{y \in S} P(x, y) = 1\ \ \forall\, x \in S. \tag{12.11}$$

De hecho, cualquier matriz cuadrada de elementos no–negativos y cuyas filas suman 1 se dice que es **estocástica** o **aleatoria**.

Ejemplo 12.3. Una caminata aleatoria. Sea $\xi_\bullet = \{\xi_0, \xi_1, \ldots\}$ una sucesión de vv.aa. i.i.d. con distribución

$$\mathrm{P}(\xi_0 = 1) = p, \quad \mathrm{P}(\xi_0 = -1) = q := 1 - p. \tag{12.12}$$

Sea $\mathrm{X}_n := \xi_0 + \cdots + \xi_n$ para $n = 0, 1, \ldots$. Nótese que

$$\mathrm{X}_{n+1} = \mathrm{X}_n + \xi_{n+1}$$

y que ξ_{n+1} es independiente de $\mathrm{X}_0, \ldots, \mathrm{X}_n$ para todo n. De aquí se sigue la propiedad de Markov (8) porque

$$\begin{aligned}
\mathrm{P}(\mathrm{X}_{n+1} = {} & y|\mathrm{X}_0 = x_0, \ldots, \mathrm{X}_{n-1} = x_{n-1}, \mathrm{X}_n = x) \\
= {} & \mathrm{P}(\xi_{n+1} = y - x|\mathrm{X}_0 = x_0, \ldots, \mathrm{X}_{n-1} = x_{n-1}, \mathrm{X}_n = x) \\
= {} & \mathrm{P}(\xi_{n+1} = y - x) \\
= {} & \mathrm{P}(\mathrm{X}_{n+1} = y|\mathrm{X}_n = x).
\end{aligned} \tag{12.13}$$

Es decir, $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ es una CM con espacio de estados $S = \{0, \pm 1, \pm 2, \ldots\}$. Además, por (12) y (13), las probabilidades de transición en un paso son

$$\mathrm{P}(x, y) = \mathrm{P}(\mathrm{X}_{n+1} = y|\mathrm{X}_n = x) = \mathrm{P}(\xi_0 = y - x),$$

i.e.

$$\mathrm{P}(x, y) = \begin{cases} p & \text{si} \quad y = x + 1, \\ q & \text{si} \quad y = x - 1. \end{cases}$$

Por lo tanto, la matriz de transición de $\mathrm{X}_\bullet$ es una matriz infinita cuya fila x, para cada $x \in S$, es de la forma

$$(\cdots 0 \ \mathrm{P}(x, x-1) \ 0 \ \mathrm{P}(x, x+1) \ 0 \cdots)$$

con $\mathrm{P}(x, x-1) = q$, $\mathrm{P}(x, x+1) = p$. (En el Ejercicio 6 se considera una ligera generalización de este ejemplo.) $\square$

Sea $\mathrm{X}_\bullet = \{\mathrm{X}_0, \mathrm{X}_1, \ldots\}$ una CM con espacio de estados S. A la v.a. X_0 se le llama el **estado inicial** de la CM y a su distribución, digamos

$$\pi(x) := \mathrm{P}(\mathrm{X}_0 = x) \quad \forall\, x \in S,$$

se le llama la **distribución inicial**. Si además $\mathrm{X}_\bullet$ tiene matriz de transición P, entonces se dice que $\mathrm{X}_\bullet$ es una CM (π, P). La siguiente proposición establece que π y P determinan la distribución del vector $(\mathrm{X}_0, \ldots, \mathrm{X}_n)$ para todo n.

Nota. El nombre "cadena de Markov" surgió precisamente de expresiones como (14).

Proposición 12.4. Un PE $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ es una CM (π, P) ssi para cualquier $n = 0, 1, \ldots$ y $x_0, \ldots, x_n \in S$

$$\mathrm{P}(\mathrm{X}_0 = x_0, \mathrm{X}_1 = x_1, \ldots, \mathrm{X}_n = x_n) = \pi(x_0)\mathrm{P}(x_0, x_1) \cdots \mathrm{P}(x_{n-1}, x_n). \quad (12.14)$$

Demostración. Si $\mathrm{X}_\bullet$ es una CM (π, P), entonces (14) se sigue trivialmente de la "regla de la multiplicación" (Proposición 3.2) y de la propiedad de Markov (8). Recíprocamente, si se cumple (14), sumando sobre todo $x_n \in S$ se obtiene que

$$\mathrm{P}(\mathrm{X}_0 = x_0, \ldots, \mathrm{X}_{n-1} = x_{n-1}) = \pi(x_0)\mathrm{P}(x_0, x_1) \cdots \mathrm{P}(x_{n-2}, x_{n-1}).$$

Repitiendo este argumento se llega a que $\mathrm{P}(\mathrm{X}_0 = x_0) = \pi(x_0)$ para todo $x_0 \in S$; es decir, π es la distribución inicial de $\mathrm{X}_\bullet$. Finalmente, usando (14) y la definición de probabilidad condicional (vea la ecuación (1) en la sección 3) se obtiene la propiedad de Markov (8) y que

$$\mathrm{P}(\mathrm{X}_{n+1} = y | \mathrm{X}_0 = x_0, \ldots, \mathrm{X}_n = x_n) = \mathrm{P}(x_n, y). \; \square$$

$\square$

La Proposición 12.4 se puede extender como sigue.

Proposición 12.5. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ un PE con espacio de estados S numerable. Las siguientes afirmaciones son equivalentes:

(a) $\mathrm{X}_\bullet$ es una CM (π, P).

(b) $\mathrm{P}(\mathrm{X}_{n+1} = y_1, \ldots, \mathrm{X}_{n+m} = y_m | \mathrm{X}_0 = x_0, \ldots, \mathrm{X}_{n-1} = x_{n-1}, \mathrm{X}_n = x)$

$$= \mathrm{P}(x, y_1)\mathrm{P}(y_1, y_2) \cdots \mathrm{P}(y_{m-1}, y_m)$$

para todo $n \geq 0, m \geq 1$, y estados $x_0, \ldots, x_{n-1}, x, y_1, \ldots, y_m$ en S.

(c) Si $A_0, \ldots, A_{n-1}$ son subconjuntos de S, entonces

$$\mathrm{P}(\mathrm{X}_{n+1} = y_1, \ldots, \mathrm{X}_{n+m} = y_m | \mathrm{X}_0 \in A_0, \ldots, \mathrm{X}_{n-1} \in A_{n-1}, \mathrm{X}_n = x)$$

$$= \mathrm{P}(x, y_1)\mathrm{P}(y_1, y_2) \cdots \mathrm{P}(y_{m-1}, y_m)$$

para todo $n \geq 1, m \geq 1$, y estados $x, y_1, \ldots, y_m$ en S.

(d) Si $A_0, \ldots, A_{n-1}, B_1, \ldots, B_m$ son subconjuntos de S,

$$\mathrm{P}(\mathrm{X}_{n+1} \in B_1, \ldots, \mathrm{X}_{n+m} \in B_m | \mathrm{X}_0 \in A_0, \ldots, \mathrm{X}_{n-1} \in A_{n-1}, \mathrm{X}_n = x)$$
$$= \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} \mathrm{P}(x, y_1)\mathrm{P}(y_1, y_2) \cdots \mathrm{P}(y_{m-1}, y_m).$$

La demostración de 12.5 se deja al lector (Ejercicio 7).

Además de las probabilidades de transición de un paso consideraremos las **probabilidades de transición en n pasos** $\mathrm{P}_n(x, y)$, definidas como

$$\mathrm{P}_n(x, y) := \mathrm{P}(\mathrm{X}_n = y | \mathrm{X}_0 = x) \quad \forall\, n = 0, 1, \ldots$$

con $\mathrm{P}_0(x, y) := \delta(x, y)$ (la delta de Kronecker, que algunas veces escribimos como δ_{xy}), y $\mathrm{P}_1(x, y) := \mathrm{P}(x, y)$. Nótese que, como la CM es homogénea,

$$\mathrm{P}_n(x, y) = \mathrm{P}(\mathrm{X}_{n+m} = y | \mathrm{X}_m = x) \quad \forall\, n, m \geq 0. \tag{12.15}$$

Asimismo, definimos la **matriz de transición de n pasos**

$$\mathrm{P}_n := [\mathrm{P}_n(x, y)] \quad \forall\, n = 0, 1, \ldots,$$

con $\mathrm{P}_0 := I$ (la matriz identidad) y $\mathrm{P}_1 := \mathrm{P}$, la matriz de transición de un paso. El siguiente teorema establece, en particular, que P_n se puede calcular multiplicando P consigo misma n veces, i.e.

$$\mathrm{P}_n = \mathrm{P}^n \quad \forall\, n = 0, 1, \ldots. \tag{12.16}$$

Teorema 12.6. *Para cualquier $n, m = 0, 1, \ldots$*

$$\mathrm{P}_{n+m} = \mathrm{P}_n \cdot \mathrm{P}_m, \tag{12.17}$$

i.e.

$$\mathrm{P}_{n+m}(x, y) = \sum_{z \in S} \mathrm{P}_n(x, z)\mathrm{P}_m(z, y) \quad \forall\, x, y \in S. \tag{12.18}$$

Nótese que (17) implica (16). A la ecuación (17) o a (18) se le llama la **ecuación de Chapman–Kolmogorov**.

Demostración. Calcularemos (18):

$$\begin{aligned} \mathrm{P}_{n+m}(x, y) &:= \mathrm{P}(\mathrm{X}_{n+m} = y | \mathrm{X}_0 = x) \\ &= \sum_{z \in S} \mathrm{P}(\mathrm{X}_n = z, \mathrm{X}_{n+m} = y | \mathrm{X}_0 = x). \end{aligned} \tag{12.19}$$

Pero, por el Ejercicio 2(a),

$$
\begin{aligned}
\mathrm{P}(\mathrm{X}_n\!=\!z, \mathrm{X}_{n+m}\!=\!y|\mathrm{X}_0\!=\!x) &= \mathrm{P}(\mathrm{X}_n\!=\!z|\mathrm{X}_0\!=\!x)\mathrm{P}(\mathrm{X}_{n+m}\!=\!y|\mathrm{X}_0\!=\!x, \mathrm{X}_n\!=\!z) \\
&= \mathrm{P}_n(x,z)\mathrm{P}_m(z,y), \qquad (12.20)
\end{aligned}
$$

en donde hemos usado la Proposición 12.5(d) y la ecuación (15). Finalmente, sustituyendo (20) en (19) se obtiene (18). □ □

Ejemplo 12.7. Una caminata aleatoria restringida. Considérese la CM $\mathrm{X}_n := \xi_0 + \cdots + \xi_n$ del Ejemplo 12.3, pero en lugar de (12) las vv.aa. ξ_n tienen distribución

$$
\mathrm{P}(\xi_0 = 1) = p, \quad \mathrm{P}(\xi_0 = 0) = 1 - p =: q.
$$

Entonces $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ tiene espacio de estados $S = \{0, 1, \ldots\}$ y probabilidades de transición

$$
\mathrm{P}(x,y) = \mathrm{P}(\xi_{n+1} = y - x) = \begin{cases} q & \text{si} \quad y = x, \\ p & \text{si} \quad y = x+1 \end{cases}
$$

para todo $x \in S$. Es decir, la matriz de transición P en un paso es de la forma

$$
\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\
\left[\begin{array}{ccccc}
q & p & 0 & 0 & \cdots \\
0 & q & p & 0 & \cdots \\
0 & 0 & q & p & \cdots \\
0 & 0 & 0 & q & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdots
\end{array}\right]
\end{array}
$$

Para calcular las probabilidades de transición de n pasos ($n = 2, 3, \ldots$) podemos calcular P^n y usar (16) para obtener $\mathrm{P}_n(x,y)$. Sin embargo, en este ejemplo particular es más fácil observar que $\mathrm{X}_n = \mathrm{X}_0 + Y_n$, en donde $Y_n := \xi_1 + \cdots + \xi_n$ es una v.a. binomial con parámetros n y p. Luego, como $\mathrm{X}_0 = \xi_0$ y Y_n ($n = 1, 2, \ldots$) son independientes vemos que

$$
\mathrm{P}_n(x,y) = \mathrm{P}(\mathrm{X}_n = y|\mathrm{X}_0 = x) = \mathrm{P}(\mathrm{X}_0 + Y_n = y|\mathrm{X}_0 = x) = \mathrm{P}(Y_n = y - x).
$$

Finalmente, usando el hecho de que $Y_n \sim$ Bin (n,p) obtenemos

$$
\mathrm{P}_n(x,y) = \binom{n}{y-x} p^{y-x} q^{n-y+x} \quad \forall\, y = x, x+1, \ldots, x+n,
$$

para cualquier estado $x \in S$. □

Para cada $n = 0, 1, \ldots$, denotaremos por $\pi_n = \{\pi_n(x), x \in S\}$ la **distribución de** X_n, i.e.

$$\pi_n(x) := \mathrm{P}(\mathrm{X}_n = x). \tag{12.21}$$

En particular, π_0 es la distribución inicial de la CM $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$. Interpretaremos π_n como un "vector fila". Se tiene entonces que

$$\pi_n = \pi_0 \mathrm{P}_n = \pi_0 \mathrm{P}^n \quad \forall\, n = 0, 1, \ldots, \tag{12.22}$$

o bien que

$$\pi_n = \pi_{n-1} \mathrm{P} \quad \forall\, n = 1, 2, \ldots. \tag{12.23}$$

En efecto, como

$$\pi_n(y) = \mathrm{P}(\mathrm{X}_n = y) = \sum_{x \in S} \mathrm{P}(\mathrm{X}_0 = x, \mathrm{X}_n = y)$$

se sigue que

$$\pi_n(y) = \sum_{x \in S} \pi_0(x) \mathrm{P}_n(x, y) \quad \forall\, y \in S,$$

lo cual da (22). Por otra parte, $\pi_n = \pi_0 \mathrm{P}^n = \pi_0 \mathrm{P}^{n-1} \mathrm{P} = \pi_{n-1} \mathrm{P}$ y se obtiene (23).

Ejemplo 12.8. CM con dos estados. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ una CM con espacio de estados $S = \{0, 1\}$ y probabilidades de transición

$$\mathrm{P}(0, 1) =: p \quad \text{y} \quad \mathrm{P}(1, 0) =: q,$$

en donde p, q y $\lambda := 1 - p - q$ son números positivos. La matriz de transición P es

$$\begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \end{array}$$

Demuestre que $\pi_n(x) := \mathrm{P}(\mathrm{X}_n = x)$ y la matriz de transición de n pasos satisfacen:

(a) $\pi_n(0) = \dfrac{q}{p+q} + \lambda^n \left(\pi_0(0) - \dfrac{q}{p+q} \right)$, de modo que

$$\pi_n(1) = 1 - \pi_n(0) = \frac{p}{p+q} + \lambda^n \left(\pi_0(1) - \frac{p}{p+q} \right)$$

(b) $\mathrm{P}^n = \dfrac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \dfrac{\lambda^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$.

Solución. (a) Por (23):

$$\pi_n(0) = \pi_{n-1}(0)(1-p) + \pi_{n-1}(1)q = \pi_{n-1}(0)(1-p) + (1-\pi_{n-1}(0))q,$$

i.e.

$$\pi_n(0) = \lambda\, \pi_{n-1}(0) + q \quad \forall\, n = 1, 2, \ldots.$$

Iterando esta relación se obtiene (a).

(b) Como (22) se cumple para *cualquier* distribución inicial, tomando $\pi_0 = (1,0)$ vemos que $\mathrm{P}_n(0,0) = \pi_n(0)$ y, por lo tanto, $\mathrm{P}_n(0,1) = 1 - \pi_n(0) = \pi_n(1)$. Luego, por (a),

$$\mathrm{P}_n(0,0) = \frac{q}{p+q} + \lambda^n \frac{p}{p+q}, \quad \mathrm{P}_n(0,1) = \frac{p}{p+q} - \lambda^n \frac{p}{p+q}.$$

Análogamente, tomando $\pi_0 = (0,1)$, de (a) y (22):

$$\mathrm{P}_n(1,1) = \pi_n(1) = \frac{p}{p+q} + \lambda^n \frac{q}{p+q},$$

$$\mathrm{P}_n(1,0) = 1 - \mathrm{P}_n(1,1) = \frac{q}{p+q} - \lambda^n \frac{q}{p+q}. \quad \square$$

En la siguiente proposición se identifica una familia muy grande de CMs.

Proposición 12.9. Sean $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ y $Y_\bullet = \{Y_n, n = 0, 1, \ldots\}$ dos PEs, y supóngase que $Y_0, Y_1, \ldots$ son independientes. Si X_0 es independiente de $Y_\bullet$ y, además, existe una función de Borel $G : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ tal que

$$\mathrm{X}_{n+1} = G(\mathrm{X}_n, Y_n) \quad \forall\, n = 0, 1, \ldots, \tag{12.24}$$

entonces $\mathrm{X}_\bullet$ es una CM.

Demostración. Verificaremos que $\mathrm{X}_\bullet$ satisface la propiedad de Markov (6). Para tal fin, primero nótese que para cualquier $n = 0, 1, \ldots$, las vv.aa. $\{\mathrm{X}_0, \ldots, \mathrm{X}_n\}$ y $\{Y_n, Y_{n+1}, \ldots\}$ son *independientes*. Por lo tanto, para cualquier conjunto de Borel B,

$$\begin{aligned}
\mathrm{P}(\mathrm{X}_{n+1} \in B | \mathrm{X}_0 = x_0, \ldots, \mathrm{X}_n = x_n) \\
&= \mathrm{P}[G(x_n, Y_n) \in B | \mathrm{X}_0 = x_0, \ldots, \mathrm{X}_n = x_n] \\
&= \mathrm{P}[G(x_n, Y_n) \in B] \qquad \text{(por independencia)} \\
&= \mathrm{P}[G(\mathrm{X}_n, Y_n) \in B | \mathrm{X}_n = x_n] \\
&= \mathrm{P}(\mathrm{X}_{n+1} \in B | \mathrm{X}_n = x_n). \quad \square
\end{aligned}$$

$\square$

Ejemplo 12.10. Casos especiales de (24). En los ejemplos siguientes, $\{Y_n\}$ es una sucesión de vv.aa. i.i.d. y, además, X_0 es independiente de $\{Y_n\}$.

(a) Proceso autoregresivo de primer orden. Este tipo de procesos, también conocidos como **procesos con "ruido aditivo"**, resulta cuando (24) es de la forma

$$\mathrm{X}_{n+1} = G(\mathrm{X}_n) + Y_n.$$

De hecho, el caso más común es el de los **sistemas lineales**, en los que $G(x) = ax$, es decir,

$$\mathrm{X}_{n+1} = a\mathrm{X}_n + Y_n \tag{12.25}$$

para alguna constante a. Si X_0 es una v.a. gaussiana o una constante, y además el "ruido" Y_n en (25) es gaussiano, entonces las vv.aa. X_n también son gaussianas y por tal motivo se dice que $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ es un **proceso de Gauss–Markov**.

(b) Sistemas de producción. Para cada $n = 0, 1, \ldots$, sea X_n el nivel de inventario, o "stock", de un cierto producto al inicio del periodo $[n, n+1)$, y sea Y_n la demanda del producto en ese mismo periodo. Se supone que el inventario inicial X_0 es independiente de la demanda Y_n para todo $n \geq 0$. Además, supóngase que en el periodo $[n, n+1)$ se produce una cantidad $f(\mathrm{X}_n)$ de artículos. Entonces el proceso $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ de inventarios satisface la ecuación

$$\mathrm{X}_{n+1} = \mathrm{X}_n + f(\mathrm{X}_n) - Y_n$$

y, por ser de la forma (24), es una CM. A la función $f(x)$ se le llama una *estrategia de producción*. Por ejemplo, una estrategia muy usual es la regla (s, S), con $0 < s < S$, definida como

$$\begin{aligned} f(x) &:= S - x \quad \text{si} \quad x \leq s, \\ &:= 0 \qquad\quad\ \text{si} \quad x > s. \end{aligned}$$

En otras palabras, si el nivel de inventario es $x \leq s$, entonces se produce la cantidad $f(x) = S - x$ con el fin de elevar el inventario del nivel x al nivel $x + f(x) = S$. Por el contrario, si $x > s$, la regla (s, S) es **no** producir, i.e. $f(x) = 0$.

(c) Una presa con capacidad C. Para cada $n = 0, 1, \ldots$, sea X_n el volumen de agua en una presa al inicio del periodo $[n, n+1)$ y sea Y_n la entrada de agua (por lluvia, escurrimientos, etc.) en dicho periodo. Asimismo, supóngase que la descarga (por ejemplo, para irrigación o para producir energía eléctrica) es $f(\mathrm{X}_n)$. Entonces la evolución del volumen satisface

$$\mathrm{X}_{n+1} = \min(\mathrm{X}_n - f(\mathrm{X}_n) + Y_n, C),$$

en donde C es la capacidad de la presa. $\square$

Definición 12.11. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ un PE y T una v.a. con valores en el conjunto $\{0, 1, \ldots\} \cup \{+\infty\}$. Decimos que T es un **tiempo de paro** para $\mathrm{X}_\bullet$ si para cualquier $n = 0, 1, \ldots$, el evento $\{T = n\}$ depende sólo de $\mathrm{X}_0, \mathrm{X}_1, \ldots, \mathrm{X}_n$. Si T toma valores en $\{0, 1, \ldots\}$ casi seguramente, es decir $P(T < \infty) = 1$, entonces decimos que T es un tiempo de paro **finito**.

Ejemplo 12.12. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ un PE con espacio de estados S, y sea A un subconjunto de S.

(a) El **primer tiempo de llegada** a A es la v.a.

$$H_A := \inf\{n \geq 0 \mid \mathrm{X}_n \in A\},$$

con $H_A := \infty$ si $\mathrm{X}_n \notin A$ para todo n. (Esta última condición es consistente con la convención $\inf \emptyset := \infty$.) El primer tiempo de llegada es un tiempo de paro para $\mathrm{X}_\bullet$ porque

$$\{H_A = n\} = \{\mathrm{X}_0 \notin A, \ldots, \mathrm{X}_{n-1} \notin A, \mathrm{X}_n \in A\}$$

(b) El **tiempo del primer paso** a A, definido como

$$T_A := \inf\{n \geq 1 \mid \mathrm{X}_n \in A\},$$

también es un tiempo de paro para $\mathrm{X}_\bullet$.

(c) El **tiempo de la última salida** de A,

$$L_A := \sup\{n \geq 0 \mid \mathrm{X}_n \in A\},$$

en general **no** es un tiempo de paro porque el evento $\{L_A = n\}$ depende de que X_{n+m} para $m \geq 1$ visite a A o no. $\square$

Teorema 12.13. (Propiedad fuerte de Markov). *Sea* $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ *una CM y* T *un tiempo de paro finito para* $\mathrm{X}_\bullet$. *Entonces*

$$\mathrm{P}(\mathrm{X}_{T+m} = y \mid \mathrm{X}_k = x_k \quad \forall\, 0 \leq k < T, \mathrm{X}_T = x) = \mathrm{P}(\mathrm{X}_{T+m} = y \mid \mathrm{X}_T = x) \tag{12.26}$$

para todo $x, y \in S, m = 0, 1, \ldots$. *(Vea también el Ejercicio 11.)*

Demostración. Denotemos por I y II el lado izquierdo y el lado derecho de (26), respectivamente. Para demostrar que I = II, nótese que si B es el evento

$$B := \{\mathrm{X}_k = x_k \quad \forall\, 0 \leq k < T\},$$

entonces

$$\begin{aligned} \mathrm{I} &= \mathrm{P}(\mathrm{X}_{T+m} = y \mid B \cap \{\mathrm{X}_T = x\}) \\ &= \mathrm{P}(B \cap \{\mathrm{X}_T = x\} \cap \{\mathrm{X}_{T+m} = y\})/\mathrm{P}(B \cap \{\mathrm{X}_T = x\}). \end{aligned} \tag{12.27}$$

Pero

$$\begin{aligned} \mathrm{P}(B\cap\{\mathrm{X}_T=x\}\cap\{\mathrm{X}_{T+m}=y\}) &= \sum_{n=0}^{\infty} \mathrm{P}(\{T=n\}\cap B\cap\{\mathrm{X}_T=x\}\cap\{\mathrm{X}_{T+m}=y\}) \\ &= \sum_{n=0}^{\infty} \mathrm{P}(\{T = n\} \cap B \cap \{\mathrm{X}_T = x\})\mathrm{P}_m(x, y) \\ &= \mathrm{P}_m(x, y)\mathrm{P}(B \cap \{\mathrm{X}_T = x\}), \end{aligned}$$

y comparando con (27), vemos que $\mathrm{I} = \mathrm{P}_m(x, y)$. Análogamente se demuestra que

$$\mathrm{II} = \mathrm{P}(\mathrm{X}_{T+m} = y \mid \mathrm{X}_T = x) = \mathrm{P}_m(x, y). \quad \square \tag{12.28}$$

$\square$

Observación 12.14. Si $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ es una CM con matriz de transición $\mathrm{P} = [\mathrm{P}(x, y)]$ y T es un tiempo de paro finito para $\mathrm{X}_\bullet$, entonces, condicionado a que $\mathrm{X}_T = x$, el PE $Y_m := \mathrm{X}_{T+m}$ $(m = 0, 1, \ldots)$ es una CM con estado inicial $Y_0 = x$ y matriz de transición P. Esta conclusión se sigue de (26) y (28).

Ejemplo 12.15. Una CM "parcialmente observable". Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ una CM con espacio de estados S, un conjunto numerable, y sea A un subconjunto de S. Si $\mathrm{X}_\bullet$ se puede observar sólo cuando toma valores en A, entonces obtenemos una nueva CM $Y_\bullet = \{Y_k, k = 0, 1, \ldots\}$ definida como sigue. Sea

$$T(0) := \inf\{n \geq 0 \mid \mathrm{X}_n \in A\}$$

y para $k = 0, 1, \ldots$, sea

$$T(k+1) := \inf\{n > T(k) \mid \mathrm{X}_n \in A\}.$$

En particular, si $\mathrm{X}_0 \in A$, entonces $T(0) = 0$ y para $k \geq 1, T(k)$ es el tiempo de la k–ésima visita de $\mathrm{X}_\bullet$ al conjunto A. Para cada $k \geq 0, T(k)$ es un tiempo de paro el cual supondremos que es finito. En este caso, $Y_k := \mathrm{X}_{T(k)}$ $(k = 0, 1, \ldots)$ define una CM con valores en A y probabilidades de transición

$$\mathrm{P}'(x, y) = \mathrm{P}_x(\mathrm{X}_{T(1)} = y) \quad \forall\, x, y \in A, \tag{12.29}$$

en donde $\mathrm{P}_x(\cdot) := \mathrm{P}(\cdot|\mathrm{X}_0 = x)$. En efecto, por la propiedad fuerte de Markov, para todo $x_0, \ldots, x_k$, y en A,

$$\begin{aligned}
\mathrm{P}(Y_{k+1} &= y \mid Y_0 = x_0, \ldots, Y_k = x_k) \\
&= \mathrm{P}(\mathrm{X}_{T(k+1)} = y \mid \mathrm{X}_{T(0)} = x_0, \ldots, \mathrm{X}_{T(k)} = x_k) \\
&= \mathrm{P}(\mathrm{X}_{T(k+1)} = y \mid \mathrm{X}_{T(k)} = x_k) \\
&= \mathrm{P}(\mathrm{X}_{T(1)} = y \mid \mathrm{X}_{T(0)} = x_k) \\
&= \mathrm{P}'(x_k, y),
\end{aligned}$$

es decir, se cumple (29). También se puede demostrar que la matriz de transición $\mathrm{P}' = [\mathrm{P}'(x, y)]$ de $Y_\bullet$ está relacionada con la matriz $\mathrm{P} = [\mathrm{P}(x, y)]$ de $\mathrm{X}_\bullet$ de la siguiente manera: para cada $y \in A$, el vector $\{\mathrm{P}'(x, y), x \in A\}$ es solución de la ecuación

$$\mathrm{P}'(x, y) = \mathrm{P}(x, y) + \sum_{z \notin A} \mathrm{P}(x, z)\mathrm{P}'(z, y). \tag{12.30}$$

(Ejercicio 9.) $\square$

Definición 12.16. (Distribución invariante o estacionaria) Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ una CM (π_0, P) con espacio de estados S, y sea $\pi^* = \{\pi^*(x), x \in S\}$ una distribución de probabilidad sobre S, i.e.

$$\pi^*(x) \geq 0 \quad \forall\, x \in S, \quad \text{y} \quad \sum_{x \in S} \pi^*(x) = 1. \tag{12.31}$$

Decimos que π^* es una **distribución invariante** (o **distribución estacionaria**) para $\mathrm{X}_\bullet$ si

$$\pi^* = \pi^* \mathrm{P}, \tag{12.32}$$

es decir,

$$\pi^*(y) = \sum_{x \in S} \pi^*(x) \mathrm{P}(x,y) \quad \forall\, y \in S. \tag{12.33}$$

El nombre de distribución "invariante" se debe a que si la distribución inicial de $\mathrm{X}_\bullet$ es $\pi_0 = \pi^*$, entonces *la distribución* π_n *de* X_n *es* π^* *para todo* $n = 0, 1, \ldots$ En símbolos:

$$\pi_0 = \pi^* \Rightarrow \pi_n = \pi^* \quad \forall\, n \geq 0. \tag{12.34}$$

Esto se sigue de (32) y (23). En general, una CM puede **no** admitir una distribución invariante.

Por ejemplo, es evidente que la CM definida como $\mathrm{X}_{n+1} := \mathrm{X}_n + 1$ para todo $n \geq 0$, con $\mathrm{X}_0 := 0$, no admite una distribución invariante.

Ejemplo 12.17. CM con dos estados. Sea $\mathrm{X}_\bullet$ la CM del Ejemplo 12.8, cuya matriz de transición P es

$$\begin{array}{cc} & \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{array}{c} 0 \\ \\ 1 \end{array} & \left[\begin{array}{cc} 1-p & p \\ \\ q & 1-q \end{array} \right]. \end{array}$$

De (31) y (33) vemos que $\mathrm{X}_\bullet$ admite una distribución invariante $\pi^* = (\pi^*(0), \pi^*(1))$ si

$$\pi^*(0) = \pi^*(0)(1-p) + \pi^*(1)q, \quad \pi^*(1) = \pi^*(0)p + \pi^*(1)(1-q), \tag{12.35}$$

con $\pi^*(0), \pi^*(1) \geq 0$ y $\pi^*(0) + \pi^*(1) = 1$. En este ejemplo, sustituyendo $\pi^*(1) = 1 - \pi^*(0)$ en (35) se obtiene que $\pi^* = (q/(p+q), p/(p+q))$ es la única distribución invariante de $\mathrm{X}_\bullet$ $\quad \square$.

Ejemplo 12.18. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ una CM cuyo espacio de estados S es un subconjunto de $\mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$. Decimos que $\mathrm{X}_\bullet$ es una **cadena de nacimiento y muerte** (abreviado: cadena de NM) si

$$\mathrm{P}(x,y) \geq 0 \quad \text{sólo si} \quad |x-y| \leq 1.$$

Por ejemplo, supóngase que $S = \{0, 1, \ldots\}$ y

$$\mathrm{P}(x,y) = \begin{cases} q_x & \text{si} \quad y = x-1 \ \ \forall \ \ x \geq 1, \quad \text{con} \ \ q_0 := 0, \\ p_x & \text{si} \ \ y = x+1 \ \ \forall \ \ x \geq 0, \\ r_x & \text{si} \ \ y = x \ \ \forall \ \ x \geq 0, \end{cases}$$

en donde q_x (el parámetro de "muerte"), p_x (el parámetro de "nacimiento") y r_x son números no–negativos y tales que $p_x + q_x + r_x = 1$. (Puesto que $r_x = 1 - p_x - q_x$, para especificar una cadena de NM basta especificar los parámetros p_x y q_x.) Deseamos encontrar condiciones bajo las que $\mathrm{X}_\bullet$ tiene una distribución invariante, para lo cual supondremos que p_x y q_x son números positivos para todo $x \in S$, excepto $q_0 := 0$.

Ahora, para $y = 0$, la ecuación (33) resulta

$$\pi^*(0) = \pi^*(0) r_0 + \pi^*(1) q_1.$$

O bien, como $r_0 + p_0 = 1$,

$$\pi^*(1) q_1 = \pi^*(0) p_0. \tag{12.36}$$

Para $y \geq 1$, de (33) obtenemos

$$\pi^*(y) = \pi^*(y-1) p_{y-1} + \pi^*(y+1) q_{y+1} + \pi^*(y) r_y.$$

Equivalentemente, como $r_y = 1 - p_y - q_y$,

$$\pi^*(y+1) q_{y+1} = \pi^*(y) q_y + \pi^*(y) p_y - \pi^*(y-1) p_{y-1} \quad \forall \ y \geq 1. \tag{12.37}$$

Por otra parte, de (36) y (37), un argumento de inducción da que

$$\pi^*(y) q_y = \pi^*(y-1) p_{y-1} \quad \forall \ y \geq 1,$$

es decir,

$$\pi^*(y) = \pi^*(y-1) p_{y-1} / q_y \quad \forall \ y \geq 1,$$

e iterando esta relación vemos que

$$\pi^*(y) = \pi^*(0) \alpha(y) \quad \forall y \geq 0, \tag{12.38}$$

donde $\alpha(0) := 1$ y

$$\alpha(y) := (p_0 p_1 \cdots p_{y-1}) / (q_1 q_2 \cdots q_y) \quad \text{si} \quad y \geq 1.$$

Resumiendo, hemos demostrado que la ecuación (33) implica (38) y, recíprocamente, se puede ver que (38) implica (33). Por lo tanto, para ver que $\pi^* = \{\pi^*(y),\, y \in S\}$ es una distribución invariante sólo faltaría verificar que π^* es una distribución de probabilidad, i.e. $\pi^*(y) \geq 0$ para todo $y \in S$, y $\sum_y \pi^*(y) = 1$. La primera de estas condiciones es obvia. Con respecto a la segunda condición nótese que

$$\sum_{y=0}^{\infty} \pi^*(y) = \pi^*(0)\overline{\alpha}, \quad \text{con} \quad \overline{\alpha} := \sum_{y=0}^{\infty} \alpha(y) = 1 + \sum_{y=1}^{\infty} \alpha(y). \tag{12.39}$$

Si $\overline{\alpha} = \infty$, es evidente que (usando la convención $0 \cdot \infty = 0$)

$$\sum_{y=0}^{\infty} \pi^*(y) = \begin{cases} 0 & \text{si} \quad \pi^*(0) = 0, \\ \infty & \text{si} \quad \pi^*(0) > 0, \end{cases}$$

de modo que π^* **no** es una distribución de probabilidad. Por otra parte, si $\overline{\alpha} < \infty$ o, equivalentemente,

$$\sum_{y=1}^{\infty} \alpha(y) < \infty, \tag{12.40}$$

vemos de (39) que

$$\sum_{y=0}^{\infty} \pi^*(y) = 1 \quad \text{ssi} \quad \pi^*(0) = 1/\overline{\alpha}.$$

Por lo tanto, concluimos que $X_\bullet$ tiene una (única) distribución invariante si y sólo si se cumple (40), en cuyo caso tal distribución está dada por (38), i.e.

$$\pi^*(y) = \alpha(y)/\overline{\alpha} \quad \forall\, y \geq 0. \tag{12.41}$$

Caso especial: parámetros de NM constantes. Supóngase que

$$p_x = p \quad \text{y} \quad q_x = q \quad \forall\, x \in S, \quad \text{excepto} \quad q_0 := 0. \tag{12.42}$$

En este caso, $\alpha(y) = (p/q)^y$ para todo $y \geq 0$; por lo tanto, (40) se cumple ssi $p < q$. En otras palabras, para una cadena de NM que satisface (42), la condición $p < q$ es necesaria y suficiente para que la cadena tenga una distribución invariante. $\Box$

Ejercicios § 12

12.1. Demuestre que si $\mathrm{X} \in L_1$ es una v.a. discreta con valores enteros no–negativos, entonces

$$E\mathrm{X} = \sum_{n=0}^{\infty} \mathrm{P}(\mathrm{X} > n) = \sum_{n=1}^{\infty} \mathrm{P}(\mathrm{X} \geq n).$$

12.2. Todos los eventos que se mencionan a continuación pertenecen a la misma σ–álgebra. Demuestre:

(a) $\mathrm{P}(A_1 \cap \cdots \cap A_n | B) = \mathrm{P}(A_1|B)\mathrm{P}(A_2|B \cap A_1) \cdots \mathrm{P}(A_n|B \cap A_1 \cap \cdots \cap A_{n-1})$.

(b) Si $B_1, B_2, \ldots$ son eventos ajenos tales que $\mathrm{P}(A|B_i) = p$ para todo $i = 1, 2, \ldots$, entonces $\mathrm{P}(A| \cup_i B_i) = p$.

(c) Si $A_1, A_2, \ldots$ son eventos ajenos, entonces $\mathrm{P}(\cup_i A_i|B) = \sum_i \mathrm{P}(A_i|B)$.

(d) Si $E_1, E_2, \ldots$ forman una partición de Ω, entonces

$$\mathrm{P}(A|B) = \sum_i \mathrm{P}(E_i|B)\mathrm{P}(A|B \cap E_i).$$

12.3. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, \, n \geq 0\}$ una CM con matriz de transición P, y sea $Y_n := \mathrm{X}_{kn}$ para algún entero $k > 0$. Demuestre que $Y_\bullet = \{Y_n\}$ es una CM con matriz de transición P^k.

12.4. Considere una CM finita o numerable con espacio de estados S y matriz de transición P. Demuestre que una distribución de probabilidad π sobre S es invariante para P ssi $\pi(I - P + E) = e$, en donde I es la matriz identidad, $E = (e_{ij})$ es la matriz con $e_{ij} = 1$ para todo $i, j \in S$, y $e = (e_i, i \in S)$ es el vector con $e_i = 1$ para todo $i \in S$.

12.5. Sea P la matriz de transición de una CM con espacio de estados S numerable, y sea π una distribución sobre S. Se dice que π y P **están en balance** (o que satisfacen las **ecuaciones de balance**) si

$$\pi_i \mathrm{P}(i,j) = \pi_j \mathrm{P}(j,i) \quad \forall \, i, j \in S.$$

Demuestre que si π y P están en balance, entonces π es invariante para P.

12.6. Considérese la CM $\mathrm{X}_n = \xi_0 + \cdots + \xi_n$ del Ejemplo 12.3 excepto que la densidad común del "ruido" $\{\xi_n\}$ es $f(k) = \mathrm{P}(\xi_0 = k)$ para cualquier entero k. Calcule las probabilidades de transición.

12.7. Demuestre la Proposición 12.5. (*Sugerencia*: use el Ejercicio 12.2.)

12.8. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ una CM con matriz de transición P. Sea $n \geq 0$ un entero arbitrario y defina $Y_m := \mathrm{X}_{n+m}$ para $m = 0, 1, \ldots$. Demuestre: dado que $\mathrm{X}_n = x$, el PE $Y_\bullet = \{Y_m\}$ es una CM con estado inicial $Y_0 = x$ y matriz de transición P y, además, $Y_\bullet$ es independiente de las v.a. $\mathrm{X}_0, \ldots, \mathrm{X}_n$; es decir, si A es cualquier evento determinado por las vv.aa. $\mathrm{X}_0, \ldots, \mathrm{X}_{n-1}$, entonces

$$\begin{aligned}
&\mathrm{P}(\{\mathrm{X}_n = y_n, \ldots, \mathrm{X}_{n+m} = y_{n+m}\} \cap A | \mathrm{X}_{n-1} = x) \\
&= \mathrm{P}(\mathrm{X}_n = y_n, \ldots, \mathrm{X}_{n+m} = y_{n+m} | \mathrm{X}_n = x) \mathrm{P}(A | \mathrm{X}_{n-1} = x).
\end{aligned}$$

12.9. En el Ejemplo 15, demuestre que se satisface la ecuación (30).

12.10. Una cola a tiempo discreto. Considérese un sistema de espera con un servidor. Los "clientes" o "trabajos" llegan al sistema de acuerdo con un proceso de arribos $\{Y_0, Y_1, \ldots\}$, en donde Y_n es el número de arribos en el periodo $[n, n+1)$ y los cuales empezarán a ser atendidos, si el servidor está disponible, a partir del siguiente periodo. Suponga que las vv.aa. Y_n son i.i.d. con distribución

$$\mathrm{P}(Y_0 = k) = a_k \quad \text{para} \quad k = 0, 1, \ldots.$$

En cada periodo $[n, n+1)$, el servidor puede atender sólo un cliente a la vez y la "completación de servicio" es una v.a. Bernoulli S_n con parámetro p; es decir, $S_n = 1$ (con probabilidad p) si se completa el servicio, en cuyo caso el cliente servido abandona el sistema, y $S_n = 0$ (con probabilidad $q := 1 - p$) si no se completa el servicio, en cuyo caso el cliente continúa su servicio en el siguiente periodo. Supóngase que los PEs $\{Y_n\}$ y $\{S_n\}$ son independientes entre sí, y también independientes del número inicial, X_0, de clientes en el sistema. Sea X_n el número de clientes en el sistema al inicio del periodo $[n, n+1)$. Nótese que el PE $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ satisface que

$$\mathrm{X}_{n+1} = \begin{cases} Y_n & \text{si} \quad \mathrm{X}_n = 0, \\ \mathrm{X}_n + Y_n - S_n & \text{si} \quad \mathrm{X}_n > 0. \end{cases}$$

Demuestre que $X_\bullet$ es una CM con valores en $\{0, 1, \ldots\}$ y probabilidades de transición

$$P(0, y) = a_y$$

y para $x > 0$:

$$P(x, y) = \begin{cases} 0 & \text{si} \quad y < x - 1, \\ p \cdot a_0 & \text{si} \quad y = x - 1, \\ p \cdot a_{y-x+1} + q \cdot a_{y-x} & \text{si} \quad y \geq x. \end{cases}$$

12.11. En el contexto del Teorema 12.13, demuestre que condicionado a que $X_T = x$, el PE $Y_m := X_{T+m}(m = 0, 1, \ldots)$ es una CM con estado inicial x y la misma matriz de transición $P = [P(x, y)]$ de $X_\bullet$; además, $Y_\bullet = \{Y_0, Y_1, \ldots\}$ es independiente de $X_0, \ldots, X_T$. (*Sugerencia*: use (27) y (28).) Compare este ejercicio con el Ejercicio 8.

12.12. Distribución límite. Sea $X_\bullet = \{X_n\}$ una CM (π_0, P) con espacio de estados S numerable. Se dice que $X_\bullet$ tiene una **distribución límite** $\pi^* = \{\pi^*(y), y \in S\}$ si π^* es una distribución de probabilidad sobre S tal que

$$\lim_{n \to \infty} P_n(x, y) = \pi^*(y) \quad \forall\, x, y \in S.$$

(a) Demuestre que la CM definida como $X_{n+1} = X_n + 1$ para todo n, *no* tiene una distribución límite.

(b) Para cada $n = 0, 1, \ldots$, sea π_n la distribución de X_n. Demuestre que si $X_\bullet$ tiene una distribución límite π^*, entonces

$$\lim_{n \to \infty} \pi_n(y) = \pi^*(y) \quad \forall\, y \in S.$$

(*Sugerencia:* use (22) y el Teorema de Convergencia Dominada.) Si, además, $X_\bullet$ es una CM finita, entonces π^* es una distribución invariante para $X_\bullet$.

(c) Calcule la distribución límite de la CM en el Ejemplo 12.8 y compárela con el resultado del Ejemplo 12.17.

12.13. La cadena de Ehrenfest. Se tienen d esferas numeradas $1, \ldots, d$, distribuidas en dos cajas marcadas 1 y 2. Sea X_0 el número de esferas en la caja 1 en el tiempo 0. En cada tiempo $n = 1, 2, \ldots$ se selecciona al azar un número del conjunto $\{1, \ldots, d\}$ y la esfera marcada con ese número se saca de la caja en que se encuentra y se coloca en la otra caja. Sea X_n el número de esferas en la caja 1 al tiempo n. Demuestre que $\{X_n\}$ es una CM con espacio de estados $S = \{0, 1, \ldots, d\}$ y probabilidades de transición

$$\mathrm{P}(x, y) = \begin{cases} x/d & \text{si} \quad y = x - 1, \\ 1 - x/d & \text{si} \quad y = x + 1, \\ 0 & \text{en c.c.} \end{cases}$$

Además, demuestre que:

(a) $E(X_{n+1}|X_n) = \alpha X_n + 1$, con $\alpha = (d-2)/d$, y

$$EX_n = \frac{1}{1-\alpha} + \alpha^n \left(EX_0 - \frac{1}{1-\alpha} \right) \quad \forall\, n = 0, 1, \ldots.$$

(b) Si $X_0 \sim \mathrm{Bin}(d, 1/2)$, entonces X_1 tiene la misma distribución que X_0.

Nota: La cadena de Ehrenfest es un modelo de difusión de un gas a través de una membrana porosa.

12.14. Considere una CM con espacio de estados $S = \{0, 1, 2\}$ y matriz de transición

$$\mathrm{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1-p & 0 & p \\ 0 & 1 & 0 \end{bmatrix}$$

(a) Encuentre P_2. (b) Demuestre que $\mathrm{P}_4 = \mathrm{P}_2$. (c) Encuentre P_n para todo $n \geq 1$.

12.15. Considere una CM con probabilidades de transición

$$\mathrm{P}(x, y) = \begin{cases} p & \text{si} \quad y = x + 1, \\ 1-p & \text{si} \quad y = 0 \end{cases}$$

para todo $x = 0, 1, \ldots$. Demuestre que existe, y calcule, una distribución invariante.

12.16. Considere una CM con probabilidades de transición $\mathrm{P}(x, y) = \alpha_y$ para todo x, y. Demuestre que la CM tiene una única distribución invariante dada por $\pi^*(y) = \alpha_y$ para todo y.

12.17. Demuestre que una combinación convexa de distribuciones invariantes es una distribución invariante. Es decir, si $\pi = \{\pi(x), x \in S\}$ y $\pi' = \{\pi'(x), x \in S\}$ son distribuciones invariantes y $\alpha \in [0,1]$, entonces

$$\pi_\alpha(x) := (1-\alpha)\pi(x) + \alpha\pi'(x) \quad \forall\, x \in S,$$

es una distribución invariante.

12.18. Considere la cadena de NM del Ejemplo 12.18 con parámetros $p_0 = 1$ y

$$p_x = p > 0, \quad q_x = q = 1-p \quad \forall\, x \geq 1.$$

Encuentre la distribución invariante, si es que existe.

12.19. Procesos de ramificación. Un proceso de ramificación es una CM $\mathrm{X}_\bullet = \{\mathrm{X}_n,\, n = 0, 1, \ldots\}$ en donde X_n representa el número de individuos (o "partículas") de una cierta población al tiempo n. Cada individuo da origen a un número aleatorio Y de individuos en la siguiente generación, donde Y es una v.a. discreta con densidad $f(y) = \mathrm{P}(Y = y)$ para $y = 0, 1, \ldots$. Luego, si $\mathrm{X}_n = x$, entonces $\mathrm{X}_{n+1} = Y_1 + \cdots + Y_x$, donde $Y_1, Y_2, \ldots$ son i.i.d. con densidad común f, y las probabilidades de transición son

$$\mathrm{P}(x, y) = \mathrm{P}(\mathrm{X}_{n+1} = y | \mathrm{X}_n = x) = \mathrm{P}(Y_1 + \cdots + Y_x = y).$$

En general, dado el número inicial X_0 de individuos, el proceso de ramificación se puede describir como $\mathrm{X}_{n+1} = Y_1 + Y_2 + \cdots + Y_{\mathrm{X}_n}$ para todo $n = 0, 1, \ldots$.

Supóngase que Y tiene media μ y varianza σ^2 finita. Demuestre:

(a) $E(\mathrm{X}_{n+1}|\mathrm{X}_n) = \mu \mathrm{X}_n$, y $E\mathrm{X}_n = \mu^n E\mathrm{X}_0$.

(b) $E(\mathrm{X}_{n+1}^2|\mathrm{X}_n) = \sigma^2 \mathrm{X}_n + \mu^2 \mathrm{X}_n^2$, y $E\mathrm{X}_{n+1}^2 = \mu^n \sigma^2 E\mathrm{X}_0 + \mu^2 E\mathrm{X}_n^2$.

(c) Si $\mathrm{X}_0 = x$, entonces

$$E\mathrm{X}_n^2 = x\sigma^2[\mu^{n-1} + \cdots \mu^{2(n-1)}] + x^2\mu^{2n} \quad \forall\, n \geq 1.$$

(d) Si $\mathrm{X}_0 = 1$ y $Y_n := \mathrm{X}_n / E(\mathrm{X}_n)$, para $n = 0, 1, \ldots$, entonces $E[Y_{n+1}|Y_n] = Y_n$.

13 Clasificación de estados de una CM

Contenido: Conjunto cerrado, estados comunicantes, estado absorbente, cadena irreducible, estado transitorio, estado recurrente, recurrente positivo y recurrente nulo.

En esta sección $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ es una CM con espacio de estados S a lo más numerable.

Sean x, y dos estados. Decimos que x **se comunica con** y (notación: $x \to y$) si existe un entero $n \geq 0$ tal que $\mathrm{P}_n(x, y) > 0$. Si $x \to y$ y $y \to x$ se dice que x, y son estados **comunicantes**, y escribimos $x \leftrightarrow y$.

Proposición 13.1. Las siguientes afirmaciones son equivalentes para estados distintos x, y:

(a) $x \to y$.

(b) Existe $n \geq 0$ y estados $x_0, \ldots, x_n$, con $x_0 = x$ y $x_n = y$ tales que

$$\mathrm{P}(x_0, x_1)\mathrm{P}(x_1, x_2) \cdots \mathrm{P}(x_{n-1}, x_n) > 0.$$

(c) $\mathrm{P}_x(\mathrm{X}_n = y \text{ para algún } n \geq 0) > 0$. (Recuérdese que $\mathrm{P}_x(A) := \mathrm{P}(A|\mathrm{X}_0 = x)$ para cualquier evento A.)

Demostración. Por la ecuación de Chapman–Kolmogorov,

$$\mathrm{P}_n(x, y) = \sum_{x_1} \mathrm{P}(x, x_1)\mathrm{P}^{n-1}(x_1, y) = \sum_{x_1}\sum_{x_2} \mathrm{P}(x, x_1)\mathrm{P}(x_1, x_2)\mathrm{P}^{n-2}(x_2, y);$$

en general

$$\mathrm{P}_n(x, y) = \sum_{x_1} \cdots \sum_{x_{n-1}} \mathrm{P}(x, x_1)\mathrm{P}(x_1, x_2) \cdots \mathrm{P}(x_{n-1}, y).$$

De aquí se sigue la equivalencia de (a) y (b). Por otra parte,

$$\mathrm{P}_n(x, y) \leq \mathrm{P}_x(\mathrm{X}_k = y \;\text{ para algún }\; k \geq 0) \leq \sum_{k=0}^{\infty} \mathrm{P}_k(x, y),$$

lo cual implica la equivalencia de (b) y (c). □ □

Proposición 13.2. La relación $\leftrightarrow$ es de equivalencia, es decir, es reflexiva ($x \leftrightarrow x$), simétrica ($x \leftrightarrow y$ ssi $y \leftrightarrow x$) y transitiva ($x \leftrightarrow y$ e $y \leftrightarrow z$, entonces $x \leftrightarrow z$).

Como la relación $\leftrightarrow$ es de equivalencia, define una partición de S en conjuntos ajenos llamados **clases comunicantes**: dos estados pertenecen a la misma clase ssi son estados comunicantes. Decimos que una clase C es **cerrada** si

$$x \in C,\ x \to y \quad \text{implica} \quad y \in C.$$

Equivalentemente, C es una clase cerrada si $\mathrm{P}(x, y) = 0$ para todo $x \in C$, $y \notin C$. En particular, si x es un estado **absorbente**, es decir $\mathrm{P}(x, x) = 1$, entonces $\{x\}$ es una clase cerrada.

Ejemplo 13.3. Considérese una CM con $S = \{1, 2, 3, 4\}$ y matriz de transición

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \surd & \surd & \surd & 0 \\ 0 & 0 & \surd & \surd \\ 0 & 0 & \surd & \surd \end{bmatrix}$$

en donde $\surd$ representa números positivos. Las clases comunicantes son $C_1 = \{1\}, C_2 = \{2\}$ y $C_3 = \{3, 4\}$, y C_1 y C_3 son cerradas. $\Box$

Una CM en la que S es la única clase comunicante se dice que es **irreducible**. La CM en el ejemplo anterior no es irreducible. La CM con dos estados y la cadena de NM en los Ejemplos 12.17 y 12.18, respectivamente, sí son irreducibles. Otra CM irreducible es aquella que tiene matriz de transición de la forma

$$\begin{bmatrix} 0 & \surd & 0 & \surd \\ \surd & \surd & 0 & 0 \\ 0 & \surd & \surd & \surd \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Definición 13.4. Sea f_{xy} la probabilidad de ir de x a y en un tiempo finito, i.e.

$$f_{xy} := \mathrm{P}_x(\mathrm{X}_n = y \quad \text{para algún} \quad n \geq 1). \tag{13.1}$$

Decimos que x es **recurrente** (o **persistente**) si $f_{xx} = 1$, y **transitorio** en caso contrario, i.e. $f_{xx} < 1$.

La probabilidad f_{xy} en (1) se puede escribir en varias formas equivalentes. En efecto, como en el Ejemplo 12.12(b), sea T_y el **tiempo del primer paso** al estado y, i.e.

$$T_y := \inf\{n \geq 1 | \mathrm{X}_n = y\}.$$

Entonces (1) se puede escribir como

$$f_{xy} = \mathrm{P}_x(T_y < \infty) = \sum_{n=1}^{\infty} \mathrm{P}_x(T_y = n). \tag{13.2}$$

Por otra parte, si $N(y)$ es el **número total de visitas** al estado y, i.e.

$$N(y) := \sum_{n=1}^{\infty} I_y(\mathrm{X}_n), \tag{13.3}$$

en donde $I_y(x) := \delta_{xy}$ es la función indicadora de $\{y\}$, entonces

$$f_{xy} = \mathrm{P}_x(N(y) \geq 1). \tag{13.4}$$

También conviene considerar el **número esperado de visitas** a y, dado el estado inicial x, definido como

$$V(x, y) := E_x[N(y)].$$

Por (3),

$$V(x, y) = \sum_{n=1}^{\infty} E_x[I_y(\mathrm{X}_n)] = \sum_{n=1}^{\infty} \mathrm{P}_n(x, y). \tag{13.5}$$

Teorema 13.5. *(a) Las siguientes afirmaciones* $(\mathrm{a}_1) - (\mathrm{a}_4)$ *son equivalentes:*

(a_1) *y es recurrente.*

(a_2) $f_{yy} = \mathrm{P}_y(T_y < \infty) = 1.$

(a_3) $\mathrm{P}_y(N(y) = \infty) = 1.$

(a_4) $V(y, y) = \infty.$

(b) Las afirmaciones (b_1)*–*(b_4) *son equivalentes:*

(b_1) *y es transitorio.*

(b$_2$) $f_{yy} = \mathrm{P}_y(T_y < \infty) < 1$.

(b$_3$) $\mathrm{P}_y(N(y) = \infty) = 0$.

(b$_4$) $V(y, y) < \infty$.

En particular, cualquier estado es transitorio o es recurrente.

La equivalencia de (a$_1$)–(a$_2$) y de (b$_1$)–(b$_2$) se sigue de (2) y de la Definición 13.4. Para completar la demostración del Teorema 13.5 necesitamos varios resultados preliminares.

Lema 13.6. Para todo $n \geq 1$:

(a) $\mathrm{P}_x(N(y) \geq n) = f_{xy} f_{yy}^{n-1}$; por lo tanto,

(b) $\mathrm{P}_x(N(y) = n) = \mathrm{P}_x(N(y) \geq n) - \mathrm{P}_x(N(y) \geq n+1) = f_{xy}(1 - f_{yy}) f_{yy}^{n-1}$,

(c) $V(x, y) = E_x[N(y)] = \sum_{n=1}^{\infty} n\mathrm{P}_x[N(y) = n] = f_{xy}/(1 - f_{yy})$.

Luego, si y es **transitorio**, entonces para todo $x \in S$:

(d) $V(x, y) = \dfrac{f_{xy}}{1 - f_{yy}} < \infty$, y

(e) $\mathrm{P}_x(N(y) < \infty) = 1$.

Demostración. Basta demostrar (a), lo cual haremos por inducción. Por (4), (a) se cumple para $n = 1$. Supongamos que (a) se satisface para algún $n \geq 1$. Sea

$$f_{xy}(k) := \mathrm{P}_x(T_y = k) \tag{13.6}$$

y nótese que, por (2),

$$f_{xy} := \sum_{k=1}^{\infty} f_{xy}(k). \tag{13.7}$$

Por lo tanto,

$$\begin{aligned}
\mathrm{P}_x(N(y) \geq n+1) &= \sum_{k=1}^{\infty} \mathrm{P}_x(T_y = k,\ N(y) \geq n+1) \\
&= \sum_{k=1}^{\infty} \mathrm{P}_x(T_y = k)\mathrm{P}(N(y) \geq n+1 | \mathrm{X}_0 = x,\ T_y = k) \\
&= \sum_{k=1}^{\infty} f_{xy}(k)\mathrm{P}_y(N(y) \geq n)
\end{aligned}$$

(por la propiedad fuerte de Markov)

$$
\begin{aligned}
&= f_{yy}^n \sum_{k=1}^{\infty} f_{xy}(k) \qquad \text{[por la hipótesis de inducción]} \\
&= f_{yy}^n f_{xy} \qquad\qquad\quad \text{[por (7)].}
\end{aligned}
$$

Esto demuestra (a) para $n+1$. □ □

Lema 13.7. Para todo $|z| < 1$, considere las "funciones generadoras"

$$
\begin{aligned}
\mathrm{P}_{xy}(z) &:= \sum_{n=0}^{\infty} z^n \mathrm{P}_n(x,y) = \delta(x,y) + \sum_{n=1}^{\infty} z^n \mathrm{P}_n(x,y), \\
F_{xy}(z) &:= \sum_{n=1}^{\infty} z^n f_{xy}(n),
\end{aligned}
$$

con $f_{xy}(n)$ como en (6). Entonces para todo $x, y \in S$ y $|z| < 1$:

$$
\mathrm{P}_{xy}(z) = \delta(x,y) + F_{xy}(z)\mathrm{P}_{yy}(z). \tag{13.8}
$$

Por lo tanto

(a) $\mathrm{P}_{xy}(z) = F_{xy}(z)\mathrm{P}_{yy}(z)$ si $x \neq y$,

(b) $\mathrm{P}_{yy}(z) = 1 + F_{yy}(z)\mathrm{P}_{yy}(z)$, i.e. $\mathrm{P}_{yy}(z) = [1 - F_{yy}(z)]^{-1}$.

Demostración. Como

$$
\mathrm{P}_n(x,y) = \mathrm{P}_x(\mathrm{X}_n = y) = \sum_{k=1}^{n} \mathrm{P}_x(T_y = k,\, \mathrm{X}_n = y),
$$

de la propiedad fuerte de Markov se sigue que

$$
\mathrm{P}_n(x,y) = \sum_{k=1}^{n} f_{xy}(k)\mathrm{P}_{n-k}(y,y) \quad \forall\, n \geq 1. \tag{13.9}
$$

Por lo tanto,

$$\begin{aligned}
\sum_{n=1}^{\infty} z^n \mathrm{P}_n(x,y) &= \sum_{n=1}^{\infty} z^n \sum_{k=1}^{n} f_{xy}(k) \mathrm{P}_{n-k}(y,y) \\
&= \sum_{k=1}^{\infty} f_{xy}(k) \sum_{n=k}^{\infty} z^n \mathrm{P}_{n-k}(y,y) \\
&= \sum_{k=1}^{\infty} z^k f_{xy}(k) \sum_{n=k}^{\infty} z^{n-k} \mathrm{P}_{n-k}(y,y) \\
&= F_{xy}(z) \mathrm{P}_{yy}(z). \quad \square
\end{aligned}$$

□

También usaremos el siguiente resultado de Análisis Matemático.

Lema 13.8. (Teorema de Abel) Sea $\{a_n\}$ una sucesión de números no negativos y sea

$$A(z) := \sum_{n=0}^{\infty} z^n a_n \quad \text{para} \quad |z| < 1.$$

Entonces

$$\lim_{z \uparrow 1} A(z) = \sum_{n=0}^{\infty} a_n.$$

Con estos preliminares podemos completar la demostración del Teorema 13.5 como sigue.

Demostración. Demostración del Teorema 13.5

(a) Ya mencionamos que (a$_1$) y (a$_2$) son equivalentes. Por otra parte, se sigue del Lema 13.6(a) que

$$\mathrm{P}_x(N(y) = \infty) = \lim_{n \to \infty} \mathrm{P}_x(N(y) \geq n) = f_{xy} \cdot \lim_{n \to \infty} f_{yy}^{n-1}. \tag{13.10}$$

En particular, con $x = y$ vemos que (a$_2$) y (a$_3$) son equivalentes. Ahora, considere las funciones generadoras $\mathrm{P}_{xy}(z)$ y $F_{xy}(z)$ del Lema 13.7. Por (5), (7) y el Lema 4.8,

$$\lim_{z \uparrow 1} \mathrm{P}_{xy}(z) = \delta(x,y) + V(x,y) \tag{13.11}$$

y

$$\lim_{z \uparrow 1} F_{xy}(z) = f_{xy}. \tag{13.12}$$

En particular, tomando $x = y$ y usando la igualdad $\mathrm{P}_{yy}(z) = [1 - F_{yy}(z)]^{-1}$ en el Lema 13.7(b) obtenemos la equivalencia de (a_2) y (a_4).

(b) Ya mencionamos la equivalencia de (b_1) y (b_2), mientras que la de (b_2) y (b_4) se sigue de (11)–(12) con $x = y$. Finalmente, la equivalencia de (b_2) y (b_3) se obtiene de (10). $\square$ $\square$

Corolario 13.9. *Si y es recurrente, entonces*

(a) $\mathrm{P}_x(N(y) = \infty) = f_{xy} = \mathrm{P}_x(T_y < \infty) \;\; \forall\, x \in S$.

(b) $$V(x,y) = 0 \quad \textit{si } f_{xy} = 0, \\ \qquad\; = \infty \quad \textit{si } f_{xy} > 0.$$

Si además $y \leftrightarrow x$, entonces

(c) $f_{xy} = f_{yx} = 1$, *y*

(d) *x es recurrente; esto implica, en particular, que si la CM es irreducible, entonces todos los estados son recurrentes o todos son transitorios.*

Demostración. Supóngase que y es recurrente. Entonces $f_{yy} = 1$, así que (a) se sigue de (10).

Para demostrar (b) supóngase primero que $f_{xy} = 0$. Luego, de (6) y (7) vemos que $f_{xy}(k) = 0$ para todo $k \geq 1$, lo cual implica, por (9), que $\mathrm{P}_n(x,y) = 0$ para todo $n \geq 1$. Por lo tanto, de (5) concluímos que $V(x,y) = 0$. Por otra parte, si $f_{xy} > 0$, vemos de (10) que $\mathrm{P}_x(N(y) = \infty) > 0$, que a su vez implica

$$V(x,y) = E_x[N(y)] = \infty. \tag{13.13}$$

(c) Si $y \leftrightarrow x$, entonces $f_{xy} = \mathrm{P}_x(N(y) \geq 1) > 0$. Además, $\mathrm{P}_y(N(y) \geq n) = 1$. Con esto, podemos usar la demostración del Lema 13.6 para mostrar que $f_{xy} = 1$.

(d) Sean n y m tales que $\mathrm{P}_n(x,y) > 0$ y $\mathrm{P}_m(y,x) > 0$. Entonces para cualquier entero $r \geq 1$, $\mathrm{P}_{n+r+m}(x,x) \geq \mathrm{P}_n(x,y)\mathrm{P}_r(y,y)\mathrm{P}_m(y,x)$ de modo que

$$\sum_{r=1}^{\infty} \mathrm{P}_{n+r+m}(x,x) \geq \mathrm{P}_n(x,y)\mathrm{P}_m(y,x) \sum_{r=1}^{\infty} \mathrm{P}_r(y,y) = \infty. \tag{13.14}$$

Esto implica que $V(x,x) = \infty$ y por el Teorema 13.5(a) concluimos que x es recurrente. $\square$ $\square$

Corolario 13.10. *Sea C una clase comunicante. Entonces:*

(a) *Todos los estados en C son recurrentes o todos son transitorios. En el primer caso se dice que C es una clase recurrente, y en el segundo que C es transitoria.*

(b) *Si C es recurrente entonces es cerrada.*

(c) *Si C es una clase cerrada finita, entonces C es recurrente.*

Demostración. El inciso (a) es consecuencia del Corolario 13.9(d) o de la desigualdad en (14) que se puede escribir como

$$\sum_{r=1}^{\infty} \mathrm{P}_r(y,y) \leq \frac{1}{\mathrm{P}_n(x,y)\mathrm{P}_m(y,x)} \sum_{r=1}^{\infty} \mathrm{P}_{n+r+m}(x,x).$$

(b) Supóngase que C es recurrente pero no es cerrada. Entonces existen $x \in C$, $y \notin C$ y $m \geq 1$ tales que $\mathrm{P}_x(\mathrm{X}_m = y) > 0$. Entonces, como

$$\mathrm{P}_x(\{\mathrm{X}_m = y\} \cap \{N(x) = \infty\}) = 0,$$

se obtiene que $\mathrm{P}_x(N(x) = \infty) < 1$ y, por lo tanto, x no es recurrente, lo cual implica que tampoco C es recurrente.

(c) Si C es una clase cerrada finita, entonces contiene al menos un estado recurrente. Luego, por (a), C es recurrente. $\square$ $\square$

Definimos el **tiempo medio de retorno al estado** x (o **tiempo medio de recurrencia** de x) como

$$m_x := E_x(T_x) = \sum_{n=1}^{\infty} n\mathrm{P}_x(T_x = n). \tag{13.15}$$

Por el Teorema 13.5(b$_2$), si x es transitorio entonces $f_{xx} = \mathrm{P}_x(T_x < \infty) < 1$ ó, equivalentemente, $\mathrm{P}_x(T_x = \infty) > 0$. Por lo tanto, si x es transitorio, necesariamente $m_x = \infty$. Sin embargo, si x es recurrente puede ocurrir que (a) $m_x < \infty$, ó (b) $m_x = \infty$. En el caso (a) se dice que x es **recurrente positivo**, y en el caso (b) que x es **recurrente nulo**.

El siguiente resultado es útil para calcular m_x, en particular.

Proposición 13.11. Sean x, y estados arbitrarios. Entonces

(a) $\mathrm{P}_x(T_y = 1) = \mathrm{P}(x, y)$, y para $n \geq 2$

$$\mathrm{P}_x(T_y = n) = \sum_{z \neq y} \mathrm{P}(x, z)\mathrm{P}_z(T_y = n - 1); \tag{13.16}$$

en particular,

$$\mathrm{P}_x(T_x = n) = \sum_{z \neq x} \mathrm{P}(x, z)\mathrm{P}_z(T_x = n - 1). \tag{13.17}$$

(b) $E_x(T_y) = \mathrm{P}(x, y) + \sum_{z \neq y} \mathrm{P}(x, z)[f_{zy} + E_z(T_y)]$; en particular, para $x = y$,

$$m_x = \mathrm{P}(x, x) + \sum_{z \neq x} \mathrm{P}(x, z)[f_{zx} + E_z(T_x)]. \tag{13.18}$$

Demostración. (a) Para cualquier $n \geq 2$,

$$\mathrm{P}_x(T_y = n) = \sum_{z \neq y} \mathrm{P}_x(\mathrm{X}_1 = z,\, T_y = n). \tag{13.19}$$

Pero, para $z \neq y$

$$\begin{aligned}
\mathrm{P}_x(\mathrm{X}_1 = z,\, T_y = n) &= \mathrm{P}_x(\mathrm{X}_1 = z)\mathrm{P}(T_y = n | \mathrm{X}_0 = x,\, \mathrm{X}_1 = z) \\
&= \mathrm{P}(x, z)\mathrm{P}(T_y = n | \mathrm{X}_1 = z) \\
&= \mathrm{P}(x, z)\mathrm{P}_z(T_y = n - 1).
\end{aligned}$$

Sustituyendo este resultado en (19) obtenemos (16).

(b) Nótese que

$$\begin{aligned}
\sum_{n=2}^{\infty} n\, \mathrm{P}_z(T_y = n - 1) &= \sum_{n=1}^{\infty} (1 + n)\mathrm{P}_z(T_y = n) \\
&= \sum_{n=1}^{\infty} \mathrm{P}_z(T_y = n) + \sum_{n=1}^{\infty} n\, \mathrm{P}_z(T_y = n) \\
&= f_{zy} + E_z(T_y). \qquad (13.20)
\end{aligned}$$

Luego

$$\begin{aligned}
E_x(T_y) &= \sum_{n=1}^{\infty} n\,\mathrm{P}_x(T_y = n) \\
&= \mathrm{P}(x,y) + \sum_{n=2}^{\infty} n\,\mathrm{P}_x(T_y = n) \\
&= \mathrm{P}(x,y) + \sum_{n=2}^{\infty} n \sum_{z \neq y} \mathrm{P}(x,z)\mathrm{P}_z(T_y = n-1) \quad \text{[por (16)]} \\
&= \mathrm{P}(x,y) + \sum_{z \neq y} \mathrm{P}(x,z) \sum_{n=2}^{\infty} n\,\mathrm{P}_z(T_y = n-1),
\end{aligned}$$

y usando (20) se obtiene (b). $\square$ $\square$

Los ejemplos siguientes ilustran el uso de los resultados en la Proposición 13.11.

Ejemplo 13.12. Caminata aleatoria. Considérese la caminata aleatoria del Ejemplo 12.3, i.e.

$$\mathrm{X}_{n+1} = \mathrm{X}_n + Y_{n+1} \quad \forall\, n = 0, 1, \ldots, \;\; \text{con} \;\; \mathrm{X}_0 = Y_0,$$

en donde $Y_\bullet = \{Y_n\}$ es una sucesión de vv.aa. i.i.d. con distribución

$$\mathrm{P}(Y_0 = 1) = p, \quad \mathrm{P}(Y_0 = -1) = q = 1 - p, \quad \text{para} \quad 0 < p < 1.$$

El espacio de estados S de $\mathrm{X}_\bullet$ es el conjunto de todos los números enteros y las probabilidades de transición son

$$\mathrm{P}(x,y) = \begin{cases} p & \text{si} \quad y = x+1, \\ q & \text{si} \quad y = x-1. \end{cases} \tag{13.21}$$

(Equivalentemente, $\mathrm{X}_\bullet$ es una cadena de NM con parámetros $p_x = p$, $q_x = q$ y $r_x = 0$; vea el Ejemplo 12.18). Como p y q son ambos positivos, $\mathrm{X}_\bullet$ es irreducible. Por lo tanto, por el Corolario 13.9(d) todos los estados son recurrentes o todos son transitorios. Luego, para ver cuál de estos dos casos ocurre basta considerar un estado cualquiera, digamos $x = 0$. Es decir, por el Teorema 13.5 basta verificar (por ejemplo) si el número esperado de visitas

$$V(0,0) = \sum_{n=1}^{\infty} \mathrm{P}_n(0,0)$$

es finito o infinito. Con este fin, obsérvese primero que $\mathrm{P}_n(0,0) = 0$ si n es *impar*. Por otra parte, si $\mathrm{X}_0 = 0$, entonces $\mathrm{X}_\bullet$ está de nuevo en el estado 0 en $2n$ pasos (o sea, $\mathrm{X}_{2n} = 0$) ssi la CM se mueve n veces "a la derecha" (con probabilidad p en cada paso) y n veces "a la izquierda" (con probabilidad q en cada paso). Luego, $\mathrm{P}_{2n}(0,0)$ se puede escribir como la probabilidad binomial

$$\mathrm{P}_{2n}(0,0) = \binom{2n}{n} p^n q^n,$$

así que $V(0,0) = \sum_{n=1}^{\infty} \mathrm{P}_{2n}(0,0)$. Ahora consideremos el límite

$$\mathrm{P}_{2(n+1)}(0,0)/\mathrm{P}_{2n}(0,0) \to 4pq$$

cuando $n \to \infty$ y observe que $4pq = 4p(1-p) \leq 1$ con igualdad ssi $p = 1/2$. Por lo tanto, si $p \neq 1/2$, entonces $4pq < 1$ y por la "prueba del cociente" para convergencia de series concluimos que $V(0,0) < \infty$; es decir, *la CM es transitoria si* $p \neq 1/2$. Sin embargo, si $p = 1/2$, entonces $4pq = 1$ y no podemos usar la prueba del cociente para ver si $V(0,0)$ converge o diverge. Este caso se puede analizar usando la **fórmula de Stirling**

$$n! \sim (2\pi n)^{1/2}(n/e)^n \quad \text{cuando} \quad n \to \infty,$$

en donde $x_n \sim y_n$ significa que $x_n/y_n \to 1$. Usando dicha formula se obtiene que

$$\binom{2n}{n} \sim 4^n(\pi n)^{-1/2}$$

y por lo tanto

$$\mathrm{P}_{2n}(0,0) \sim (4pq)^n(\pi n)^{-1/2}.$$

De aquí se sigue que $V(0,0)$ es finita o infinita cuando $4pq = 1$ ssi la serie

$$\sum_{n=1}^{\infty}(\pi n)^{-1/2}$$

converge o diverge, respectivamente. Luego, como es evidente que esta serie diverge, concluimos que $V(0,0) = \infty$ cuando $p = 1/2$. En conclusión, la caminata aleatoria $\mathrm{X}_\bullet$ es recurrente si $p = 1/2$ y transitoria si $p \neq 1/2$. Cuando $p = 1/2$ se dice que la caminata aleatoria es **simétrica**.

Ahora deseamos ver si la caminata aleatoria simétrica (por lo tanto, recurrente) es recurrente positiva o recurrente nula. Por el Ejercicio 4, basta ver si uno de los estados, por ejemplo $x = 0$, es recurrente positivo o nulo.

Anteriormente obtuvimos la identidad $P_{00}(z) = 1+F_{00}(z)P_{00}(z)$, de donde $F_{00} = 1 - 1/P_{00}(z)$, pero

$$P_{00}(z) = \sum_{n=0}^{\infty} z^n P_n(0,0) = \sum_{n=0}^{\infty} z^{2n}\binom{2n}{n}p^n q^n = \sum_{n=0}^{\infty}\binom{2n}{n}\left(pqz^2\right)^n.$$

Pero de la expansión de Taylor $(1-x)^{-1/2} = \sum_{n=0}^{\infty}\binom{2n}{n}(x/4)^n$, cuando $|x| < 1$, llegamos a que

$$F_{00}(z) = 1 - \sqrt{1-4pqz^2},\ |z| < 1.$$

De esto, como $p = 1/2$, tenemos que $P_0(T_0 < \infty) = F_{00}(1) = 1$, o sea que la cadena es recurrente. Ahora bien, del Teorema de Abel concluimos que la suma $F_{00}(z)$ es uniformente convergente en $|z| \leq 1$. Luego, podemos intercambiar la suma con la derivada $\frac{d}{dz}$, para llegar a que

$$\frac{d}{dz}F_{00}(z) = \sum_{n=1}^{\infty} nz^{n-1}f_{00}(n) = \frac{1}{2}\frac{2z}{\sqrt{1-z^2}}.$$

Al tomar el limite cuando $z \to 1$ concluimos que

$$E_0(T_0) = \sum_{n=1}^{\infty} nf_{00}(n) = \infty.$$

Por lo tanto, la cadena es recurrente nula. $\Box$

Ejemplo 13.13. Una cola en tiempo discreto. Para cada $n = 0, 1, \ldots$, sea X_n el número de trabajos en un sistema de espera con un servidor al inicio del n–ésimo periodo, $[n, n+1)$, y sea Y_n el número de arribos en ese periodo. Si el sistema no está vacío ($\mathrm{X}_n > 0$) se completa un y sólo un servicio (y el trabajo servido sale del sistema) y, además, no se aceptan nuevos trabajos, i.e.

$$\mathrm{X}_{n+1} = \mathrm{X}_n - 1 \quad \text{si} \quad \mathrm{X}_n > 0.$$

Por otra parte, si el sistema está vacío ($\mathrm{X}_n = 0$), se admite la entrada a todos los arribos, Y_n, en el periodo $[n, n+1)$, i.e.

$$\mathrm{X}_{n+1} = Y_n \quad \text{si} \quad \mathrm{X}_n = 0.$$

Supóngase que las vv.aa. Y_n son i.i.d. con distribución

$$q(y) := \mathrm{P}(Y_n = y) > 0 \quad \forall\, y = 1, 2, \ldots.$$

Por lo tanto, las probabilidades de transición de $X_\bullet$ son

$$\mathrm{P}(0,y) = q(y) \quad \forall\, y \geq 1;\ \mathrm{P}(x,x-1) = 1 \quad \forall\, x \geq 1.$$

Es evidente que $X_\bullet$ es *irreducible*. (Explique.) Por lo tanto, para verificar que $X_\bullet$ es una CM recurrente o transitoria basta analizar un estado, digamos $x = 0$. Por la Proposición 13.1(a), $\mathrm{P}_0(T_0 = 1) = \mathrm{P}(0,0) = 0$, y para $n \geq 2$

$$\mathrm{P}_0(T_0 = n) = \sum_{y \neq 0} \mathrm{P}(0,y)\mathrm{P}_y(T_0 = n-1).$$

Luego, como $\mathrm{P}(0,y) = q(y)$, y $\mathrm{P}_y(T_0 = n-1) = 1$ ssi $y = n-1$, vemos que

$$\mathrm{P}_0(T_0 = n) = q(n-1) \quad \forall\, n \geq 2.$$

De aquí se sigue que

$$f_{00} := \mathrm{P}_0(T_0 < \infty) = \sum_{n=1}^{\infty} \mathrm{P}_0(T_0 = n) = \sum_{n=1}^{\infty} q(n) = 1;$$

es decir, $x = 0$ es un estado recurrente, y por lo tanto $X_\bullet$ es *recurrente*. Ahora deseamos ver si $X_\bullet$ es recurrente positiva o recurrente nula.

Por (18),

$$m_0 := E_0(T_0) = \mathrm{P}(0,0) + \sum_{y \geq 1} \mathrm{P}(0,y)[f_{y0} + E_y(T_0)].$$

Además [por el Corolario 13.9(c)], $f_{y0} = 1$, mientras que $\mathrm{P}(0,y) = q(y)$ y $E_y(T_0) = y$ para todo $y \geq 1$. Por lo tanto

$$m_0 = \sum_{y=1}^{\infty} q(y)(1+y) = 1 + \overline{q},$$

donde $\overline{q} := E(Y_n) = \sum_{y=1}^{\infty} y\, q(y)$ es el número esperado de arribos por periodo. De aquí concluimos que el estado $x = 0$ (y por lo tanto $X_\bullet$) es *recurrente positivo* si $\overline{q} < \infty$ y *nulo* si $\overline{q} = \infty$. $\square$

Del Corolario 13.10(b) podemos concluir lo siguiente.

Corolario 13.14. *El espacio de estados S se puede particionar en forma única como*

$$S = T \cup C_1 \cup C_2 \cup \dots,$$

en donde T es el conjunto de estados transitorios, y los C_k son clases comunicantes cerradas de estados recurrentes.

A este corolario se le conoce como el **teorema de descomposición** (de S).

Ejercicios § 13

13.1. Demuestre la Proposición 14.2.

13.2. Encuentre las clases comunicantes de la CM con espacio de estados $S = \{1, \dots, 5\}$ y matriz de transición

$$\begin{bmatrix} \surd & 0 & 0 & 0 & \surd \\ 0 & \surd & 0 & \surd & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \surd & \surd & \surd & \surd \\ \surd & 0 & 0 & 0 & \surd \end{bmatrix}$$

en donde $\surd$ representa números positivos. Diga qué clases son cerradas.

13.3. Si y es un estado transitorio, entonces [por el Lema 13.6(d)] $V(x, y) < \infty$ para todo $x \in S$. De este hecho deduzca que si y es transitorio entonces

$$\lim_{n\to\infty} \mathrm{P}_n(x, y) = 0 \quad \forall\, x \in S.$$

13.4. Supóngase que $x \leftrightarrow y$. Demuestre que si x es recurrente positivo o recurrente nulo, entonces y es recurrente positivo o recurrente nulo, respectivamente. (De aquí se sigue, en particular, que en una clase recurrente todos los estados son recurrentes positivos o todos son recurrentes nulos.) (La demostración es complicada. Se sugiere que el lector consulte, por ejemplo, el libro de D.P. Heyman and M.J. Sobel, *Stochastic Models of Operations Research, Volume I*, p. 234, Theorem 7–7.)

13.5. Demuestre que en una CM irreducible y recurrente, $\mathrm{P}(T_y < \infty) = 1$ para todo estado y.

13.6. Se dice que una CM es **transitoria** si todos sus estados son transitorios. (Por ejemplo, la CM $\mathrm{X}_{n+1} := \mathrm{X}_n + 1$ con espacio de estados $S = \{0, 1, \ldots\}$ es transitoria.) Demuestre que una CM con espacio de estados *finito* no puede ser transitoria, es decir tiene al menos un estado recurrente. (*Sugerencia*: suponga que la CM es transitoria y use el Ejercicio 3.)

13.7. Sea y un estado absorbente, es decir, $\mathrm{P}(y, y) = 1$. Demuestre que: (a) y es recurrente, y (b) $\mathrm{P}_n(x, y) = \mathrm{P}_x(T_y \leq n)$ para todo $x \in S$ y $n \geq 1$. [*Sugerencia*: en (b) use la fórmula (9).]

13.8. Demuestre: (a) $f_{xy} = \mathrm{P}(x, y) + \sum_{z \neq y} \mathrm{P}(x, z) f_{zy}$.

(b) $\mathrm{P}_x(T_y \leq n) = \mathrm{P}(x, y) + \sum_{z \neq y} \mathrm{P}(x, y)\mathrm{P}_z(T_y \leq n - 1)$ para todo $n \geq 1$.

(c) $E_x(T_y) = 1 + \sum_{z \neq y} \mathrm{P}(x, z) E_z(T_y)$ si $f_{xy} = 1$.

(Compare (b) y (c) con (16) y la Proposición 13.11(b), respectivamente.)

13.9. Sea F un subconjunto no vacío del espacio de estados de una CM. Se dice que F es un conjunto **cerrado** si $\mathrm{P}(x, y) = 0$ para todo $x \in F$ y $y \notin F$. Demuestre que las siguientes afirmaciones son equivalentes:

(a) F es un conjunto cerrado.

(b) $\mathrm{P}_n(x, y) = 0 \quad \forall\, x \in F,\ y \notin F,\ n \geq 1$.

(c) $f_{xy} = 0 \quad \forall\, x \in F,\ y \notin F$.

13.10. Considérese una CM con espacio de estados $S = \{0, 1, \ldots\}$ y probabilidades de transición $\mathrm{P}(x, x + 1) = p$ (para algún $0 < p < 1$) y $\mathrm{P}(x, 0) = 1 - p$.

(a) Demuestre que la CM es irreducible.

(b) Encuentre $\mathrm{P}_0(T_0 = n)$ para todo $n \geq 1$.

(c) Demuestre que la CM es recurrente.

(d) Diga si la CM es recurrente positiva o recurrente nula — demuestre que su respuesta es correcta.

13.11. Considérese una CM $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$. Para cada $m = 1, 2, \ldots$, sea $N_m(y)$ el número de visitas de $\mathrm{X}_\bullet$ al estado y hasta el tiempo m, es decir

$$N_m(y) := \sum_{n=1}^{m} I_y(\mathrm{X}_n).$$

[Nótese que $N_m(y) \to N(y)$ cuando $m \to \infty$, con $N(y)$ como en (3).] Asimismo, sea $V_m(x, y)$ el número esperado de tales visitas, dado que $\mathrm{X}_0 = x$, i.e.

$$V_m(x, y) := E_x[N_m(y)] = \sum_{n=1}^{m} \mathrm{P}_n(x, y).$$

[Por (5), $V_m(x, y) \to V(x, y)$ cuando $m \to \infty$.] Demuestre que si y es un estado transitorio de $\mathrm{X}_\bullet$, entonces para cualquier estado inicial x:

(a) $\lim\limits_{m\to\infty} \frac{1}{m} V_m(x, y) = 0$, (b) $\lim\limits_{m\to\infty} \frac{1}{m} N_m(y) = 0$ c.s.

13.12. Sea F un conjunto cerrado de una CM (vea el Ejercicio 9) y suponga que para cada $x \notin F$ existe $y \in F$ tal que $x \to y$. Demuestre que todos los estados $x \notin F$ son transitorios.

13.13. Considérese una CM con estados en $\{0, 1, \ldots\}$ y probabilidades de transición

$$\mathrm{P}(0, y) = a_y > 0 \quad \forall\, y \geq 0; \quad \mathrm{P}(x, x) = p, \quad \mathrm{P}(x, x-1) = 1 - p \quad \forall\, x \geq 1.$$

Clasifique los estados de la CM y calcule sus tiempos medios de retorno.

13.14. Una ley cero–uno. Demuestre que para cualquier estado $x \in S$

(a) $\mathrm{P}_x(N(x) = \infty) = 0$, ó (b) $\mathrm{P}_x(N(x) = \infty) = 1$.

[De hecho, (a) se cumple ssi $f_{xx} < 1$, y (b) se cumple ssi $f_{xx} = 1$. *Sugerencia:* use el Lemma 13.6(a) o, más directamente, la ecuación (10).]

13.15. Demuestre: si x es un estado recurrente positivo y $x \leftrightarrow y$, entonces $E_y(T_x) < \infty$. [*Sugerencia:* itere la ecuación en el Ejercicio 8(c) con $y = x$, y use la Proposición 13.1(b).]

13.16. Considere una CM que tiene una distribución estacionaria $\pi^* = \{\pi^*(x),\, x \in S\}$. Demuestre que

$$\pi^*(y) = \sum_{x \in S} \pi^*(x) \frac{1}{m} \sum_{n=1}^{m} \mathrm{P}_n(x, y) \quad \forall\, y \in S.$$

De esta expresión deduzca que si y es un estado transitorio, entonces $\pi^*(y) = 0$. (*Sugerencia:* use el Ejercicio 11(a).)

14 Distribución límite de una cadena de Markov

Contenido: Convergencia de probabilidades de transición, distribución límite, periodicidad.

En esta sección, $\mathrm{X}_\bullet = \{\mathrm{X}_n,\, n = 0, 1, \ldots\}$ es una CM definida sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$ y con espacio de estados S a lo más numerable. Para cada par de estados x, y deseamos encontrar condiciones para la existencia de límites de la forma

$$\lim_{n\to\infty} \mathrm{P}_n(x, y) \tag{14.1}$$

o

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathrm{P}_k(x, y). \tag{14.2}$$

Por el Lema de Toeplitz (Ejercicio 15.9 más adelante), la existencia del límite en (1) implica la convergencia en (2). El problema es que (1) se cumple sólo bajo condiciones muy restrictivas y, por lo tanto, conviene estudiar primero el límite en (2), el cual *siempre existe*.

Ejemplo 14.1. Sea $\mathrm{X}_\bullet$ una CM con espacio de estados $S = \{0, 1\}$ y matriz de transición

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Si $x = 0$ ó 1, entonces

$$\mathrm{P}_n(x, x) = \begin{cases} 1 & \text{si} \quad n \quad \text{es par}, \\ 0 & \text{si} \quad n \quad \text{es impar}, \end{cases}$$

mientras que si $x \neq y$ entonces

$$\mathrm{P}_n(x, y) = \begin{cases} 1 & \text{si} \quad n \quad \text{es impar}, \\ 0 & \text{si} \quad n \quad \text{es par}. \end{cases}$$

De aquí se sigue que el límite en (1) **no** existe, mientras que el límite en (2) es $1/2$ para cualquier x, y. (La no existencia del límite en (1) se debe a que la CM es "periódica", concepto que estudiaremos más adelante.) $\square$

Usaremos la notación en el Ejercicio 13.11: para cada $n = 1, 2, \ldots, N_n(y)$ es el número de visitas de $X_\bullet$ al estado y hasta el tiempo n, i.e.

$$N_n(y) := \sum_{k=1}^{n} I_y(X_k),$$

mientras que

$$V_n(x, y) := E_x[N_n(y)] = \sum_{k=1}^{n} P_k(x, y)$$

es el número esperado de visitas al estado y hasta el tiempo n, dado que el estado inicial es $X_0 = x$. Además, definimos

$$\overline{N}_n(y) := \frac{1}{n} N_n(y), \quad \overline{V}_n(x, y) := \frac{1}{n} V_n(x, y) = \frac{1}{n} \sum_{k=1}^{n} P_k(x, y). \tag{14.3}$$

A. Convergencia de promedios

El resultado principal en relación con los promedios (o "sumas de Cesàro") en (3) es el siguiente.

Teorema 14.2. *Para todo $x, y \in S$ los límites*

$$\lim_{n\to\infty} \overline{N}_n(y) =: \overline{\pi}(x, y) \quad c.s., \tag{14.4}$$

$$\lim_{n\to\infty} \overline{V}_n(x, y) =: \pi(x, y) \tag{14.5}$$

existen y

$$\pi(x, y) = E_x[\overline{\pi}(x, y)]. \tag{14.6}$$

Además, si y es recurrente se cumple que

$$(i) \quad \overline{\pi}(x, y) = I_{\{T_y < \infty\}}/m_y, \qquad (ii) \quad \pi(x, y) = f_{xy}/m_y \tag{14.7}$$

en donde $m_y := E_y(T_y)$ es el tiempo medio de retorno a y. Por lo tanto, para todo $x \in S$:

(a) $\pi(x, y) = 0$ *si y es transitorio o recurrente nulo, y*

(b) $\pi(x, y) = f_{xy}/m_y$ *si y es recurrente positivo.*

Nota 14.3. En la siguiente demostración se usa el hecho de que la ley fuerte de los grandes números se cumple inclusive si la media es infinita, siempre y cuando esté bien definida.

Demostración. Por el teorema de convergencia acotada, si (4) se cumple entonces también se cumplen (5) y (6). Por otra parte, si y es *transitorio*, del Ejercicio 13.11 se tiene que $\overline{\pi}(x, y) = 0$ c.s. y $\pi(x, y) = 0$. Por lo tanto, para demostrar el teorema basta suponer que y es *recurrente* y verificar (4) y (7). Primero consideraremos el caso en el que $\mathrm{X}_0 = y$.

Sea $T_y(0) := 0$, y para $r = 1, 2, \ldots$ sea $T_y(r)$ el *tiempo de la r–ésima visita al estado y*, el cual se puede definir recursivamente como

$$T_y(r) := \inf\{n \geq T_y(r-1) + 1 \mid \mathrm{X}_n = y\} \quad \forall\, r = 1, 2, \ldots$$

Como $T_y(0) := 0$, el tiempo de la primera visita a y coincide con el tiempo del primer paso a y, es decir, $T_y(1) \equiv T_y := \inf\{n \geq 1 \,|\, \mathrm{X}_n = y\}$. Alternativamente, podemos escribir

$$T_y(r) := \inf\{n \geq 1 \mid N_n(y) = r\} \quad \forall\, r = 1, 2, \ldots. \tag{14.8}$$

(Nótese que $T_y(r) < \infty$ para todo r, porque $\mathrm{X}_0 = y$ y y es recurrente.) Ahora consideremos, para $r = 1, 2, \ldots,$

$$W_y(r) := T_y(r) - T_y(r-1) = \inf\{n \geq 1 \mid \mathrm{X}_{T_y(r-1)+n} = y\} \tag{14.9}$$

el *tiempo de espera* entre la $(r-1)$–ésima visita a y y la r–ésima visita (también llamado la longitud de la r–ésima excursión a y). Por el Ejercicio 4, las vv.aa. $W_y(r)$, para $r = 1, 2, \ldots,$ son i.i.d. con distribución

$$\mathrm{P}[W_y(r) = n] = \mathrm{P}_y(T_y = n),$$

de modo que $E_y[W_y(1)] = E_y(T_y) = m_y$. Luego, como

$$T_y(r) = W_y(1) + \cdots + W_y(r),$$

de la ley fuerte de los grandes números (vea 9.12) se sigue que

$$\lim_{r\to\infty} \frac{1}{r} T_y(r) = m_y \quad \text{c.s.} \tag{14.10}$$

Por otro lado, se sigue de (8) que si $N_n(y) = r$ entonces $T_y(r) \leq n < T_y(r+1)$, i.e.

$$T_y(N_n(y)) \leq n < T_y(N_n(y) + 1).$$

Luego, como $N_n(y) \to N(y) = \infty$ cuando $n \to \infty$, se obtiene que

$$\frac{T_y(N_n(y))}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{T_y(N_n(y) + 1)}{N_n(y)}$$

y tomando el límite cuando $n \to \infty$ vemos de (10) que

$$\lim_{n\to\infty} \frac{n}{N_n(y)} = m_y \quad \text{c.s.},$$

es decir

$$\lim_{n\to\infty} \overline{N}_n(y) = 1/m_y \quad \text{c.s.} \tag{14.11}$$

De aquí se siguen (4) y (7) cuando $\mathrm{X}_0 = y$ es recurrente.

Finalmente, suponga que el estado inicial es $\mathrm{X}_0 = x$ arbitrario. Entonces, por el Corolario 13.9, puede ocurrir que la CM alcance o no el estado y (i.e. que estén o no comunicados), pero en cualquier caso $\overline{N}_n(y)$ satisface (4) con $\overline{\pi}(x, y)$ como en (7). □ □

Corolario 14.4. *Si la CM tiene una distribución invariante* $\pi^* = \{\pi^*(y),\ y \in S\}$*, entonces:*

(a) $\pi^*(y) = 0$ *si* y *es transitorio o recurrente nulo, y*

(b) $\pi^*(y) = \dfrac{1}{m_y} \sum_{x\in S} \pi^*(x) f_{xy}$ *si* y *es recurrente positivo.*

Demostración. Como π^* es invariante,

$$\pi^*(y) = \sum_{x\in S} \pi^*(x) \mathrm{P}_k(x, y) \quad \forall\ k = 0, 1, \ldots;\ y \in S.$$

Por lo tanto, sumando de $k = 1$ a $k = n$ y multiplicando por $1/n$, se sigue de (3) que

$$\pi^*(y) = \sum_{x\in S} \pi^*(x) \overline{V}_n(x, y) \quad \forall\ n = 1, 2, \ldots;\ y \in S. \tag{14.12}$$

Luego, como $0 \leq \overline{V}_n(x,y) \leq 1$, del Teorema de Convergencia Dominada y (5), al tomar $n \to \infty$ en (12) concluimos que

$$\pi^*(y) = \sum_{x \in S} \pi^*(x)\pi(x,y) \quad \forall\, y \in S. \tag{14.13}$$

Por lo tanto, (a) y (b) se siguen del Teorema 14.2. $\square$ $\square$

Teorema 14.5. *Si la CM es irreducible, entonces las siguientes condiciones son equivalentes:*

(a) *La CM es recurrente positiva.*

(b) *La CM tiene una única distribución invariante, digamos π^*. Además, en este caso $\pi^*(x) = 1/m_x > 0$ para todo $x \in \mathrm{X}$.*

Demostración. (b) implica (a). Si π^* es una distribución invariante, necesariamente $\pi^*(y) > 0$ para algún estado y, y entonces, por el Corolario 14.4, y es recurrente positivo. De hecho, por la parte (b) de dicho corolario,

$$0 < \pi^*(y) = 1/m_y,$$

así que $m_y < \infty$. De aquí se obtiene (a), por la irreducibilidad de la CM.

(a) implica (b). Supóngase que la CM es recurrente positiva. Entonces $f_{xy} = 1$ para cualesquiera dos estados x, y de modo que si la CM tiene una distribución invariante, entonces, por (12) y el Teorema 14.2(b),

$$\pi^*(y) = \lim_{n\to\infty} \sum_{x \in S} \pi^*(x)\overline{V}_n(x,y) = 1/m_y \quad \forall\, y \in S. \tag{14.14}$$

Nótese que $\pi^*(y)$ coincide con el límite $\pi(x,y)$ en (5)–(7). Por lo tanto, si π' es alguna otra distribución invariante, el argumento usado en (12) y (13) da que

$$\pi'(y) = \sum_{x \in S} \pi'(x)\pi(x,y) = \sum_{x \in S} \pi'(x)\pi^*(y) = \pi^*(y) \quad \forall\, y \in S. \tag{14.15}$$

En otras palabras, si la CM tiene una distribución invariante π^*, entonces π^* es *única* y está dada por (14).

Ahora demostraremos que, efectivamente, (14) define una distribución invariante. Con este fin probaremos primero que

$$\text{(a)} \sum_{y \in S} \pi^*(y) \leq 1, \quad \text{y} \quad \text{(b)}\ \pi^*(y) = \sum_{x \in S} \pi^*(x)\mathrm{P}(x,y) \ \ \forall\, y \in S. \tag{14.16}$$

Después, definiendo $\overline{\pi} := \sum_y \pi^*(y)\,(> 0)$ y $\pi'(y) := \pi^*(y)/\overline{\pi}$ para todo y, concluiremos que π' es una distribución invariante, así que, por (15), $\overline{\pi} = 1$, es decir $\pi' = \pi^*$ es la única distribución invariante.

Para demostrar (16)(a) primero obsérvese que, por (3),

$$\sum_{y \in S} \overline{V}_n(x, y) = \frac{1}{n} \sum_{k=1}^{n} \sum_{y \in S} \mathrm{P}_k(x, y) = 1 \quad \forall \in S.$$

De aquí se sigue (16)(a) porque usando (14) y el Lema de Fatou obtenemos

$$1 = \liminf_{n \to \infty} \sum_{y \in S} \overline{V}_n(x, y) \geq \sum_{y \in S} \liminf_{n \to \infty} \overline{V}_n(x, y) = \sum_{y \in S} \pi^*(y).$$

Para demostrar (16)(b) primero usamos la ecuación de Chapman – Kolmogorov para escribir

$$\mathrm{P}_{k+1}(x, y) = \sum_{z \in S} \mathrm{P}_k(x, z)\mathrm{P}(z, y)$$

y después un cálculo directo da

$$\frac{n+1}{n} \cdot \overline{V}_{n+1}(x, y) - \frac{1}{n}\mathrm{P}(x, y) = \sum_{z \in S} \overline{V}_n(x, z)\mathrm{P}(z, y).$$

Por lo tanto, usando de nuevo (14) y el Lema de Fatou vemos que

$$\pi^*(y) \geq \sum_{z \in S} \pi^*(z)\mathrm{P}(z, y) \quad \forall\, y \in S. \tag{14.17}$$

Si la igualdad **no** se cumple en (17) para algún $y \in S$, i.e.

$$\pi^*(y) > \sum_{z \in S} \pi^*(z)\mathrm{P}(z, y),$$

entonces

$$\sum_{y \in S} \pi^*(y) > \sum_{z \in S} \pi^*(z) \sum_{y \in S} \mathrm{P}(z, y) = \sum_{z \in S} \pi^*(z),$$

o sea $\overline{\pi} > \overline{\pi}$. Esta contradicción da que la igualdad se cumple en (17) para todo $y \in S$. □ □

Del Teorema 14.5 y la expresión (7)(i) concluimos lo siguiente.

Corolario 14.6. *Si la CM es irreducible y recurrente positiva, entonces para cualquier estado inicial* $\mathrm{X}_0 = x$,

$$\lim_{n\to\infty} \overline{N}_n(y) = 1/m_y \quad c.s. \quad \forall\, y \in S.$$

B. Distribuciones límite

La convergencia en (1) está asociada con el concepto de "periodicidad" de una CM. Para introducir este concepto recordemos que dados dos enteros positivos d, n se dice que d es **divisor** de n si n/d es un entero. Además, si J es un conjunto no vacío de enteros positivos, definimos el **máximo común divisor** (m.c.d.) de J como el mayor entero positivo d que es divisor de n para todo $n \in J$. Nótese que

$$1 \leq \text{m.c.d.}(J) \leq \min\{n \,|\, n \in J\}.$$

Por ejemplo, si $1 \in J$, entonces m.c.d.$(J)=1$. Asimismo, si $J=\{4, 6, 10, 12\}$, entonces m.c.d.$(J)=2$.

Definición 14.7. Sea $x \in S$ un estado de la CM, y $J_x := \{n > 0 \,|\, \mathrm{P}_n(x, x) > 0\}$. Definimos el **periodo** $d(x)$, o simplemente d, de x como

$$d := \text{m.c.d.}(J_x).$$

En otras palabras, x tiene periodo d si $\mathrm{P}_n(x, x) = 0$ cuando $n \neq d, 2d, 3d, \ldots$, y d es el mayor entero con esta propiedad. Cada estado de la CM del Ejemplo 14.1 es periódico con periodo $d = 2$.

Si $d(x) = 1$, se dice que x es un estado **aperiódico**. Por ejemplo, si $\mathrm{P}(x, x) > 0$, entonces x es aperiódico.

Ejemplo 14.8. Considérese una cadena de NM con parámetros $p_x := \mathrm{P}(x, x+1)$ y $q_x := \mathrm{P}(x, x-1)$, ambos positivos. Un estado x es aperiódico si $r_x := \mathrm{P}(x, x) = 1-(p_x+q_x) > 0$, y periódico con periodo $d = 2$ si $r_x = 0$ $\quad \Box$.

La siguiente proposición demuestra que la periodicidad es una "propiedad de clase", es decir, dos estados en una misma clase comunicante tienen el mismo periodo. En particular, *en una CM irreducible todos los estados tienen el mismo periodo.*

Proposición 14.9. Si $x \leftrightarrow y$, entonces $d(x) = d(y)$.

Demostración. Sean n, m enteros tales que $\mathrm{P}_n(x,y) > 0$ y $\mathrm{P}_m(y,x) > 0$. Entonces, por la ecuación de Chapman–Kolmogorov,

$$\mathrm{P}_{n+m}(x,x) \geq \mathrm{P}_n(x,y)\mathrm{P}_m(y,x) > 0$$

de modo que $d(x)$ es divisor de $n+m$. Asimismo, si $\mathrm{P}_k(y,y) > 0$, entonces

$$\mathrm{P}_{n+k+m}(x,x) \geq \mathrm{P}_n(x,y)\mathrm{P}_k(y,y)\mathrm{P}_m(y,x) > 0,$$

así que $d(x)$ también es divisor de $n+k+m$. Esto implica que $d(x)$ es divisor de k porque $k = (n+k+m) - (n+m)$. Luego, como $k \in J_y$, concluimos que $d(x) \leq d(y)$. Análogamente se demuestra que $d(y) \leq d(x)$. $\Box$ $\Box$

La demostración del siguiente resultado se puede ver, por ejemplo, en el libro de Norris (1997), página 43, Theorem 1.8.4.

Proposición 14.10. Supóngase que la CM es irreducible. Existe un entero $d \geq 1$ y una partición

$$S = C_0 \cup C_1 \cup \cdots \cup C_{d-1}$$

de S tal que, definiendo $C_{nd+r} := C_r$,

(a) $\mathrm{P}_n(x,y) > 0$ sólo si $x \in C_r$ y $y \in C_{r+n}$ para algún r;

(b) $\mathrm{P}_{nd}(x,y) > 0$ para todo n suficientemente grande, para todo $x, y \in C_r$ y todo r.

El entero $d \geq 1$ en 14.10 es el periodo de la CM. Como ejemplo considérese una CM con espacio de estados $S = \{1,2,3\}$ y matriz de transición

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

En este caso $d = 3$.

Teorema 14.11. *Supóngase que la CM es irreducible y recurrente positiva, y sea π^* su distribución invariante, es decir $\pi^*(y) = 1/m_y$ para todo $y \in S$.*

(a) *Si la CM es aperiódica, entonces*

$$\lim_{n\to\infty} \mathrm{P}_n(x,y) = \pi^*(y) \quad \forall\, x, y \in S; \tag{14.18}$$

de hecho, para cualquier distribución inicial π_0,

$$\pi_n(y) = \mathrm{P}(\mathrm{X}_n = y) \to \pi^*(y) \quad \textit{cuando} \quad n \to \infty, \;\; \forall\, y \in S. \tag{14.19}$$

(b) *Si la CM es periódica con periodo d, entonces para cualquier par de estados x, y, existe un entero r, con $0 \leq r < d$, tal que* $\mathrm{P}_n(x,y) = 0$ *a menos que $n = md + r$ para algún entero $m \geq 0$, y*

$$\lim_{m\to\infty} \mathrm{P}_{md+r}(x,y) = d\pi^*(y). \tag{14.20}$$

Demostración. Sólo demostraremos el inciso (a). La demostración de (b) se puede ver en el libro de Norris (1997), página 44, Theorem 1.8.5, o en el de Hoel, Port y Stone (1972), p. 73, Theorem 7.

Demostración de (a). Sea $Y_\bullet = \{Y_n\}$ una CM con la misma matriz de transición que $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$, pero independiente de $\mathrm{X}_\bullet$. Sea $Z_\bullet = (\mathrm{X}_\bullet, Y_\bullet)$ la CM en $S \times S := \{(x,y) \,|\, x, y \in S\}$ con probabilidad de transición

$$\overline{\mathrm{P}}((x,y),(z,w)) = \mathrm{P}(x,z)\mathrm{P}(y,w).$$

Parte 1. La CM $Z_\bullet$ es irreducible y recurrente positiva. Como P es irreducible, dados cualesquiera dos pares de estados (x_1, x_2) y (y_1, y_2), existen enteros n y m tales que $\mathrm{P}_n(x_1,x_2) > 0$ y $\mathrm{P}_m(y_1,y_2) > 0$. Además, como x_2 y y_2 son aperiódicos, por la Proposición 14.10(b) (con $d = 1$) existe un entero $n_0 > 0$ tal que

$$\mathrm{P}_{m+r}(x_2,x_2) > 0 \quad \text{y} \quad \mathrm{P}_{n+r}(y_2,y_2) > 0 \quad \forall\, r \geq n_0.$$

Por lo tanto

$$\overline{\mathrm{P}}_{n+m+r}((x_1,y_1),(x_2,y_2)) > 0 \tag{14.21}$$

porque $\mathrm{P}_{n+m+r}(x_1,x_2) \geq \mathrm{P}_n(x_1,x_2)\mathrm{P}_{m+r}(x_2,x_2)$ y similarmente para (y_1,y_2). Por (21), $Z_\bullet$ es irreducible. Luego, para ver que $Z_\bullet$ es recurrente positiva, por el Teorema 14.5 basta verificar que $\overline{\pi}(x,y) = \pi^*(x)\pi^*(y)$ es una distribución invariante, lo cual es inmediato porque

$$\begin{aligned} \overline{\pi}(x,y) &= \pi^*(x)\pi^*(y) = \sum_z \pi^*(z)\mathrm{P}(z,x) \cdot \sum_w \pi^*(w)\mathrm{P}(w,y) \\ &= \sum_z \sum_w \overline{\pi}(z,w)\overline{\mathrm{P}}((z,w),(x,y)). \end{aligned}$$

Parte 2. Fíjese un estado arbitrario $a \in S$, y sea

$$T(a) := \inf\{n \geq 1 \,|\, \mathrm{X}_n = Y_n = a\}$$

el tiempo del primer paso de $Z_\bullet$ al estado $(a,a) \in S \times S$. Como $Z_\bullet$ es irreducible y recurrente positiva, se sigue del Ejercicio 13.5 que $T(a) < \infty$ c.s. Esto implica que el tiempo de paro

$$T := \inf\{n \geq 1 \,|\, \mathrm{X}_n = Y_n\} < \infty \quad \text{c.s.} \tag{14.22}$$

para cualquier estado inicial $(x,y) \in S \times S$, porque $T \leq T(a)$. Por otra parte, como $\mathrm{X}_\bullet$ y $Y_\bullet$ tienen la misma matriz de transición, vemos que para $1 \leq k \leq n$

$$\begin{aligned}
\mathrm{P}(\mathrm{X}_n = y,\, T = k) &= \sum_{x \in S} \mathrm{P}(\mathrm{X}_n = y,\, T = k,\, \mathrm{X}_k = Y_k = x) \\
&= \sum_{x \in S} \mathrm{P}(T = k,\, \mathrm{X}_k = Y_k = x)\mathrm{P}_{n-k}(x,y) \\
&= \mathrm{P}(Y_n = y,\, T = k).
\end{aligned}$$

Por lo tanto, sumando sobre $k = 1, \ldots, n$,

$$\mathrm{P}(\mathrm{X}_n = y,\, T \leq n) = \mathrm{P}(Y_n = y,\, T \leq n). \tag{14.23}$$

Parte 3. Por (23):

$$\begin{aligned}
\mathrm{P}(\mathrm{X}_n = y) &= \mathrm{P}(\mathrm{X}_n = y,\, T \leq n) + \mathrm{P}(\mathrm{X}_n = y,\, T > n) \\
&= \mathrm{P}(Y_n = y,\, T \leq n) + \mathrm{P}(\mathrm{X}_n = y,\, T > n) \\
&\leq \mathrm{P}(Y_n = y) + \mathrm{P}(\mathrm{X}_n = y,\, T > n),
\end{aligned}$$

es decir, $\mathrm{P}(\mathrm{X}_n = y) \leq \mathrm{P}(Y_n = y) + \mathrm{P}(\mathrm{X}_n = y,\, T > n)$. Análogamente,

$$\mathrm{P}(Y_n = y) \leq \mathrm{P}(\mathrm{X}_n = y) + \mathrm{P}(Y_n = y,\, T > n),$$

de modo que

$$|\,\mathrm{P}(\mathrm{X}_n = y) - \mathrm{P}(Y_n = y)| \leq \mathrm{P}(\mathrm{X}_n = y,\, T > n) + \mathrm{P}(Y_n = y,\, T > n).$$

y

$$\sum_{y \in S} |\,\mathrm{P}(\mathrm{X}_n = y) - \mathrm{P}(Y_n = y)| \leq 2\mathrm{P}(T > n). \tag{14.24}$$

Finalmente, tómese $\mathrm{X}_0 = x$ y $Y_0 \sim \pi^*$. Entonces $\mathrm{P}(\mathrm{X}_n = y) = \mathrm{P}_n(x,y)$ y $\mathrm{P}(Y_n = y) = \pi^*(y)$ para todo $n \geq 0$ y $y \in S$, así que (24) se reduce a

$$\sum_{y \in S} |\,\mathrm{P}_n(x,y) - \pi^*(y)| \leq 2\mathrm{P}(T > n)$$

y, por (22),

$$\sum_{y\in S} |\,\mathrm{P}_n(x,y)-\pi^*(y)| \to 0 \quad \text{cuando} \quad n\to\infty. \tag{14.25}$$

Esto implica (18), que a su vez, junto con el Ejercicio 12.12, implica (19). $\square$

Observación 14.12. Si π y π' son dos distribuciones de probabilidad sobre S, definimos la distancia en **variación total** entre π y π' como

$$\|\,\pi-\pi'\,\| := \sum_{y\in S} |\,\pi(y)-\pi'(y)|.$$

Luego, (25) afirma que para cualquier estado inicial x

$$\|\,\mathrm{P}_n(x,\cdot)-\pi^*(\cdot)\,\| \to 0, \tag{14.26}$$

que de hecho es una condición mucho más fuerte que la convergencia "local" en (18) porque

$$\sup_{y\in S} |\,\mathrm{P}_n(x,y)-\pi^*(y)\,| \le \|\,\mathrm{P}_n(x,\cdot)-\pi^*(\cdot)\,\| \quad \forall\, x\in S;$$

es decir, la convergencia en (18) resulta ser *uniforme en y*. Por otra parte, para cualquier distribución inicial π_0, tenemos

$$\pi_n(y) := \sum_{x\in S}\pi_0(x)\mathrm{P}_n(x,y) \quad \text{y} \quad \pi^*(y) = \sum_{x\in S}\pi_0(x)\pi^*(y).$$

De aquí se puede deducir fácilmente (Ejercicio 6) que

$$\sup_{y\in S} |\,\pi_n(y)-\pi^*(y)| \le \|\,\pi_n-\pi^*\,\| \to 0 \tag{14.27}$$

cuando $n\to 0$. $\square$

Ejercicios § 14

14.1. Use la ecuación (9) en la Sección 13 para demostrar que

(a) $\sum_{n=1}^{\infty}\mathrm{P}_n(x,y) = f_{xy}\sum_{n=0}^{\infty}\mathrm{P}_n(y,y)$ y, por lo tanto,

(b) $V(x,y) = f_{xy}[1 + V(y,y)]$.

Además, si el límite $\lim_{n\to\infty} \mathrm{P}_n(y,y) =: \pi(y)$ existe, entonces

(c) $\lim_{n\to\infty} \mathrm{P}_n(x,y) = f_{xy}\pi(y)$,

(d) $\lim_{n\to\infty} \overline{V}_n(x,y) = f_{xy}\pi(y)$, con $\overline{V}_n(x,y)$ como en (3).

14.2. Sea $\{a_0, a_1, \ldots\}$ una sucesión de números no negativos y defina su *función generadora* $A(z)$ como

$$A(z) := \sum_{n=0}^{\infty} z^n a_n \quad \text{para} \quad |z| < 1.$$

(a) Demuestre que para cualquier número a

$$(1-z)A(z) = a + (1-z)\sum_{n=0}^{\infty} z^n (a_n - a),$$

y que, por lo tanto, si $a_n \to a$, entonces $(1-z)A(z) \to a$ cuando $z \uparrow 1$.

(b) Sea $s_n := a_0 + \cdots + a_n$ y observe que $a_n = s_n - s_{n-1}$ para $n \geq 1$. Demuestre que

$$A(z) = (1-z)\sum_{n=0}^{\infty} z^n s_n,$$

de modo que

$$A(z) = a + (1-z)\sum_{n=0}^{\infty} z^n (s_n - a) \quad \forall\, a \in \mathbb{R}.$$

Concluya que si $\lim_{n\to\infty} s_n = \sum_{n=0}^{\infty} a_n = a$, con $a \in \mathbb{R}$, entonces $A(z) \to a$ cuando $z \uparrow 1$.

Nota. El recíproco de (b) también se cumple en el siguiente sentido: si el límite

$$\lim_{z\uparrow 1} A(z) = a \quad \text{con} \quad a \in \mathbb{R}$$

existe, entonces $s_n \to \sum_{n=0}^{\infty} a_n = a$.

(c) Sea s_n como en (b) y sea $\overline{s}_n := s_n/(n+1)$. Demuestre que para cualquier $a \in \mathbb{R}$

$$A(z) = (1-z)^{-1}a + (1-z)\sum_{n=0}^{\infty}(n+1)z^n(\overline{s}_n - a),$$

de modo que

$$(1-z)A(z) = a + (1-z)^2\sum_{n=0}^{\infty}(n+1)z^n(\overline{s}_n - a).$$

Concluya que $\overline{s}_n \to a$ cuando $n \to \infty$ ssi $(1-z)A(z) \to a$ cuando $z \uparrow 1$.

14.3. Sea $y \in S$ un estado arbitrario. Demuestre que si el límite

(a) $$\lim_{n\to\infty} \overline{V}_n(y,y) = \pi(y)$$

existe, entonces

(b) $$\lim_{n\to\infty} \overline{V}_n(x,y) = f_{xy}\pi(y) \quad \forall\, x \in S.$$

[*Sugerencia:* sean $\mathrm{P}_{xy}(z)$ y $F_{xy}(z)$ las funciones generadoras en el Lema 13.7, y nótese que por el Ejercicio 2(c) las condiciones (a) y (b) son equivalentes a

$$\lim_{z\uparrow 1}(1-z)\mathrm{P}_{yy}(z) = \pi(y) \quad \text{y} \quad \lim_{z\uparrow 1}(1-z)\mathrm{P}_{xy}(z) = f_{xy}\pi(y),$$

respectivamente. Finalmente, multiplique por $1-z$ ambos miembros de la ecuación (8) en la Sección 13 y tome el límite cuando $z \uparrow 1$.]

14.4. Sea y un estado recurrente y sean $T_y(r)$ y $W_y(r)$ como en (8) y (9), respectivamente. Supóngase que $\mathrm{X}_0 = y$, así que $T_y(r) < \infty$ para todo $r \geq 1$. Demuestre: para $r = 1, 2, \ldots$

(a) $\mathrm{P}[W_y(r+1) = w_{r+1} \,|\, W_y(k) = w_k \text{ para } 1 \leq k \leq r] = \mathrm{P}_y[W_y(1) = w_{r+1}] = \mathrm{P}_y(T_y = w_{r+1})$, en donde la segunda igualdad se debe a que $W_y(1) = T_y(1) = T_y$;

(b) $\mathrm{P}_y[W_y(1) = w_1, \ldots, W_y(r) = w_r] = \mathrm{P}_y[W_y(1) = w_1] \cdots \mathrm{P}_y[W_y(r) = w_r]$. (*Sugerencia:* use inducción y el inciso (a).)

De (a) y (b) concluya que $W_y(1), W_y(2), \ldots$ son i.i.d. con distribución

$$\mathrm{P}[W_y(r) = n] = \mathrm{P}_y(T_y = n) = f_{yy}(n).$$

14.5. (a) Demuestre que para cada $x \in S$ y $n \geq 1$ fijos, la función $\overline{V}_n(x, \cdot)$ definida en (3) es una distribución de probabilidad sobre S, i.e.

$$\overline{V}_n(x,y) \geq 0 \quad \forall\, y \in S, \quad \text{y} \quad \sum_{y\in S} \overline{V}_n(x,y) = 1.$$

(b) Demuestre que si el espacio S de estados es **finito**, entonces la CM tiene al menos un estado recurrente positivo. Si además la CM es irreducible, entonces todos sus estados son recurrentes positivos.

14.6. Demuestre que (27) se cumple bajo las hipótesis del Teorema 15.10. (*Sugerencia:* use (26) y el Teorema de Convergencia Dominada.)

14.7. Suponga que el límite $\lim_{n\to\infty} \overline{V}_n(x,y) =: \pi(y)$ existe para cada $x, y \in S$. Para cada $n = 0, 1, \ldots$, sea $\pi_n = \{\pi_n(y),\, y \in S\}$ la distribución de X_n. Demuestre que

$$\lim_{m\to\infty} \frac{1}{m} \sum_{n=1}^{m} \pi_n(y) = \pi(y) \quad \forall\, y \in S.$$

14.8. Considérese una CM irreducible con espacio de estados S y matriz de transición P. Se dice que P (o la CM) es **dóblemente estocástica** si la suma de los elementos de cada *columna* es igual a 1, es decir,

$$\sum_{x\in S} \mathrm{P}(x,y) = 1 \quad \forall\, y \in S.$$

Demuestre que si P es dóblemente estocástica y S tiene un número finito N de elementos, entonces $\pi(y) := 1/N$ para todo $y \in S$, define una distribución invariante. Además, los tiempos medios de retorno son $m_y = N$ para cada estado y.

14.9. Considere la CM con espacio de estados $\{1, 2, 3, 4\}$ y matriz de transición

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}.$$

(a) Determine cuales estados son transitorios y cuales son recurrentes.

(b) Calcule f_{xy} para cada x, y.

14.10. Considere la CM con espacio de estados $\{1, \ldots, 5\}$ y matriz de transición

$$\begin{bmatrix} 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 1/4 & 3/4 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Demuestre que la CM es irreducible, y encuentre el periodo y la distribución invariante.

14.11. Sean a y b números arbitrarios. Demuestre que

(a) $\max(a, b) + \min(a, b) = a + b$, y $\max(a, b) - \min(a, b) = |a - b|$;

(b) $\min(a, b) = \frac{1}{2}(a + b - |a - b|)$;

(c) $|a - b| = a + b - 2 \min(a, b)$.

14.12. Sea S un conjunto a lo más numerable, y sea $\mathbb{P}(S)$ el conjunto de todas las distribuciones de probabilidad $\pi = \{\pi(y),\, y \in S\}$ sobre S. Dadas dos distribuciones π_1 y π_2 en $\mathbb{P}(S)$ definimos la distancia en **variación total** entre π_1 y π_2 como en la observación 15.10, i.e.

$$\| \pi_1 - \pi_2 \| := \sum_{y \in S} |\pi_1(y) - \pi_2(y)|.$$

Demuestre que, efectivamente, esta es una función de distancia en $\mathbb{P}(S)$, es decir,

(a) $\| \pi_1 - \pi_2 \| = 0$ ssi $\pi_1(y) = \pi_2(y) \; \forall \, y \in S$,

(b) $\| \pi_1 - \pi_2 \| = \| \pi_2 - \pi_1 \|$, y

(c) $\| \pi_1 - \pi_2 \| \leq \| \pi_1 - \pi_3 \| + \| \pi_3 - \pi_2 \| \;\; \forall \, \pi_1, \pi_2, \pi_3$.

14.13. Considere las siguientes condiciones (a) y (b):

(a) Existe un entero $\nu \geq 1$ y un número $\rho > 0$ tales que

$$\sum_{y \in S} \min\{\mathrm{P}_\nu(x, y),\, \mathrm{P}_\nu(z, y)\} \geq \rho \quad \forall \, x, z \in S.$$

(b) Existe un entero $\nu \geq 1$ y un número $0 \leq \alpha < 1$ tales que

$$\| \pi^x - \pi^z \| \leq 2\alpha \quad \forall\, x, z \in S,$$

en donde π^x y π^z son las distribuciones de probabilidad

$$\pi^x := \{\mathrm{P}_\nu(x,y),\, y \in S\}, \quad \pi^z := \{\mathrm{P}_\nu(z,y),\, y \in S\},$$

y $\| \cdot \|$ distancia en variación total (Ejercicio 12).

Demuestre que (a) y (b) son equivalentes. (*Sugerencia:* use el Ejercicio 11(b).)

Nota. Las condiciones (a) y (b), entre otras, se usan para estudiar **ergodicidad geométrica** de CMs, es decir, el caso en el que en (26) se tiene

$$\| \mathrm{P}_n(x,\cdot) - \pi^*(\cdot) \| \leq M\rho^n \quad \forall\, n \geq 1,$$

en donde $M \geq 0$ y $0 < \rho < 1$ son constantes. Vea, por ejemplo, el artículo:

– A. Federgruen, A. Hordijk and H.C. Tijms, *A note on simultaneous recurrence conditions on a set of denumerable stochastic matrices*, J. Appl. Probab., **15** (1978), pp. 842–847.

14.14. Considere una CM con espacio de estados $S = \{1, 2, 3\}$ y matriz de transición

$$\begin{bmatrix} 0.80 & 0.05 & 0.15 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0 & 0.95 \end{bmatrix}.$$

(a) Demuestre que la CM es irreducible y aperiódica y calcule su distribución invariante.

Además, demuestre que:

(b) Existe un entero $\nu \geq 1$, un número $\rho > 0$ y un estado $s \in S$ tales que

$$\mathrm{P}_\nu(x,s) \geq \rho \quad \forall\, x \in S.$$

(c) Para una CM *arbitraria*, si se cumple (b) entonces también se cumplen las condiciones (equivalentes) (a) y (b) del Ejercicio 13.

15 Martingalas

Contenido. Martingalas, submartingalas y supermartingalas, teoremas de convergencia, tiempos de paro, teorema de muestreo opcional.

Para motivar los conceptos introducidos en esta sección, considérese una sucesión $\{X_n\}$ de vv.aa. en L_1, en donde X_n representa el capital de un jugador después de n jugadas. Sea X_0 el capital inicial del jugador. Si

$$E(X_{n+1}|X_0, \ldots, X_n) = X_n \quad \forall n = 0, 1, \ldots, \tag{15.1}$$

se dice que el juego es "honesto" o que $\{X_n\}$ es una **martingala**. En este caso, la propiedad de la esperanza iterada da que $EX_{n+1} = EX_n$ para todo $n = 0, 1, \ldots$, de modo que la ganancia esperada del jugador permanece constante:

$$EX_n = EX_0 \quad \forall\, n = 0, 1, \ldots.$$

Por otra parte, si

$$E(X_{n+1}|X_0, \ldots, X_n) \geq X_n \quad \forall\, n = 0, 1, \ldots, \tag{15.2}$$

se dice que el juego "está a favor" del jugador o que $\{X_n\}$ es una **submartingala**, y la ganancia esperada es no decreciente porque

$$EX_{n+1} \geq EX_n \quad \forall\, n = 0, 1, \ldots.$$

Finalmente, se dice que el juego "está en contra" del jugador o que $\{X_n\}$ es una **supermartingala** si

$$E(X_{n+1}|X_0, \ldots, X_n) \leq X_n \quad \forall\, n = 0, 1, \ldots, \tag{15.3}$$

en cuyo caso la ganacia esperada es no creciente pues

$$EX_{n+1} \leq EX_n.$$

Los conceptos en (1), (2), (3) se pueden extender a colecciones más generales de vv.aa. como sigue.

Definición 15.1. Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y T un subconjunto de $\mathbb{R}$ (por ejemplo, $T = \{0, 1, 2, \ldots\}$, $T = [a, b]$ ó $T = (-\infty, \infty)$). Sea $\{\mathrm{X}_t, t \in T\}$ una familia de vv.aa. sobre Ω y $\{\mathcal{F}_t, t \in T\}$ una familia de sub–σ–álgebras de $\mathcal{F}$. Decimos que:

(a) $\{\mathcal{F}_t, t \in T\}$ es una **filtración** de $\mathcal{F}$ si la familia es no decreciente, en el sentido de que

$$\mathcal{F}_s \subset \mathcal{F}_t \quad \forall\, s, t \in T, \ \text{ con } \ s < t;$$

(b) la familia $\{\mathrm{X}_t, t \in T\}$ **está adaptada** a la filtración $\{\mathcal{F}_t, t \in T\}$ si X_t es $\mathcal{F}_t$–medible para todo $t \in T$;

(c) $\{\mathrm{X}_t, t \in T\}$ es una **martingala con respecto a** $\{\mathcal{F}_t, \ t \in T\}$ (o que $\{\mathrm{X}_t, \mathcal{F}_t, \ t \in T\}$ es una martingala) si

(c$_1$) $\{\mathcal{F}_t, t \in T\}$ es una filtración de $\mathcal{F}$,

(c$_2$) $\{\mathrm{X}_t, t \in T\}$ está adaptada a $\{\mathcal{F}_t, t \in T\}$,

(c$_3$) X_t está en L_1 $\forall\, t \in T$, y

(c$_4$) $E(\mathrm{X}_t|\mathcal{F}_s) = \mathrm{X}_s$ $\forall\, s, t \in T$, con $s \leq t$.

Si la igualdad en (c$_4$) se sustituye por $\geq$ ó $\leq$, es decir, para todo $s \leq t$

$$E(\mathrm{X}_t|\mathcal{F}_s) \geq \mathrm{X}_s \quad \text{ó} \quad E(\mathrm{X}_t|\mathcal{F}_s) \leq \mathrm{X}_s,$$

se dice entonces que $\{\mathrm{X}_t, \mathcal{F}_t, t \in T\}$ es una **submartingala** o una **supermartingala**, respectivamente.

En esta sección consideraremos únicamente el caso en el que T es un conjunto de números enteros. Si $\mathrm{X}_\bullet := \{\mathrm{X}_n\}$ es una sucesión de vv.aa., entonces la familia $\mathcal{F}_\bullet^{\mathrm{X}} := \{\mathcal{F}_n^{\mathrm{X}}\}$ con

$$\mathcal{F}_n^{\mathrm{X}} := \sigma\{\mathrm{X}_0, \ldots, \mathrm{X}_n\} \quad \forall\, n = 0, 1, \ldots \tag{15.4}$$

se llama la **filtración natural** de $\mathrm{X}_\bullet$. Nótese que, efectivamente, $\{\mathcal{F}_n^{\mathrm{X}}\}$ es una filtración (porque $\mathcal{F}_n^{\mathrm{X}} \subset \mathcal{F}_{n+1}^{\mathrm{X}}$ para todo n) y que $\mathrm{X}_\bullet$ está adaptada a $\mathcal{F}_\bullet^{\mathrm{X}}$ (porque X_n es $\mathcal{F}_n^{\mathrm{X}}$–medible para todo n). Por lo tanto, la condición (1) coincide con la Definición 15.1(c) pues

$$E(\mathrm{X}_{n+1}|\mathcal{F}_n^{\mathrm{X}}) = \mathrm{X}_n \quad \forall\, n = 0, 1, \ldots, \tag{15.5}$$

y similarmente para (2) y (3).

Cuando $X_\bullet = \{X_n\}$ es una martingala con respecto a su filtración natural, se dice simplemente que $X_\bullet$ es una **martingala**. Para sub o supermartingalas se usa una terminología similar.

Si $\{X_n\}$ es una martingala con respecto a alguna filtración $\{\mathcal{F}_n\}$, entonces $\{X_n\}$ es una martingala con respecto a su filtración natural $\{\mathcal{F}_n^X\}$. Este hecho también se cumple para sub o supermartingalas.

Observación 15.2. (a) $\{X_n, \mathcal{F}_n,\ n = 0, 1, \ldots\}$ es una martingala ssi

$$E(X_{n+k}|\mathcal{F}_n) = X_n \quad \forall\, n \geq 0,\ k \geq 1.$$

Esta condición también se cumple para sub o supermartingalas reemplazando la igualdad por $\geq$ ó $\leq$, respectivamente.

(b) $\{X_n\}$ es una submartingala (con respecto a $\{\mathcal{F}_n\}$, digamos) ssi $\{-X_n\}$ es una supermartingala.

(c) Sean $X_0, X_1, \ldots$ vv.aa. en L_1. Si $X_{n+1} = X_n$ para todo n (en particular si $X_n \equiv c$, una constante, para todo n), entonces $\{X_n\}$ es una martingala. Si $X_{n+1} \geq X_n$ ó $X_{n+1} \leq X_n$ para todo n, entonces $\{X_n\}$ es una submartingala o una supermartingala, respectivamente. En otras palabras, una sucesión *monótona* de vv.aa. es una submartingala o una supermartingala dependiendo de que la sucesión sea creciente o decreciente, respectivamente.

Ejemplo 15.3. Sea $X_\bullet = \{X_1, X_2, \ldots\}$ una sucesión de vv.aa. independientes en L_1. Demuestre que:

(a) Si $EX_n = 0$ para todo n, entonces la sucesión $S_\bullet := \{S_n\}$, con $S_n := X_1 + \cdots + X_n$, es una martingala. (Más generalmente, $S_n := \sum_{k=1}^{n}(X_k - EX_k)$ para $n = 1, 2, \ldots$, es una martingala, aún si $EX_n \neq 0$.)

(b) Si $EX_n = 1$ para todo n, y $Y_n := X_1 \cdots X_n$, entonces $Y_\bullet := \{Y_n\}$ es una martingala.

Solución. (a) Nótese que $\sigma\{X_1, \ldots, X_n\} = \sigma\{S_1, \ldots, S_n\}$. (Explique.) Es evidente que $S_\bullet$ satisface las condiciones (c_1) a (c_3) de la Definición 15.1 y, por otra parte, como $S_{n+1} = S_n + \mathrm{X}_{n+1}$, obtenemos (c$_4$) pues

$$\begin{aligned} E(S_{n+1}|S_1,\ldots,S_n) &= E(S_n + \mathrm{X}_{n+1}|\mathrm{X}_1,\ldots,\mathrm{X}_n) \\ &= S_n + E\mathrm{X}_{n+1} \qquad \text{(explique)} \\ &= S_n. \end{aligned}$$

(b) Sea $\mathcal{F}_n^Y := \sigma\{Y_1,\ldots,Y_n\}$ la filtración natural de $Y_\bullet$, de modo que $Y_\bullet$ está adaptada a $\{\mathcal{F}_n^Y\}$. Además, Y_n está en L_1 para todo n porque $E|Y_n| = E|\mathrm{X}_1|\cdots E|\mathrm{X}_n| < \infty$. Finalmente,

$$\begin{aligned} E(Y_{n+1}|\mathcal{F}_n) &= E(Y_n\mathrm{X}_{n+1}|\mathcal{F}_n) \\ &= Y_n E(\mathrm{X}_{n+1}) \qquad \text{(explique)} \\ &= Y_n. \quad \square \end{aligned}$$

Por lo tanto, $Y_\bullet$ es una martingala.

Proposición 15.4. (a) Sea $\{\mathrm{X}_n\}$ una martingala y $h : \mathbb{R} \to \mathbb{R}$ una función convexa tal que $h(\mathrm{X}_n)$ está en L_1 para todo n. Entonces $\{h(\mathrm{X}_n)\}$ es una submartingala.

(b) Si $\{\mathrm{X}_n\}$ es una submartingala y h es una función convexa, no–decreciente y tal que $h(\mathrm{X}_n)$ está en L_1 para todo n, entonces también $\{h(\mathrm{X}_n)\}$ es una submartingala.

Demostración. Sea $\mathcal{F}_n^{\mathrm{X}} := \sigma\{\mathrm{X}_1,\ldots,\mathrm{X}_n\}$. Por la desigualdad de Jensen 11.12(g),

$$\begin{aligned} E[h(\mathrm{X}_{n+1})|\mathcal{F}_n^{\mathrm{X}}] &\geq h[E(\mathrm{X}_{n+1}|\mathcal{F}_n^{\mathrm{X}})] \\ &= h(\mathrm{X}_n). \end{aligned}$$

Esto demuestra (a). La demostración de (b) es similar. $\square$ $\square$

Ejemplo 15.5. Si $\{\mathrm{X}_n\}$ es una martingala, entonces $\{\mathrm{X}_n^+\}$, $\{\mathrm{X}_n^2\}$ y $\{e^{\alpha \mathrm{X}_n}\}$ son submartingalas para cualquier $\alpha \in \mathbb{R}$, en donde $\mathrm{X}_n^+ := \max\{\mathrm{X}_n, 0\}$. Asimismo, si $\{\mathrm{X}_n\}$ es una submartingala, entonces también lo son $\{\mathrm{X}_n^+\}$ y $\{e^{\alpha \mathrm{X}_n}\}$ para $\alpha \geq 0$.

Ahora queremos mostrar resultados importantes sobre convergencia de martingalas, para lo cual necesitaremos la siguiente desigualdad cuya prueba sigue la idea de la desigualdad de Kolmogorov.

Teorema 15.6. (desigualdad de Doob-Kolmogorov). *Si $\{X_n\}$ es una martingala, entonces para todo $\epsilon > 0$*

$$P(\max_{k\leq n} |X_k| \geq \epsilon) \leq \frac{1}{\epsilon} E(|X_n| I_A) \leq \frac{E(|X_n|)}{\epsilon},$$

donde $A := \{\max_{k=1,\ldots,n} |X_k| \geq \epsilon\}$.

Proof. Sean

$$A_k := \{|X_1| < \epsilon, \ldots, |X_{k-1}| < \epsilon, |X_k| \geq \epsilon\},\ \ k = 1, \ldots, n.$$

Luego, como $\{|X_n|\}$ es también submartingala y $A_k \in \mathcal{F}_k$ $k = 1, 2, \ldots, n$,

$$E(|X_n| I_{A_k}) \geq E(|X_k| I_{A_k}) \geq \epsilon P(A_k).$$

Y como los conjuntos A_k's son disjuntos y la union de todos es A, la suma sobre k de la desigualdad anterior da el resultado. □

Teorema 15.7. (Convergencia de martingalas en L_2.) *Sea $\{X_n\}$ una martingala con $E(X_n^2) \leq c < \infty$ para todo n. Entonces existe una v.a. X tal que $X_n \to X$ c.s. y en L_2. Además, $EX_n = EX$ para todo n.*

Demostración. Vemos que $E(X_m(X_{m+n} - X_m)) = 0$, luego

$$E(X_{m+n}^2) = E(X_m^2) + E((X_{m+n} - X_m)^2). \tag{15.6}$$

Esto dice que $\{E(X_n^2)\}_{n=0}^{\infty}$ no es decreciente. Además, por hipótesis, está acotada, así que $E(X_n^2)$ converge a una constante $C \leq c$. Probemos que para casi toda $\omega \in \Omega$, $\{X_n(\omega)\}_{n=0}^{\infty}$ es de Cauchy, lo cual implicaría que $X_n \to X$ c.s., con $X(\omega) := \lim_{n\to\infty} X_n(\omega)$ cuando $\omega \in A$ y $X(\omega) := 0$ cuando $\omega \notin A$, donde

$$A := \{\omega \in \Omega : \{X_n(\omega)\} \text{ es de Cauchy}\}.$$

Notamos que

$$A = \{\omega \in \Omega : (\forall \epsilon > 0)\ \exists m \text{ tal que } (\forall k, j \geq 0)|X_{m+k}(\omega) - X_{m+j}(\omega)| < \epsilon\}.$$

Por la desigualdad del triangulo

$$|X_{m+k} - X_{m+j}| \leq |X_{m+k} - X_m| + |X_{m+j} - X_m|.$$

Por lo tanto,

$$\begin{aligned} A &= \{\omega \in \Omega : (\forall \epsilon > 0)\ \exists m \text{ tal que } (\forall k \geq 0)|X_{m+k}(\omega) - X_m(\omega)| < \epsilon\}. \\ &= \bigcap_{\epsilon>0} \bigcup_{m\geq 0} \{|X_{m+k} - X_m| < \epsilon,\ k \geq 0\}. \end{aligned}$$

Así, $A^c = \bigcup_{\epsilon>0} \bigcap_{m\geq 0} A_m(\epsilon)$, donde

$$A_m(\epsilon) := \{|X_{m+k} - X_m| \geq \epsilon \text{ para algún } k \geq 0\}.$$

Como $A_m(\epsilon_2) \subset A_m(\epsilon_1)$ cuando $\epsilon_2 \geq \epsilon_1$, entonces

$$P(A^c) = \lim_{\epsilon\to 0} P(\bigcap_{m\geq 0} A_m(\epsilon)) \leq \lim_{\epsilon\to 0} \lim_{m\to\infty} P(A_m(\epsilon)).$$

Basta ver que $\lim_{m\to\infty} P(A_m(\epsilon)) = 0$ para todo $\epsilon > 0$.

Definamos para cada $m \geq 0$, $Y_n^{(m)} := X_{m+n} - X_m$, $n = 1, 2, \ldots$, y veamos que $\{Y_n^{(m)}\}_{n=1}^{\infty}$ es una martingala c.r.a su filtración natural. En efecto,

$$E(Y_{n+1}^{(m)}|Y_1^{(m)}, \ldots, Y_n^{(m)}) = E(E(Y_{n+1}^{(m)}|\mathcal{F}_{m+n})|Y_1^{(m)}, \ldots, Y_n^{(m)}) = Y_n^{(m)}.$$

Ahora, aplicando la desigualdad de Doob-Kolmogorov (junto con la desigualdad de Hölder) a $\{Y_n^{(m)}\}_{n=1}^{\infty}$ tenemos

$$P\left(\max_{k\leq n} |X_{m+k} - X_m| \geq \epsilon\right) \leq \frac{1}{\epsilon}\sqrt{E((X_{m+n} - X_m)^2)}.$$

Usando (15.6) y tomando $n \to \infty$, llegamos a que

$$P(A_m(\epsilon)) \leq \frac{1}{\epsilon}\sqrt{(C - E(X_m^2))}.$$

Y si ahora $m \to \infty$, entonces $P(A_m(\epsilon)) \to 0$. Esto implica que $\{X_n\}$ es de Cauchy c.s., lo cual demuestra la primera conclusión del teorema. Finalmente, usando el Lema de Fatou,

$$E((X_n - X)^2) \leq \liminf_m E((X_n - X_m)^2) = C - E(X_n^2).$$

Tomando el límite cuando $n \to \infty$ se obtiene que $X_n \to X$ en L_2 y también que $EX_n = EX$ para todo n. □

Ejemplo 15.8. Sea $\{\mathrm{X}_n\}$ una sucesión de vv.aa. tales que $E(\mathrm{X}_n|\mathrm{X}_1,\ldots,\mathrm{X}_{n-1}) = 0$ para todo n, y sea $S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n$. Demuestre que si $\sum_{k=1}^{\infty} E(\mathrm{X}_k^2)/k^2 < \infty$, entonces $n^{-1}S_n \to 0$ c.s.

Demostración. Sea $Y_n := \sum_{k=1}^{n} \mathrm{X}_k/k$. Entonces $\{Y_n\}$ es una martingala porque

$$\begin{aligned} E(Y_n|Y_1,\ldots,Y_{n-1}) &= E(Y_{n-1} + \mathrm{X}_n/n|Y_1,\ldots,Y_{n-1}) \\ &= Y_{n-1}. \qquad \text{(explique.)} \end{aligned}$$

Además, existe una constante c tal que $E(Y_n^2) \leq c$ para todo n, pues

$$\begin{aligned} E(Y_n^2) &= \sum_{k=1}^{n} E(\mathrm{X}_k^2)/k^2 + \sum_{j\neq k} E(\mathrm{X}_j\mathrm{X}_k)/jk \\ &= \sum_{k=1}^{n} E(\mathrm{X}_k^2)/k^2 \quad \text{(explique)} \\ &\leq \sum_{k=1}^{\infty} E(\mathrm{X}_k^2)/k^2 < \infty. \end{aligned}$$

Por lo tanto $\{Y_n\}$ satisface las hipótesis del Teorema 15.7, así que existe una v.a. Y tal que, en particular, $Y_n \to Y$ c.s. Luego, por el Lema de Kronecker (Ejercicio 15.10),

$$\frac{1}{n}\sum_{k=1}^{n} k(\mathrm{X}_k/k) = \frac{1}{n}S_n \to 0 \quad \text{c.s.} \qquad \square$$

$\square$

Como caso especial del ejemplo anterior, observe que si $\{\mathrm{X}_n\}$ es una sucesión de vv.aa. independientes con media 0 y $\sum_{k=1}^{\infty} E(\mathrm{X}_k^2)/k^2 < \infty$, entonces $n^{-1}S_n \to 0$ c.s.

Ahora mostraremos el teorema de convergencia cuando las variables estan en L_1, para lo cual necesitamos el concepto de "número de cruces".

Definición 15.9. Sea $X_\bullet := \{X_n\}$ un proceso estocástico y sean $a, b \in \mathbb{R}$ con $a < b$. A la variable aleatoria

$$U_n([a,b], X_\bullet) := \max\left\{k \in \mathbb{N} : \begin{array}{c} \exists\ 0 \leq t_1 < s_1 < \ldots < t_k < s_k \leq n \\ \text{tal que } X_{t_i} \leq a \text{ y } X_{s_i} \geq b,\ i = 1,\ldots,k \end{array}\right\}$$

se le llama el **número de cruces superiores** de $[a, b]$, del proceso $X_\bullet$. Simplemente cuenta cuantas veces antes de n el proceso cruza el intervalo completo de abajo hacia arriba.

Nota 15.10. Los tiempos de cruce podemos definirlos como

$t_1 := \inf\{n \geq 0 : X_n \leq a\},\ s_1 := \inf\{n > t_1 : X_n \geq b\},\ t_2 := \inf\{n > s_1 : X_n \leq a\}, \ldots$

Lema 15.11. Si $X_\bullet := \{X_n\}_{n=0}^\infty$ es una submartingala c.r.a $\{\mathcal{F}_n\}$, entonces

$$(b-a)E(U_n([a,b], X_\bullet)) \leq E\left(\max(X_n - a, 0)\right) - E\left(\max(X_0 - a, 0)\right).$$

Demostración. Observamos que $Z_n := \max(X_n - a, 0)$ define una submartingala c.r.a $\{\mathcal{F}_n\}$, denotemosla $Z_\bullet$. Observemos también que

$$U_n([a,b], X_\bullet) = U_n([0, b-a], Z_\bullet),\ n = 0, 1, \ldots$$

Ahora definamos $C_1 := I_{\{X_0 \leq a\}}$ e inductivamente

$$C_n := I_{\{C_{n-1}=1\}} I_{\{X_{n-1} < b\}} + I_{\{C_{n-1}=0\}} I_{\{X_{n-1} \leq a\}},$$

por lo que $\{C_n\}_{n=1}^\infty$ es previsible, i.e. $C_n \sim \mathcal{F}_{n-1}$. Tenemos que

$$(b-a)U_n([0, b-a], Z_\bullet) \leq \sum_{i=1}^n (Z_i - Z_{i-1})C_i.$$

Sin embargo, usando que $E(Z_i - Z_{i-1} | \mathcal{F}_{i-1}) \geq 0$, tenemos

$$\begin{aligned} E((Z_i - Z_{i-1})C_i) &= E(E((Z_i - Z_{i-1})C_i | \mathcal{F}_{i-1})) \\ &= E(C_i E((Z_i - Z_{i-1}) | \mathcal{F}_{i-1})) \\ &\leq E(E(Z_i - Z_{i-1} | \mathcal{F}_{i-1})) \\ &= E(Z_i) - E(Z_{i-1}). \end{aligned}$$

Así, al tomar la esperanza surge la cota de una suma telescópica. □

Corolario 15.12. *Con las hipótesis anteriores y suponiendo que $E(X_n) \leq M < \infty$ para toda n, se tiene que*

$$P(U_\infty([a,b], X_\bullet) = \infty) = 0.$$

Demostración. Como $E(\max(X_n - a, 0)) \leq M + |a|,\ n = 0, 1, \ldots$, por el lema anterior

$$E(U_\infty([a,b], X_\bullet)) < \infty.$$

Lo cual implica el resultado. $\square$

Teorema 15.13. (Convergencia de Doob en L_1). *Sea $\{\mathrm{X}_n, \mathcal{F}_n\}$ una submartingala tal que $M := \sup_n E|\mathrm{X}_n| < \infty$. Entonces existe una v.a. $\mathrm{X} \in L_1$ tal que $\mathrm{X}_n \to \mathrm{X}$ c.s. y, además, $E|\mathrm{X}| \leq \sup_n E|\mathrm{X}_n|$.*

Demostración. Tenemos primero que

$$\begin{aligned} A &:= \{\omega : \{X_n(\omega)\} \text{ no converge en } \mathbb{R}\} \\ &= \{\omega : \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} \\ &= \bigcup_{a<b\in\mathbb{Q}} A_{a,b}, \end{aligned}$$

donde

$$A_{a,b} := \{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\}.$$

Como $A_{a,b} \subset \{U_\infty([a,b], X_\bullet) = \infty\}$, por el corolario anterior, $P(A_{a,b}) = 0$, luego $P(A) = 0$. Es decir, $X(\omega) := \lim_{n\to\infty} X_n(\omega)$ existe para casi toda $\omega \in \Omega$. Además, por el Lema de Fatou,

$$E(|X|) \leq \liminf_n E(|X_n|) \leq M < \infty.$$

$\square$

Observación 15.14. (a) Nótese que $|\mathrm{X}_n| = \mathrm{X}_n^+ + \mathrm{X}_n^- \geq \mathrm{X}_n^+$, de modo que $E|\mathrm{X}_n| \geq E\mathrm{X}_n^+$. Por otra parte, como $|\mathrm{X}_n| = 2\mathrm{X}_n^+ - \mathrm{X}_n$ (explique), vemos que si $\{\mathrm{X}_n\}$ es una submartingala, entonces

$$E|\mathrm{X}_n| = 2E\mathrm{X}_n^+ - E\mathrm{X}_n \leq 2E\mathrm{X}_n^+ - E\mathrm{X}_1.$$

Por lo tanto, la condición $\sup_n E|\mathrm{X}_n| < \infty$ en el Teorema 15.13 es equivalente a

$$\sup_n E\mathrm{X}_n^+ < \infty.$$

De aquí se sigue que cualquier submartingala negativa converge c.s.

(b) Para *supermartingalas*, el resultado análogo a 15.13 es el siguiente. Si $\{\mathrm{X}_n, \mathcal{F}_n\}$ es una supermartingala y $\sup_n E(\mathrm{X}_n^-) < \infty$, entonces existe una v.a.

X tal que $X_n \to X$ c.s. En particular, cualquier supermartingala no–negativa converge c.s.

(c) En el Teorema 15.13, si $\{X_n\}$ es uniformemente integrable (u.i.), entonces se puede mostrar también que $X_n \to X$ en L_1. Recuerde que una sucesión de vv.aa. es u.i. si y solo si

$$\sup_n E(|X_n| I_{\{|X_n| \geq r\}}) \to 0, \quad \text{cuando } r \to \infty.$$

Definición 15.15. Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y $\{\mathcal{F}_n\}$ una filtración de $\mathcal{F}$. Decimos que una v.a. $T : \Omega \to \mathbb{N} \cup \{+\infty\}$ (con $\mathbb{N} := \{0, 1, \ldots\}$) es un **tiempo de paro** con respecto a $\{\mathcal{F}_n\}$ si el evento $\{T \leq n\}$ está en $\mathcal{F}_n$ para todo $n \in \mathbb{N}$. Asimismo, un tiempo de paro T es **finito** si $\mathrm{P}(T < \infty) = 1$, y **acotado** si existe una constante c tal que $\mathrm{P}(T \leq c) = 1$. Además, si T es un tiempo de paro finito definimos la v.a. X_T como

$$X_T(\omega) := X_{T(\omega)}(\omega).$$

Es decir,

$$X_T(\omega) = \sum_{n=0}^{\infty} X_n(\omega) I_{\{T(\omega)=n\}}. \tag{15.7}$$

Dado un tiempo de paro T, definimos

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid A \cap \{T \leq n\} \in \mathcal{F}_n \; \forall\, n\}. \tag{15.8}$$

Se puede demostrar que $\mathcal{F}_T$ es una σ–álgebra y que X_T es $\mathcal{F}_T$–medible. (Ejercicio 12.)

Proposición 15.16. Si $P(T < \infty) = 1$, entonces

$$\sigma(\{X_T\}) \subset \mathcal{F}_T.$$

Demostración. Tenemos que

$$\sigma(\{X_T\}) = \{X^{-1}(B) : B \subset \mathbb{R} \text{ de Borel}\}.$$

Como $X_T = \sum_{k=0}^{\infty} X_k I_{\{T=k\}}$, entonces

$$A := X_T^{-1}(B) = \bigcup_{k=0}^{\infty} \{T = k, X_k \in B\}.$$

Por lo que

$$A \bigcap \{T = k\} = \{T = k\} \bigcap \{X_k \in B\} \in \mathcal{F}_k$$

para cada $k = 0, 1, \ldots$. □

Teorema 15.17. *Sea $\{X_n\}$ una submartingala c.r.a $\mathcal{F}_\bullet := \{\mathcal{F}_n\}$ y T un tiempo de paro c.r.a $\mathcal{F}_\bullet$. Entonces $\{X_{T\wedge n}\}$ es una submartingala c.r.a $\mathcal{F}_\bullet$.*

Demostración. Sea $Z_n := X_{T\wedge n}$, entonces

$$Z_n = \sum_{k=0}^{n-1} X_k I_{\{T=k\}} + X_n I_{\{T\geq n\}},$$

por lo que $Z_n \sim \mathcal{F}_n$ y $E(|Z_n|) \leq \sum_{k=0}^{n} E(|X_k|) < \infty$. Además,

$$Z_{n+1} - Z_n = (X_{n+1} - X_n) I_{\{T>n\}},$$

luego

$$E(Z_{n+1} - Z_n|\mathcal{F}_n) = I_{\{T>n\}} E(X_{n+1} - X_n|\mathcal{F}_n) \geq 0,$$

pues $E(X_{n+1}|\mathcal{F}_n) \geq X_n$. □

Teorema 15.18. *Sean $\{X_n\}$ y $\{Y_n\}$ submartingalas c.r.a $\mathcal{F}_\bullet$ y T un tiempo de paro c.r.a $\mathcal{F}_\bullet$. Suponga que $X_T = Y_T$ en $\{T < \infty\}$, entonces la siguiente es también es una submartingala c.r.a $\mathcal{F}_\bullet$*

$$Z_n := \begin{cases} X_n & n < T \\ Y_n & n \geq T \end{cases}, n = 0, 1, \ldots$$

Demostración. Tenemos que $Z_n = X_n I_{\{T>n\}} + Y_n I_{\{T\leq n\}}$, por lo que $Z_n \sim \mathcal{F}_n$ y $E(|Z_n|) < \infty$. Además,

$$\begin{aligned} E(Z_{n+1}|\mathcal{F}_n) &= E(X_{n+1} I_{\{T>n+1\}} + Y_{n+1} I_{\{T\leq n+1\}}|\mathcal{F}_n) \\ &= E(X_{n+1}(I_{\{T>n\}} - I_{\{T=n+1\}}) + Y_{n+1}(I_{\{T\leq n\}} + I_{\{T=n+1\}})|\mathcal{F}_n) \\ &\geq Z_n - E((X_{n+1} - Y_{n+1}) I_{\{T=n+1\}}|\mathcal{F}_n) = Z_n. \end{aligned}$$

$\square$

Nos interesa ahora mostrar el teorema de muestreo opcional de Doob, para ellos tenemos primero los siguientes dos lemas.

Lema 15.19. Sean $\mathcal{F}_1$ y $\mathcal{F}_2$ dos σ-álgebras en $\mathcal{F}$, y X y Y dos variables aleatorias en L_1. Suponga que hay $A \in \mathcal{F}_1 \bigcap \mathcal{F}_2$ tal que

$$(\forall \omega \in A)\ X(\omega) = Y(\omega),$$

y

$$\mathcal{F}_1 \bigcap A = \mathcal{F}_2 \bigcap A.$$

Entonces

$$(\forall \omega \in A)\ E[X|\mathcal{F}_1](\omega) = E[Y|\mathcal{F}_2](\omega).$$

Lema 15.20. Sean $S \leq T$ dos tiempos de paro. Se cumple que

$$\mathcal{F}_T \bigcap \{T = k\} = \mathcal{F}_k \bigcap \{T = k\} \text{ y } \mathcal{F}_S \subset \mathcal{F}_T.$$

Teorema 15.21. (Muestreo opcional, Doob). *Sea X una martingala y $\mathcal{F}_n$ su filtración. Sea $N < \infty$ y $S \leq T \leq N$ dos tiempos de paro. Entonces*

$$E[X_T|\mathcal{F}_S] = X_S \ c.s.$$

Demostración. Por los dos lemas anteriores,

$$\begin{aligned}(\forall \omega \in \{T = k\}) E[X_N|\mathcal{F}_T](\omega) &= E[X_N|\mathcal{F}_k](\omega) \\ &= X_k(\omega) = X_T(\omega).\end{aligned}$$

Luego

$$\begin{aligned} &= \sum_{k=1}^{N} E[X_N|\mathcal{F}_T] I_{\{T=k\}} \\ E[X_N|\mathcal{F}_T] &= \sum_{k=1}^{N} X_T I_{\{T=k\}} \\ &= X_T \ c.s. \end{aligned}$$

De nuevo usando un lema,

$$\begin{aligned} E[X_T|\mathcal{F}_S] &= E[E[X_N|\mathcal{F}_T]|\mathcal{F}_S] \\ &= E[X_N|\mathcal{F}_S] \\ &= X_S \ c.s. \end{aligned}$$

$\square$

Nota 15.22. Los Teoremas 15.17, 15.18 y 15.21 son válidos también para supermartingalas, con las correspondientes desigualdades, y esto implica que son también válidos para martingalas, cambiando las correspondientes desigualdades por igualadades.

Ahora queremos verificar la ecuación del teorema opcional que tiene diferentes aplicaciones. Para el siguiente teorema, podemos comparar las condiciones con el siguiente contraejemplo. Considere la caminata aleatoria simétrica S_n, con $S_0 = 0$, con el tiempo de paro $T = \inf\{n : S_n = 1\}$, y veamos que NO se cumple la ecuación

$$E[S_T] = E[S_0].$$

Teorema 15.23. ***(Paro opcional de Doob)****. Sea $\{X_n\}$ una martingala y T un timepo de paro, ambos c.r.a $\mathcal{F}_\bullet$. Si se cumple una de las siguientes, entonces*

$$E(X_T) = E(X_0).$$

i) T es acotada.
ii) $(\forall n)|X_n| \leq M < \infty$ y $P(T < \infty) = 1$.
iii) $E(T) < \infty$ y $(\forall n \geq 1)|X_n - X_{n-1}| \leq M < \infty$.
iv) $E(X_T) < \infty$, $P(T < \infty) = 1$ y $\lim_{k\to\infty} E(X_k I_{\{T\geq k\}}) = 0$.
v) $P(T < \infty) = 1$ y $\{X_n\}$ u.i.

Demostración.
i) Teorema anterior.
ii) Tenemos que $\{X_{T\wedge n}\}$ es supermartingala, entonces $E(X_{T\wedge n}|\mathcal{F}_0) \leq X_0$, luego $E(X_{T\wedge n} - X_0) \leq 0$, $n = 0, 1, \ldots$. Como $X_{T\wedge n} - X_0 \leq 2M$, por el teorema de convergencia dominada

$$E(X_T - X_0) = \lim_{n\to\infty} E(X_{T\wedge n} - X_0) \leq 0.$$

Note que de $P(T < \infty) = 1$, efectivamente $X_{T\wedge n} \to X_T$ c.s. Para la otra desigualdad usar que $\{X_{T\wedge n}\}$ es submartingala.
iii) Tenemos que

$$|X_{T\wedge n} - X_0| = \left|\sum_{K=1}^{T\wedge n}(X_k - X_{k-1})\right| \leq MT.$$

Como $E(T) < \infty$, por el teorema de convergencia dominada, y usando que $\{X_{T\wedge n}\}$ es supermartingala, entonces

$$E(X_T - X_0) = \lim_{n\to\infty} E(X_{T\wedge n} - X_0) \le 0.$$

Ahora podemos repetir el argumento con la submartingala.
iv) Escribimos

$$X_T = X_{T\wedge n} + (X_T - X_n)I_{\{T\ge n\}}.$$

Como $\{X_{T\wedge n}\}$ es supermartingala, $E(X_{T\wedge n}) \le E(X_0)$, luego

$$E(X_T) \le E(X_0) + E(X_T I_{\{T\ge n\}}) - E(X_n I_{\{T\ge n\}}).$$

Por hipótesis, $\lim_{n\to\infty} E(X_n I_{\{T\ge n\}}) = 0$. Vemos también que

$$E(X_T I_{\{T\ge n\}}) \le \sum_{k=n}^{\infty} E(|X_k| I_{\{T=k\}}).$$

Como $P(T < \infty) = 1$

$$E(|X_T|) = \sum_{k=0}^{\infty} E(|X_k| I_{\{T=k\}}),$$

entonces, de $E(|X_T|) < \infty$, se concluye que $\lim_{n\to\infty} E(X_T I_{\{T\ge n\}}) = 0$. Para la otra desigualdad usar la submartingala.
v) Para $r > 0$,

$$E(|X_n|) = E(|X_n| I_{\{|X_n|<r\}}) + E(|X_n| I_{\{|X_n|\ge r\}}).$$

Luego

$$\sup_n E(|X_n|) \le r + \sup_n E(|X_n| I_{\{|X_n|\ge r\}}),$$

y por integrabilidad uniforme $\sup_n E(|X_n|) < \infty$. Así, por el teorema de convergencia de Doob, hay $Z \in L_1$ talque $X_n \overset{c.s.}{\to} Z$ y por la Observación 15.14(c) $X_n \overset{L_1}{\to} Z$. Luego, para $A \in \mathcal{F}_N$, con $N \in \mathbb{N}$ fijo,

$$|E((X_n - Z)I_A)| \le E(|X_n - Z|) \to 0,\ n \to \infty.$$

Entonces

$$E(ZI_A) = \lim_{n\to\infty} E(X_n I_A) = \lim_{n\to\infty} E(I_A E(X_n|\mathcal{F}_N)) = E(X_N I_A),$$

por lo tanto $X_N = E(Z|\mathcal{F}_N)$.
También, para toda $A \in \mathcal{F}_T$, $A \bigcap \{T = k\} \in \mathcal{F}_k$, por lo que

$$\begin{aligned} E(X_T I_A) &= \sum_{k=0}^{\infty} E(X_k I_{A\bigcap\{T=k\}}) \\ &= \sum_{k=0}^{\infty} E(Z I_{A\bigcap\{T=k\}}) = E(Z I_A). \end{aligned}$$

Por lo tanto $X_T = E(Z|\mathcal{F}_T)$ c.s. Finalmente

$$E(X_T|\mathcal{F}_0) = E(E(Z|\mathcal{F}_T)|\mathcal{F}_0) = E(Z|\mathcal{F}_0) = X_0.$$

□

Ejercicios § 15

15.1. Sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ vv.aa. tales que las sumas $S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n$ forman una martingala. Demuestre que $E(\mathrm{X}_i \mathrm{X}_j) = 0$ si $i \neq j$.

15.2. Sean $\mathrm{X}_0, \mathrm{X}_1, \ldots$ vv.aa. en L_1 y tales que

$$E(\mathrm{X}_{n+1}|\mathrm{X}_0, \mathrm{X}_1, \ldots, \mathrm{X}_n) = a\mathrm{X}_n + b\mathrm{X}_{n-1} \quad \forall\, n \geq 1,$$

en donde $0 < a, b < 1$ y $a + b = 1$. Encuentre un valor de α para el cual la sucesión de vv.aa. $S_n := \alpha \mathrm{X}_n + \mathrm{X}_{n-1}$, para $n \geq 1$, es una martingala con respecto a $\mathcal{F}_n := \sigma\{\mathrm{X}_0, \ldots, \mathrm{X}_n\}$.

15.3. Sea $\{\mathrm{X}_1, \mathrm{X}_2, \ldots\}$ una sucesión de vv.aa. independientes tales que $E\mathrm{X}_n = m_n \neq 0$ para todo $n \geq 1$. Demuestre que la sucesión

$$Y_n := \prod_{k=1}^{n} (\mathrm{X}_k / m_k) \quad \text{para} \quad n = 1, 2, \ldots$$

es una martingala con respecto a $\mathcal{F}_n := \sigma\{\mathrm{X}_1, \ldots, \mathrm{X}_n\}$.

15.4. Sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ vv.aa. independientes tales que

$$\mathrm{P}\{\mathrm{X}_n = 1\} = p \quad \text{y} \quad \mathrm{P}\{\mathrm{X}_n = -1\} = 1 - p =: q \quad \forall\, n = 1, 2, \ldots,$$

donde $0 < p < 1$. Sean

$$S_n := \mathrm{X}_1 + \cdots + \mathrm{X}_n \quad \text{y} \quad Y_n := (q/p)^{S_n}.$$

Demuestre que $\{Y_n\}$ es una martingala con respecto a $\mathcal{F}_n := \sigma\{\mathrm{X}_1, \ldots, \mathrm{X}_n\}$.

15.5. Sean $X_1, X_2, \ldots$ vv.aa. i.i.d. con $EX_1 = 0$ y $E(X_1^2) = \sigma^2 < \infty$, y sea $S_n := X_1 + \cdots + X_n$. Demuestre que la sucesión

$$Y_n := S_n^2 - n\sigma^2 \quad \text{para} \quad n = 1, 2, \ldots$$

es una martingala con respecto a $\mathcal{F}_n := \sigma\{X_1, \ldots, X_n\}$.

15.6. Sea X una v.a. en $L_1 \equiv L_1(\Omega, \mathcal{F}, P)$ y $\{\mathcal{F}_n\}$ una filtración de $\mathcal{F}$. Demuestre que las vv.aa. $X_n := E(X|\mathcal{F}_n)$, para $n = 1, 2, \ldots,$ forman una martingala. (*Sugerencia:* use la Proposición 11.12(f).)

15.7. (**Descomposición Doob-Meyer de submartingalas**) Sea $\{X_n, \mathcal{F}_n, n \geq 1\}$ una submartingala. Además, sea

$$Y_n := X_1 + \sum_{k=1}^{n-1}[X_{k+1} - E(X_{k+1}|\mathcal{F}_k)] \;\text{ para }\; n \geq 2, \;\text{ con }\; Y_1 := X_1, \;\text{ y}$$

$$Z_n := \sum_{k=1}^{n-1}[E(X_{k+1}|\mathcal{F}_k) - X_k] \;\text{ para }\; n \geq 2, \;\text{ con }\; Z_1 := 0$$

Demuestre que:

(a) $\{Y_n, \mathcal{F}_n\}$ es una martingala,

(b) $\{Z_n\}$ es una sucesión no–decreciente de vv.aa. no–negativas y tales que Z_n es $\mathcal{F}_{n-1}$–medible para $n \geq 2$, y

(c) $X_n = Y_n + Z_n$.

15.8. Demuestre que si $\{X_n\}$ es una submartingala, entonces también lo es $\{\max(X_n, a)\}$ para cualquier $a \in \mathbb{R}$.

15.9. (**Lema de Toeplitz**) Si $\{a_n\}$ es una sucesión de números reales tales que $a_n \to a$ cuando $n \to \infty$, entonces $\frac{1}{n}\sum_{k=1}^n a_k \to a$ cuando $n \to \infty$.

15.10. (**Lema de Kronecker**) Sea $\{a_n\}$ una sucesión de números reales tal que la serie $\sum_{n=1}^\infty a_n$ converge. Entonces $\frac{1}{n}\sum_{k=1}^n k\, a_k \to 0$ cuando $n \to \infty$. (*Sugerencia:* sea s_n la suma parcial $s_n := a_1 + \cdots + a_n$ para $n \geq 1$, y $s_0 := 0$. Verifique que

$$\frac{1}{n}\sum_{k=1}^n k\, a_k = s_n - \frac{1}{n}\sum_{k=1}^{n-1} s_k$$

y después use el Ejercicio 9.)

15.11. Demuestre que T es un tiempo de paro (Definición 12.10) ssi el evento $\{T = n\}$ está en $\mathcal{F}_n$ para todo n.

15.12. Demuestre que $\mathcal{F}_T$ (definida en (7)) es una σ–álgebra y que X_T es $\mathcal{F}_T$–medible.

15.13. Sean S y T tiempos de paro con respecto a una filtración $\{\mathcal{F}_n\}$. Demuestre:

(a) Si $S \leq T$, entonces $\mathcal{F}_S \subset \mathcal{F}_T$.

(b) Si $T \equiv n$, entonces $\mathcal{F}_T = \mathcal{F}_n$.

(c) $\min(S, T)$ y $\max(S, T)$ son tiempos de paro.

Definición. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ una CM con espacio de estados S (numerable) y matriz de transición $\mathrm{P} = [\mathrm{P}(x, y)]$. Dada una función $h : S \to \mathbb{R}$ definimos

$$\mathrm{P}h(x) := E[h(\mathrm{X}_{n+1})|\mathrm{X}_n = x] = \sum_{y \in S} \mathrm{P}(x, y)h(y) \quad \forall\, x \in S.$$

Decimos que h es una función **armónica** (o **invariante**) con respecto a P si $\mathrm{P}h(x) = h(x)$ para todo $x \in S$. Por otra parte, si $\mathrm{P}h(x) \geq h(x)$ o $\mathrm{P}h(x) \leq h(x)$ para todo $x \in S$, decimos, respectivamente, que h es **sub–armónica** o que h es **super–armónica**.

15.14. Demuestre que si h es armónica (respectivamente, sub o super–armónica), entonces $h(\mathrm{X}_n)$ es una martingala (respectivamente, sub o super–martingala).

15.15. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ una CM con espacio de estados S, y sean $A \subset S$ y T_A el tiempo del primer paso a A; vea el Ejemplo 12.12(b). Para cada $x \in S$, sea

$$h(x) := \mathrm{P}_x(T_A < \infty) = \sum_{n=1}^{\infty} \mathrm{P}_x(T_A = n).$$

Demuestre que h es una función armónica y, por lo tanto (por el Ejercicio 15.14), $h(\mathrm{X}_n)$ es una martingala.

15.16. Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$ un PE con espacio de estados S, y sea $\mathrm{P} = [\mathrm{P}(x, y)]$ una matriz estocástica (con x, y en S). Sea $\mathcal{F}_n := \sigma\{\mathrm{X}_0, \ldots, \mathrm{X}_n\}$ la filtración natural de $\mathrm{X}_\bullet$. Demuestre que las siguientes afirmaciones son equivalentes:

(a) $X_\bullet$ es una CM con matriz de transición P.

(b) Para cualquier función acotada $h : S \to \mathbb{R}$, el PE definido para $n \geq 1$ como

$$M_n(h) := h(X_n) - h(X_0) - \sum_{k=0}^{n-1}[\mathrm{P}h(X_k) - h(X_k)], \quad M_0(h) := 0,$$

es una martingala con respecto a $\mathcal{F}_n$, en donde $\mathrm{P}h(x)$ es como en la definición antes del Ejercicio 15.14.

16 Procesos a tiempo continuo: introducción

Contenido: Procesos de conteo, de renovación, con incrementos independientes, con incrementos estacionarios, de Poisson, de Wiener (o movimiento browniano), de Lévy, de segundo orden, gaussianos.

En la sección 12 mencionamos que un PE cuyo conjunto T de índices es un *intervalo* se dice que es un PE a *tiempo continuo*. En lo que resta de estas notas estudiaremos PEs de este tipo y, a menos que se diga lo contrario, supondremos que T es el intervalo $[0, \infty)$.

Ejemplo 16.1. Ejemplo: procesos de conteo. Sea $N(\bullet) = \{N(t), t \geq 0\}$ un PE en el que $N(t)$ representa el número de veces que ocurre un cierto evento, digamos E^*, en el intervalo $[0, t]$ para $t > 0$, con $N(0) := 0$. En este caso se dice que $N(\cdot)$ es un **proceso de conteo** y es evidente que su espacio de estados es el conjunto $S = \{0, 1, \ldots\}$ y que tiene trayectorias *no decrecientes*, es decir,

$$N(s) \leq N(t) \quad \text{si} \quad 0 \leq s < t.$$

Sea $\tau_0 := 0$, y para $n \geq 1$ sea τ_n el tiempo en el que E^* ocurre por n–ésima vez, i.e. $\tau_n := \inf\{t \geq 0 | N(t) = n\}$. A τ_n se le llama el **tiempo del n–ésimo salto** y las vv.aa. $\Delta_n := \tau_n - \tau_{n-1}$, para $n = 1, 2, \ldots$, son los **tiempos entre saltos**. Nótese que

$$\tau_n = \tau_{n-1} + \Delta_n = \Delta_1 + \cdots + \Delta_n \tag{16.1}$$

Si $\Delta_1, \Delta_2, \ldots$ son i.i.d., se dice que $N(\cdot)$ es un **proceso de renovación**.

En el párrafo anterior, primero introdujimos el proceso de conteo $N(\cdot)$ y después definimos los tiempos de salto τ_n. Pero también puede ocurrir lo contrario, primero se dan los tiempos de salto $\tau_1, \tau_2, \cdots$ con $\tau_0 := 0$, y después definimos el proceso de conteo como

$$N(t) := \max\{n | \tau_n \leq t\} \quad \text{para} \quad t \geq 0. \tag{16.2}$$

Estas dos formulaciones son equivalentes. $\square$

Un ejemplo muy importante de proceso de conteo es el proceso de Poisson. Para introducir dicho proceso primero definiremos los siguientes conceptos.

Definición 16.2. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un PE. Decimos que $\mathrm{X}(\cdot)$ tiene:

(a) **incrementos independientes** si para cualquier entero $n \geq 1$ y cualquier colección de índices $0 \leq t_0 < t_1 < \cdots < t_n$ se cumple que las vv.aa. (o "incrementos")

$$\mathrm{X}(t_1) - \mathrm{X}(t_0),\ \mathrm{X}(t_2) - \mathrm{X}(t_1), \ldots, \mathrm{X}(t_n) - \mathrm{X}(t_{n-1})$$

son independientes;

(b) **incrementos estacionarios** si para cualquier $t \geq 0$ y $h > 0$ la distribución del incremento $\mathrm{X}(t+h) - \mathrm{X}(t)$ depende sólo de h, es decir,

$$\mathrm{X}(t+h) - \mathrm{X}(t) \sim \mathrm{X}(s+h) - \mathrm{X}(s) \quad \forall\, s, t \geq 0.$$

(Recuérdese que si Y y Z son vv.aa., entonces la notación $Y \sim Z$ significa que Y y Z tienen la misma distribución.)

Definición 16.3. Sea $N(\cdot) = \{N(t), t \geq 0\}$ un proceso de conteo y λ un número positivo. Decimos que $N(\cdot)$ es un **proceso de Poisson** con parámetro (o intensidad media) λ si

(a) $N(\cdot)$ tiene incrementos independientes, y

(b) $N(\cdot)$ tiene incrementos estacionarios tales que

$$N(t+h) - N(t) \sim \mathrm{Poi}(\lambda h) \quad \forall\, t \geq 0,\ h > 0. \tag{16.3}$$

La condición (3) significa que $N(t+h) - N(t)$ tiene distribución de Poisson con parámetro λh, de modo que

$$\mathrm{P}[N(t+h) - N(t) = k] = e^{-\lambda h}(\lambda h)^k / k! \quad \forall\, k = 0, 1, \ldots$$

Por lo tanto

$$E[N(t+h) - N(t)] = \lambda h, \quad \text{y} \quad \mathrm{Var}\,[N(t+h) - N(t)] = \lambda h.$$

En particular, de la primera de estas ecuaciones vemos que

$$\lambda = \frac{1}{h} E[N(t+h) - N(t)] \quad \forall\, t \geq 0,\ h > 0,$$

a lo cual se debe el nombre de "intensidad media" (o "tasa media") del parámetro λ. Es decir, si E^* es el evento asociado al proceso $N(\cdot)$ (como en el ejemplo 16.1), entonces λ se puede interpretar como el número promedio de veces que E^* ocurre por unidad de tiempo.

Por otra parte, como $N(0) := 0$, podemos escribir $N(t) = N(t) - N(0)$ y se sigue de (3) que $N(t)$ tiene distribución de Poisson $\mathrm{Poi}(\lambda t)$ para todo $t > 0$.

En el ejercicio 16.26 se dá una caracterización del proceso de Poisson. También veremos que dicho proceso es de renovación y, además, es un proceso markoviano de saltos.

Definición 16.4. Se dice que $W(\cdot) = \{W(t), t \geq 0\}$ es un **proceso de Wiener** — también llamado **movimiento browniano** — si satisface que

(a) $W(0) := 0$,

(b) tiene incrementos independientes, y

(c) tiene incrementos estacionarios con

$$W(t+h) - W(t) \sim N(0, \sigma^2 h) \quad \forall\, t \geq 0,\ h > 0, \tag{16.4}$$

en donde σ es una constante positiva. Si $\sigma^2 = 1$, se dice que $W(\cdot)$ es un proceso de Wiener **estándar.**

En el ejercicio 16.4 se da una caracterización del proceso de Wiener. Como $W(t) = W(t) - W(0)$, se sigue de (4) que

$$W(t) \sim N(0, \sigma^2 t) \quad \forall\, t > 0. \tag{16.5}$$

En el ejercicio 16.5 se puede demostrar que el proceso de Wiener $W(\cdot)$ tiene trayectorias continuas; sin embargo, las trayectorias *no* son diferenciables. Por otra parte, de la propiedad 16.4(b) se puede deducir que $W(\cdot)$ es una martingala (Definición 15.1) y un proceso de Markov (vea la sección 12, ecuación (16.3)). Esto es consecuencia de la siguiente observación y de la Proposición 16.6

Observación 16.5. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un PE a tiempo continuo con espacio de estados $(S, \mathcal{S})$. Por la ecuación (16.3) de la sección 12, $\mathrm{X}(\cdot)$ es de Markov si

$$P[\mathrm{X}(t) \in B | X(r)\, \forall\, 0 \leq r \leq s] = P[\mathrm{X}(t) \in B \,|\, \mathrm{X}(s)] \tag{16.6}$$

para todo $B \in \mathcal{S}$ y $0 \leq s \leq t$. Resulta que esta condición es equivalente a la siguiente: para cualquier entero $n \geq 1$ y cualquier colección de índices $0 \leq t_1 < \cdots < t_n < t_{n+1}$

$$P[\mathrm{X}(t_{n+1}) \in B|\mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)] = \mathrm{P}[\mathrm{X}(t_{n+1}) \in B|\ \mathrm{X}(t_n)]. \tag{16.7}$$

Es decir, tomando $t_n = s$ y $t_{n+1} = t$, en lugar de verificar (16.6) para la colección $\{\mathrm{X}(r),\ 0 \leq r \leq s\}$ basta considerar colecciones *finitas* $\{\mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)\}$. Análogamente, supóngase que $\mathrm{X}(\cdot)$ es una martingala con respecto a $\mathcal{F}_s := \sigma\{\mathrm{X}(r),\ 0 \leq r \leq s\}$; vea la Definición 15.1. En particular, $\mathrm{X}(t)$ está en L_1 y

$$E[\mathrm{X}(t)|\mathcal{F}_s] = \mathrm{X}(s) \quad \forall\, 0 \leq s \leq t.$$

Esta condición es equivalente a la siguiente: para cualquier entero $n \geq 1$ y cualquier colección de índices $0 \leq t_1 < \cdots < t_n < t_{n+1}$

$$E[\mathrm{X}(t_{n+1})|\mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)] = \mathrm{X}(t_n). \tag{16.8}$$

(La demostración de estos hechos se puede ver, por ejemplo, en el capítulo 4 del libro de Ash y Gardner.)

Proposición 16.6. Si $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\ t \geq 0\}$ es un PE con incrementos independientes, entonces

(a) $\mathrm{X}(\cdot)$ es un proceso de Markov.

(b) Si además $\mathrm{X}(t) \in L_1$ para todo $t \geq 0$, entonces el "proceso centrado" $\overline{\mathrm{X}}(t) := \mathrm{X}(t) - E\mathrm{X}(t)$ es una martingala.

Por lo tanto:

(c) el proceso de Wiener es de Markov y también una martingala;

(d) si $N(\cdot)$ es un proceso de Poisson con parámetro λ, entonces $N(\cdot)$ es de Markov y el proceso centrado $\overline{N}(t) := N(t) - \lambda t$ es una martingala.

Demostración. Como $\mathrm{X}(\cdot)$ tiene incrementos independientes, por el ejercicio 16.2 podemos suponer, sin pérdida de generalidad, que

$$\mathrm{X}(0) = 0. \tag{16.9}$$

Ahora, para demostrar (a), sea $n \geq 1$ un entero arbitrario, y considérense los índices $0 = t_0 < t_1 < \cdots < t_n < t_{n+1}$ y las vv.aa.

$$Y_k := \mathrm{X}(t_{k+1}) - \mathrm{X}(t_k) \quad \text{para} \quad k = 0, \ldots, n, \tag{16.10}$$

las cuales son independientes (por la independencia de los incrementos de $\mathrm{X}(\cdot)$). Esto significa que (por la Proposición 12.9) la sucesión

$$Z_{k+1} := \sum_{j=0}^{k} Y_j \quad \text{para} \quad k = 0, \ldots, n$$

es una CM; en particular,

$$\mathrm{P}(Z_{k+1} \in B | Z_0, \ldots, Z_k) = \mathrm{P}(Z_{k+1} \in B | Z_k) \tag{16.11}$$

Sin embargo, por (9) y (10), $Z_{k+1} = X(t_{k+1})$ para todo $k = 0, \ldots, n$, así que de hecho la condición (11) es la misma que (7). Esto demuestra (a).

Para demostrar (b), de nuevo supondremos (9). Además, es inmediato que el proceso centrado $\overline{\mathrm{X}}(t) := \mathrm{X}(t) - E\mathrm{X}(t)$ tiene incrementos independientes y media $E\overline{\mathrm{X}}(t) = 0$. Por lo tanto, sin pérdida de generalidad, podemos suponer que $E\mathrm{X}(t) = 0$ para todo $t \geq 0$. Ahora consideremos de nuevo las vv.aa. independientes Y_k en (10). En particular, Y_n es independiente de $Y_0, \ldots, Y_{n-1}$ y , por lo tanto, Y_n es independiente de $\mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)$. Luego, escribiendo $\mathrm{X}(t_{n+1}) = \mathrm{X}(t_n) + Y_n$ vemos que

$$\begin{aligned} E[\mathrm{X}(t_{n+1}) | \mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)] &= \mathrm{X}(t_n) + E[Y_n | \mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)] \\ &= \mathrm{X}(t_n) + E[Y_n] \quad \text{(por independencia)} \\ &= \mathrm{X}(t_n). \end{aligned}$$

Esto comprueba la condición (8), de modo que $\mathrm{X}(\cdot)$ es una martingala.

Por último, los incisos (c) y (d) se siguen de las Definiciones 16.4 y 16.3, respectivamente. □

Otras definiciones básicas son las siguientes.

Definición 16.7. Decimos que un PE $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ es un

(a) **proceso de segundo orden** (notación: $\mathrm{X}(\cdot) \in L_2$) si $\mathrm{X}(t) \in L_2$ para todo $t \geq 0$, es decir, $E(|\mathrm{X}(t)|^2) < \infty$;

(b) **proceso gaussiano** si las combinaciones lineales

$$a_1\mathrm{X}(t_1) + \cdots + a_n\mathrm{X}(t_n) \tag{16.12}$$

son vv.aa. gaussianas para cualquier colección finita de índices $0 \leq t_1 < \cdots < t_n$ y cualquier colección de números reales $a_1, \ldots, a_n$. (Nótese que en este caso, $\mathrm{X}(\cdot)$ es un proceso de segundo orden y $\mathrm{X}(t)$ es una v.a. gaussiana para todo $t \geq 0$.)

En el ejercicio 16.4 se pide demostrar, en particular, que un proceso de Wiener es gaussiano.

Si $\mathrm{X}(\cdot) \in L_2$ definimos su **función de covarianza** K_{X} como

$$K_{\mathrm{X}}(s,t) := \mathrm{Cov}(\mathrm{X}(s), \mathrm{X}(t)) \quad \forall\, s,t \geq 0,$$

es decir,

$$\begin{aligned} K_{\mathrm{X}}(s,t) &= E[(\mathrm{X}(s) - E\mathrm{X}(s)) \cdot (\mathrm{X}(t) - E\mathrm{X}(t))] \\ &= E[\mathrm{X}(s)\mathrm{X}(t)] - E\mathrm{X}(s) \cdot E\mathrm{X}(t) \end{aligned} \tag{16.13}$$

En particular, con $s = t$ obtenemos la variancia de $\mathrm{X}(t)$:

$$K_{\mathrm{X}}(t,t) = \mathrm{Var}[\mathrm{X}(t)].$$

Ejemplo 16.8. (a) Si $N(\cdot) = \{N(t), t \geq 0\}$ es un proceso de Poisson con parámetro λ, entonces $K_N(s,t) = \lambda \cdot \min(s,t)$.

(b) Si $W(\cdot) = \{W(t), t \geq 0\}$ es un proceso de Wiener con parámetro σ^2, entonces $K_W(s,t) = \sigma^2 \cdot \min(s,t)$.

Demostración. Las demostraciones de (a) y (b) son muy similares. Por consiguiente, probaremos (a) y la demostración de (b) se deja como ejercicio.

Sea $N(\cdot)$ un proceso de Poisson con parámetro λ. Entonces

$$EN(s) \cdot EN(t) = (\lambda s) \cdot (\lambda t) = \lambda^2 st. \tag{16.14}$$

Por otra parte, para $0 \leq s < t$, podemos escribir

$$N(t) = [N(t) - N(s)] + N(s)$$

y los incrementos $N(s) = N(s) - N(0)$ y $N(t) - N(s)$ son independientes. Luego

$$\begin{aligned} E[N(s)N(t)] &= E[N(s)\cdot(N(t)-N(s))] + E[N(s)^2] \\ &= EN(s)\cdot E[N(t)-N(s)] + (\lambda s)^2 + \lambda s \\ &= \lambda s \cdot \lambda(t-s) + \lambda^2 s^2 + \lambda s \\ &= \lambda^2 st + \lambda s. \end{aligned}$$

De esta relación junto con (14) y (13) obtenemos

$$K_N(s,t) = \lambda s \quad \text{si} \quad 0 \leq s < t.$$

En forma análoga se puede ver que $K_N(s,t) = \lambda t$ si $0 \leq t < s$. Combinando estos resultados se sigue de $K_N(s,t) = \lambda \cdot \min(s,t)$ para todo $s,t \geq 0$. □

Continuidad de PEs. A continuación introducimos dos tipos de continuidad de PEs; en una sección posterior introduciremos un tercer tipo — continuidad "en la media cuadrática".

Definición 16.9. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un PE definido sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$. Se dice que $\mathrm{X}(\cdot)$ es

(a) **continuo casi seguramente** (c.s) o **continuo con probabilidad 1** (c.p.1) si sus trayectorias $t \longmapsto \mathrm{X}(t,\omega)$ son continuas (como funciones de $t \geq 0$) para casi todo $\omega \in \Omega$;

(b) **estocásticamente continuo** si para cualquier $t_0 \geq 0$ y cualquier $\epsilon > 0$

$$\lim_{t \to t_0} \mathrm{P}\{|\mathrm{X}(t) - \mathrm{X}(t_0)| > \epsilon\} = 0.$$

En la definición anterior es evidente que (a) implica a (b).

Por ejemplo, sea $N(\cdot) = \{N(t), t \geq 0\}$ un proceso de Poisson con parámetro λ. Como $N(\cdot)$ es un proceso de "saltos", es evidente que *no* es continuo c.s. Sin embargo, $N(\cdot)$ es *estocásticamente continuo* porque, digamos para $0 < \epsilon < 1$,

$$\begin{aligned} \mathrm{P}\{|N(t) - N(t_0)| > \epsilon\} &= \mathrm{P}\{|N(t) - N(t_0)| \geq 1\} \\ &= 1 - e^{-\lambda|t-t_0|} \\ &\to 0 \quad \text{cuando} \quad t \to t_0. \end{aligned}$$

Para verificar continuidad c.p.1, un criterio muy útil es el siguiente — cuya demostración se puede ver prácticamente en cualquier libro avanzado sobre PEs (por ejemplo, Gikhman y Skorokhod (1969), página 170).

Teorema 16.10. *(Criterio de continuidad de Kolmogorov) Si existen constantes positivas* α*,* β *y* γ *tales que*

$$E[|\mathrm{X}(t) - \mathrm{X}(s)|^{\alpha}] \leq \beta |t - s|^{1+\gamma} \quad \forall\, t, s \geq 0, \tag{16.15}$$

entonces $\mathrm{X}(\cdot)$ *es un PE continuo c.s.*

En el Ejercicio 16.5 se pide verificar que para un proceso de Wiener la desigualdad (15) se cumple con $\alpha = 4$ y $\gamma = 1$ y, en consecuencia, *un proceso de Wiener es continuo c.s.*.

Para concluir esta sección, caracterizaremos un proceso de Poisson como un proceso de renovación cuyos tiempos entre saltos $\Delta_n := \tau_n - \tau_{n-1}$, para $n = 1, 2, \ldots$, tienen distribución exponencial.

Teorema 16.11. *Las siguientes proposiciones son equivalentes para un PE* $N(\cdot) = \{N(t), t \geq 0\}$*.*

(a) $N(\cdot)$ *es un proceso de Poisson con parámetro* λ*.*

(b) $N(\cdot)$ *es un proceso de renovación con tiempos entre saltos* $\Delta_n \sim Exp(\lambda)$*.*

Demostración. (a) $\Rightarrow$ (b) Supóngase que (a) se cumple. Entonces cada tiempo de salto τ_n ($n = 0, 1, \cdots$, con $\tau_0 = 0$) es un *tiempo de paro* para $N(\cdot)$, es decir, la ocurrencia del evento $\{\tau_n \leq t\}$ sólo depende de $\{N(s), s \leq t\}$. Además, τ_n es un tiempo de paro finito porque

$$\tau_n \leq t \quad \text{ssi} \quad N(t) \geq n. \tag{16.16}$$

Asimismo, de (1),

$$\Delta_n > t \;\text{ ssi }\; \tau_n > t + \tau_{n-1} \;\text{ ssi }\; N(t + \tau_{n-1}) - N(\tau_{n-1}) = 0, \tag{16.17}$$

y se tiene que (explique)

$$N(t + \tau_{n-1}) - N(\tau_{n-1}) \sim N(t + s) - N(s) \sim \mathrm{Poi}(\lambda t) \tag{16.18}$$

para todo $n = 1, 2, \cdots,\ s \geq 0$ y $t \geq 0$.

Ahora, para demostrar (b) verificaremos que

(b_1) los tiempos entre saltos $\Delta_1, \Delta_2, \ldots$ son independientes, y

(b_2) $\Delta_n \sim \mathrm{Exp}(\lambda)$ para todo $n = 1, 2, \ldots$.

Demostración de (b$_1$). Sean $0 < n_1 < n_2 < \cdots < n_k$ enteros positivos, y N_j el incremento

$$N_j := N(t_j + \tau_{n_j - 1}) - N(\tau_{n_j-1}) \quad \text{para} \quad j = 1, \ldots, k,$$

con $t_1, \ldots, t_k > 0$. Luego, por (17) y la independencia de incrementos de $N(\cdot)$,

$$\begin{aligned} \mathrm{P}(\Delta_{n_1} > t_1, \ldots, \Delta_{n_k} > t_k) &= \mathrm{P}(N_1 = 0, \ldots, N_k = 0) \\ &= \mathrm{P}(N_1 = 0) \cdots \mathrm{P}(N_k = 0) \\ &= \mathrm{P}(\Delta_{n_1} > t_1) \cdots \mathrm{P}(\Delta_{n_k} > t_k), \end{aligned}$$

de nuevo por (17).

Demostración de (b$_2$). Por (17) y (18),

$$\mathrm{P}(\Delta_n > t) = \mathrm{P}[N(t + \tau_{n-1}) - N(\tau_{n-1}) = 0] = e^{-\lambda t},$$

es decir, Δ_n tiene distribución Exp(λ).

(b) $\Rightarrow$ (a). Si (b) se cumple, entonces, por la perdida de memoria de la variable exponencial, para cualquier tiempo fijo $s \geq 0$, el PE $N'(\cdot)$ definido por $N'(t) := N(t+s) - N(s)$ para $t \geq 0$ es de nuevo un proceso de renovación con tiempos entre saltos $\Delta'_n \sim \Delta_n$. Por lo tanto, para tiempos cualesquiera $0 \leq r \leq s \leq s + t$ se tiene que

$$\mathrm{P}[N(t+s) - N(s) = n | N(r) = k] = \mathrm{P}[N(t) = n] \tag{16.19}$$

lo cual implica la independencia de los incrementos $N(r) = N(r) - N(0)$ y $N(t+s) - N(s)$. Para ver que $N(\cdot)$ tiene incrementos estacionarios $N(t+s) - N(t) \sim \mathrm{Poi}(\lambda t)$ nótese primero que

$$N(t) = n \quad \text{ssi} \quad \tau_n \leq t < \tau_{n+1} \tag{16.20}$$

pero además, $\tau_n = \Delta_1 + \cdots + \Delta_n$ tiene distribución gama con parámetros n, λ^{-1}. Por lo tanto, por (20),

$$\mathrm{P}[N(t) = n] = \mathrm{P}[\tau_n \leq t] - P[\tau_{n+1} \leq t] = e^{-\lambda t}(\lambda t)^n / n! \tag{16.21}$$

para $n = 1, 2, \ldots$, mientras que (como $\Delta_1 = \tau_1$)

$$\mathrm{P}[N(t) = 0] = \mathrm{P}[\Delta_1 > t] = e^{-\lambda t}. \tag{16.22}$$

Resumiendo, de (21) y (22) tenemos que $N(t) \sim \mathrm{Poi}(\lambda t)$ para todo $t \geq 0$, así que, como en la demostración de (19), para $0 \leq s \leq t$ y $n = 0, 1, \ldots$ se sigue que

$$\begin{aligned} \mathrm{P}[N(t) - N(s) = n] &= \mathrm{P}[N(t-s) = n] \\ &= e^{-\lambda(t-s)}[\lambda(t-s)]^n/n! \end{aligned}$$

Esto completa la demostración de que $N(\cdot)$ es un proceso de Poisson. □

Concluiremos esta sección con la definición de una clase muy importante de PEs.

Definición 16.12. Se dice que $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ es un **proceso de Lévy** si:

(a) $\mathrm{X}(0) = 0$ c.s.

(b) $\mathrm{X}(\cdot)$ tiene incrementos independientes y estacionarios, y

(c) $\mathrm{X}(\cdot)$ es estocásticamente continuo.

Los ejemplos más conocidos de procesos de Lévy son el proceso de Wiener y el proceso de Poisson, pero hay muchos más — vea los libros de Applebaum (2004), Bertoin (1996), Sato (1999), etc.

Notas § 16

Sean S y T espacios medibles arbitrarios y $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad.

Definición 16.13. Se dice que dos PEs $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \in T\}$ y $Y(\cdot) = \{Y(t), t \in T\}$ definidos sobre $(\Omega, \mathcal{F}, \mathrm{P})$ y con valores en S son **equivalentes** (o que $Y(\cdot)$ es una **versión** de $\mathrm{X}(\cdot)$) si

$$\mathrm{P}[\mathrm{X}(t) \neq Y(t)] = 0 \quad \forall\, t \in T.$$

En el resto de esta sección supondremos que S y T son *espacios métricos* con sus respectivas σ–álgebras de Borel $\mathcal{B}(S)$ y $\mathcal{B}(T)$, y $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \in T\}$ es un PE con espacio de estados S.

Definición 16.14. Sea S un espacio topológico. Se dice que $\mathrm{X}(\cdot)$ es **separable** si existe un subconjunto $T_0 \subset T$ denso y numerable tal que para todo $F \subset S$ cerrado, y $I \subset T$ abierto, la diferencia simétrica entre

$$\{X(t) \in F : t \in I \bigcap T_0\} \text{ y } \{X(t) \in F : t \in I\}$$

tiene medida cero.

En palabras, $\mathrm{X}(\cdot)$ es separable si, casi seguramente, el comportamiento de las trayectorias sobre T es determinado por su comportamiento en T_0. A T_0 se le llama **conjunto separante**.

Teorema 16.15. *[Gikhman y Skorokhod (1969), p.153, Teorema 2.] Supóngase que S es un espacio métrico separable y localmente compacto, y que T es un espacio métrico separable. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \in T\}$ un PE cualquiera con valores en S. Entonces existe un proceso separable $Y(\cdot) = \{Y(t), t \in T\}$ equivalente a $\mathrm{X}(\cdot)$.*

En el contexto del Teorema 16.15 la pregunta obvia es, ¿cómo se selecciona el conjunto separante T_0? Sean S, T y $\mathrm{X}(\cdot)$ como en dicho teorema. Si $\mathrm{X}(\cdot)$ es separable y estocásticamente continuo, entonces *cualquier* conjunto $T_0 \subset T$ numerable y denso en T se puede usar como conjunto separante. (La demostración de este hecho se puede ver en Gikhman y Skorokhod (1969), p.155, Teorema 5. El caso especial con $S = \mathbb{R}$ se puede ver en Ash y Gardner (1975), p.163, Teorema 4.1.6.)

En lo que sigue supondremos que $T = [0, \infty)$.

Definición 16.16. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un PE con valores en un espacio métrico completo S. Se dice que $\mathrm{X}(\cdot)$ **no tiene discontinuidades de segundo orden** si los límites laterales de $\mathrm{X}(s)$ cuando $s \to t+$ ó $s \to t-$ existen para todo $t > 0$, y también en $0+$.

En el libro de Gikhman y Skorokhod (1969), pp.160–168, se pueden ver condiciones necesarias y/o suficientes para que un PE no tenga discontinuidades de segundo orden. En particular, se tienen los siguientes dos resultados.

Proposición 16.17. Un proceso con incrementos independientes que es separable y estocásticamente continuo no tiene discontinuidades de segundo orden.

Proposición 16.18. Si un PE $\mathrm{X}(\cdot)$ es estocásticamente continuo y no tiene discontinuidades de segundo orden, entonces tiene una versión $Y(\cdot)$ cuyas trayectorias, casi seguramente, son continuas por la derecha.

Ejercicios § 16

16.1. Demuestre que un proceso de Poisson $N(\cdot)$ tiene trayectorias no decrecientes, es decir,

$$\mathrm{P}[N(t) \geq N(s)] = 1 \quad \forall\, t > s \geq 0.$$

16.2. Demuestre:

(a) Si $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ es un PE con incrementos independientes, entonces el PE $Y(\cdot)$ definido por $Y(t) := \mathrm{X}(t) - \mathrm{X}(0)$, para $t \geq 0$, tiene incrementos independientes y $Y(0) = 0$.

(b) Recíprocamente, sea $Y(\cdot) = \{Y(t), t \geq 0\}$ un PE con incrementos independientes y $Y(0) = 0$, y sea $X(0)$ una v.a. independiente de $Y(\cdot)$. Entonces el PE $X(\cdot)$ definido por $\mathrm{X}(t) := Y(t) + \mathrm{X}(0)$ para todo $t \geq 0$ tiene incrementos independientes y estado inicial $\mathrm{X}(0)$.

16.3. Demuestre el inciso (b) del Ejemplo 16.8.

16.4. Demuestre que las siguientes proposiciones (a) y (b) son equivalentes:

(a) $W(\cdot) = \{W(t), t \geq 0\}$ es un proceso de Wiener con parámetro σ^2.

(b) $W(\cdot) = \{W(t), t \geq 0\}$ es un proceso gaussiano con media y función de covarianza $EW(t) = 0$ y $K_W(s,t) = \sigma^2 \cdot \min(s,t)$ para todo $s, t \geq 0$.

16.5. Demuestre que si $W(\cdot)$ es un proceso de Wiener, existe una constante β tal que

$$E[|W(t) - W(s)|^4] \leq \beta |t-s|^2 \quad \forall\, s,t \geq 0.$$

De la Proposición 16.10 concluya que $W(\cdot)$ es un PE continuo c.s.

16.6. Sea λ una constante, y sea Y_1, Y_2 vv.aa. i.i.d. con distribución normal $N(0, \sigma^2)$. Demuestre que el PE $\mathrm{X}(\cdot) = \{\mathrm{X}(t), -\infty < t < \infty\}$ definido como

$$\mathrm{X}(t) := Y_1 \cos\lambda t + Y_2 \operatorname{sen}\lambda t$$

es un proceso gaussiano y calcule la media $E\mathrm{X}(t)$ y la función de covarianza de $\mathrm{X}(\cdot)$.

16.7. Sea $N(\cdot)$ un proceso de Poisson, y sea X_0 una v.a. independiente de $N(\cdot)$ y con distribución $\mathrm{P}(\mathrm{X}_0 = 1) = \mathrm{P}(\mathrm{X}_0 = -1) = 1/2$. En este caso el PE $\mathrm{X}(\cdot)$ definido por

$$\mathrm{X}(t) := \mathrm{X}_0 \cdot (-1)^{N(t)} \quad \forall\, t \geq 0$$

se dice que es una **señal telegráfica aleatoria**. Calcule la media y la función de covarianza de $\mathrm{X}(\cdot)$.

16.8. Calcule $E[N(s)N(s+t)]$ para $s, t \geq 0$, donde $N(\cdot)$ es un proceso de Poisson con parámetro λ.

16.9. Demuestre que si $N_1(\cdot)$ y $N_2(\cdot)$ son procesos de Poisson independientes con parámetros λ_1 y λ_2, entonces su suma $N_1(t) + N_2(t), t \geq 0$, es un proceso de Poisson con parámetro $\lambda_1 + \lambda_2$. [*Observación:* la diferencia $N_1(\cdot) - N_2(\cdot)$ *no* es un proceso de Poisson.]

16.10. Sea $N(\cdot)$ un proceso de Poisson con intensidad media λ, y sea T una v.a. positiva independiente de $N(\cdot)$. Calcule $EN(T)$ y $\mathrm{Var}[N(T)]$ en cada uno de los siguientes casos en los que T tiene distribución:

(a) exponencial con parámetro μ,

(b) uniforme en el intervalo $[a, b]$, con $0 < a < b$,

(c) continua arbitraria con densidad $f(t)$.

16.11. Sea $N(\cdot)$ un proceso de Poisson y sean $0 < s < t$, Demuestre que la distribución condicional de $N(s)$, dado que $N(t) = n$, es la distribución binomial $\mathrm{Bin}(n, p)$ con $p = s/t$, es decir,

$$\mathrm{P}[N(s) = k | N(t) = n] = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para} \quad k = 0, 1, \ldots, n.$$

16.12. Proceso de Poisson muestrado, homogéneo. Sea $N(\cdot)$ un proceso de Poisson con parámetro λ, y supóngase que $N(\cdot)$ es el proceso de conteo asociado a un cierto evento E^*. Cada ocurrencia de E^* puede ser registrada con probabilidad p, donde $0 < p < 1$, y no registrada con probabilidad $1 - p$. Suponga, además, que el registro de una ocurrencia de E^* es independiente de otras ocurrencias, y también independiente de $N(\cdot)$. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ el PE en el que $\mathrm{X}(t)$ es el número de veces que se registra E^* en el intervalo $[0, t]$. Demuestre que $\mathrm{X}(\cdot)$ es un proceso de Poisson con parámetro λp.

16.13. El número de personas que pasan frente a un restaurante se comporta como un proceso de Poisson con parámetro $\lambda = 1000$ personas por hora. Suponga que cada persona tiene probabilidad $p = 0.01$ de entrar al restaurante, y sea Y el número de personas que entran en un periodo de 10 minutos. Calcule EY, $\mathrm{Var}(Y)$ y $\mathrm{P}(Y \geq 2)$. (*Sugerencia*: use el ejercicio 12.)

16.14. El puente de Poisson. Sea $N(\cdot)$ un proceso de Poisson con parámetro λ, y sea $N_p(t) := N(t) - tN(1)$, para $0 \leq t \leq 1$, el llamado "puente de Poisson". Calcule $EN_p(t)$ y la función de covarianza de $N_p(\cdot)$.

16.15. El puente browniano. Repita el Ejercicio anterior pero para el "puente browniano" $W_p(t) := W(t) - tW(1)$, con $0 \leq t \leq 1$, en donde $W(\cdot)$ es un proceso de Wiener o movimiento browniano.

16.16. Suponga que un cierto dispositivo deja de funcionar cuando se acumula el efecto de k choques. Calcule la densidad del tiempo de vida T del dispositivo si los choques ocurren de acuerdo a un proceso de Poisson con intensidad media λ.

16.17. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un PE que tiene incrementos independientes y función media $m(t) := E\mathrm{X}(t)$ finita para todo $t \geq 0$. Demuestre que si $0 \leq t_1 < \cdots < t_n < t_{n+1}$, entonces

$$E[\mathrm{X}(t_{n+1})|\mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)] = \mathrm{X}(t_n) + m(t_{n+1}) - m(t_n).$$

Nota. En los Ejercicios 16.18 al 16.22, $W(\cdot) = \{W(t), t \geq 0\}$ es un proceso de Wiener con parámetro $\sigma^2 > 0$.

16.18. Demuestre que $E[(W(s)-W(a))\cdot(W(t)-W(a))] = \sigma^2 \cdot \min(s-a, t-a)$ para todo $s, t \geq a \geq 0$.

16.19. Demuestre que $W(1) + \cdots + W(n)$ tiene distribución $N(0, s_n)$ en donde

$$s_n := \sigma^2 \; n(n+1)(2n+1)/6.$$

[*Sugerencia:* use la independencia de incrementos de $W(\cdot)$ y la fórmula

$$\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6.]$$

16.20. Demuestre que cada uno de los siguientes PEs es de Wiener:

(a) $W_1(t) := a^{-1/2}W(at)$ para $t \geq 0$ (donde $a > 0$ es una constante), y

(b) $W_2(t) := tW(1/t)$ para $t > 0$ y $W_2(0) := 0$.

16.21. Calcule $E\mathrm{X}(t)$ y $\mathrm{Cov}(\mathrm{X}(s), \mathrm{X}(t))$ para cada uno de los siguientes procesos:

(a) $\mathrm{X}(t) := W^2(t)$ para $t \geq 0$;

(b) $\mathrm{X}(t) := e^{-at}W(e^{2at})$ para $t \geq 0$ (donde $a > 0$ es una constante);

(c) $\mathrm{X}(t) := \cos\,(W(t))$ para $t \geq 0$;

(d) $\mathrm{X}(t) := W(t+a) - W(a)$ para $t \geq 0$ (donde $a \geq 0$ es una constante);

(e) $\mathrm{X}(t) := At + W(t)$ para $t \geq 0$, donde A es una v.a. independiente de $W(\cdot)$ y tiene distribición $N(m, \sigma^2)$.

16.22. Demuestre que el PE $\mathrm{X}(t) := W^2(t) - \sigma^2 t$, $t \geq 0$, es una martingala.

Observación 16.19. Se dice que una función f es $\mathrm{o}(t)$ cuando $t \to 0$, en cuyo caso escribimos $f = \mathrm{o}(t)$, si

$$\lim_{t \to 0} \frac{f(t)}{t} = 0.$$

(En el Ejercicio 16.23 se mencionan algunos ejemplos de funciones $\mathrm{o}(t)$.)

Observación 16.20. Sea a una constante. La solución de la ecuación diferencial lineal

$$x'(t) = ax(t) + b(t) \quad \text{para} \quad t \geq 0, \quad \text{con } x(0) = x_0,$$

es la función:

$$x(t) = x_0 e^{at} + \int_0^t b(s) e^{a(t-s)}\, ds \quad \forall\, t \geq 0.$$

16.23. Demuestre: cuando $t \to 0$,

(a) Si $f(t) = t^r$, entonces $f = \mathrm{o}(t)$ si $r > 1$, pero $f \neq \mathrm{o}(t)$ si $0 \leq r \leq 1$.

(b) Si $f(t) = e^{\lambda t}$ para alguna constante λ, entonces $f(t) = 1 + \lambda t + \mathrm{o}(t)$.

(c) Una combinación lineal de funciones $\mathrm{o}(t)$ es $\mathrm{o}(t)$; es decir, si $a_1, \ldots, a_n$ son constantes y $f_1, \ldots, f_n$ son funciones $\mathrm{o}(t)$, entonces $a_1 f_1 + \cdots a_n f_n = \mathrm{o}(t)$.

16.24. Demuestre que si $N(\cdot)$ es un proceso de Poisson, entonces

(a) $\mathrm{P}[N(t) = 1] = \lambda t + \mathrm{o}(t)$, y

(b) $\mathrm{P}[N(t) \geq 2] = \mathrm{o}(t)$.

Por lo tanto, $\mathrm{P}[N(t) = 0] = 1 - \mathrm{P}[N(t) \geq 1] = 1 - \lambda t + \mathrm{o}(t)$.

16.25. Sea $N(t) = \{N(t), t \geq 0\}$ un proceso de conteo y sea

$$p_n(t) := \mathrm{P}[N(t) = n] \quad \forall\, t \geq 0,\ n = 0, 1 \ldots.$$

Sea $\lambda > 0$ una constante y suponga que, para $t \geq 0$, se satisfacen las ecuaciones diferenciales

(a) $p_0'(t) = -\lambda p_0(t)$

(b) $p_n'(t) = \lambda p_n(t) + \lambda p_{n-1}(t)$ para $n = 1, 2, \ldots$

con condiciones iniciales $p_0(0) = 1$ y $p_n(0) = 0$ para $n = 1, 2 \ldots$. Demuestre que para $t > 0$, $N(t)$ tiene distribución de Poisson con parámetro λt, i.e.

$$p_n(t) = e^{-\lambda t}(\lambda t)^n / n! \quad \forall\, t > 0,\ n = 0, 1, \ldots.$$

[*Sugerencia:* para resolver (a) use la Observación 16.20 que aparece antes del Ejercicio 16.23; después resuelva (b) por inducción.]

16.26. Demuestre que las siguientes proposiciones (a) y (b) son equivalentes para un PE $N(\cdot) = \{N(t), t \geq 0\}$:

(a) $N(\cdot)$ es un proceso de Poisson con parámetro λ.

(b) $N(\cdot)$ es un proceso de conteo tal que

- (b_1) $N(\cdot)$ tiene incrementos independientes y estacionarios;
- (b_2) $\mathrm{P}[N(t) = 1] = \lambda t + \mathrm{o}(t)$,
- (b_3) $\mathrm{P}[N(t) \geq 2] = \mathrm{o}(t)$.

[*Sugerencia:* use los Ejercicios 16.24 y 16.25. En particular, para demostrar que (b) $\Rightarrow$ (a) pruebe que las condiciones en (b) implican que se satisfacen las ecuaciones diferenciales en el Ejercicio 16.25.]

Nota 16.21. En los Ejercicios 16.27 al 16.29 $N(\cdot) = \{N(t), t \geq 0\}$ es un proceso de Poisson con parámetro λ.

16.27. Demuestre que para cualquier $s \geq 0$ fijo, el PE $N'(\cdot) = \{N'(t), t \geq 0\}$ con

$$N'(t) := N(t+s) - N(s) \quad \forall\, t \geq 0$$

es de Poisson con parámetro λ, y que $N'(\cdot)$ es independiente de $\{N(r), r \leq s\}$.

Observación 16.22. Este resultado sigue valiendo si s se sustituye por un *tiempo de paro* $T < \infty$ para $N(\cdot)$. [Una v.a. $T \geq 0$ es un **tiempo de paro** para un PE $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ si, para cualquier $s \geq 0$, la ocurrencia del evento $\{T \leq s\}$ depende sólo de $\mathrm{X}(r)$ para $r \leq s$.]

16.28. Proceso de Poisson compuesto. Sean $\mathrm{X}_1, \mathrm{X}_2, \ldots$ vv.aa. i.i.d. e independientes de $N(\cdot)$. El proceso $Y(\cdot) = \{Y(t), t \geq 0\}$ definido por:

$$Y(t) := \sum_{j=1}^{N(t)} \mathrm{X}_j, \quad \text{con} \quad Y(t) := 0 \quad \text{si} \quad N(t) = 0,$$

se llama “proceso de Poisson compuesto”. Demuestre que para todo $t \geq 0$,

(a) Si $\mathrm{X}_1 \in L_1$, entonces $EY(t) = E\mathrm{X}_1 \cdot EN(t) = \lambda t\, E\mathrm{X}_1$.

(b) Si $\mathrm{X}_1 \in L_2$, entonces $\mathrm{Var}[Y(t)] = EN(t) \cdot \mathrm{Var}(\mathrm{X}_1) + (E\mathrm{X}_1)^2 \cdot \mathrm{Var}[N(t)]$, i.e.

$$\mathrm{Var}[Y(t)] = \lambda t \cdot E(\mathrm{X}_1^2).$$

(c) Si $\phi(s)$ es la función característica de X_1, entonces la función característica de $Y(t)$ es

$$C_{Y(t)}(s) := E[e^{isY(t)}] = \exp[\lambda t(\phi(s) - 1)] \quad \forall\, s \in \mathbb{R}$$

16.29. Supóngase que $N(\cdot)$ registra el número de fallecimientos de los clientes de una compañia de seguros; es decir, $N(t)$ es el número de clientes fallecidos en el periodo $[0, t]$. Cuando un cliente fallece, la aseguradora le paga al beneficiario de la póliza la cantidad X. Supóngase que dichos pagos, digamos $\mathrm{X}_1, \mathrm{X}_2, \ldots$, son vv.aa. i.i.d. con $\mathrm{X}_n \sim \mathrm{X}$. Si además tal sucesión de pagos es independiente de $N(\cdot)$, entonces

$$Y(t) := \sum_{j=1}^{N(t)} \mathrm{X}_j$$

representa la cantidad total de pagos que ha hecho la aseguradora en el periodo $[0, t]$. Si X tiene distribución uniforme en el intervalo $[a, b]$, con $0 \leq a < b$, entonces, para cada $t \geq 0$, calcule el total esperado de pagos $EY(t)$ y la varianza de $Y(t)$.

16.30. Sea $Y(\cdot)$ el PE del ejercicio 16.29 en donde X está distribuida exponencialmente con media \$100,000 pesos. Calcule la media, la varianza y la función característica de $Y(t)$, y la probabilidad $\mathrm{P}[Y(t) = 0]$.

16.31. Demuestre que la distribución exponencial "no tiene memoria"; es decir, si $Y \sim \mathrm{Exp}(\lambda)$ entonces

$$\mathrm{P}(Y > s + t | Y > s) = \mathrm{P}(Y > t) \quad \forall\, s, t \geq 0. \qquad (*)$$

Observación 16.23. De hecho, se puede demostrar que la distribución exponencial es la *única* con la propiedad (*). Esto es consecuencia del siguiente resultado de análisis matemático (cuya demostración se puede ver en el libro de Parzen (1962), pág 121, o en el de Norris (1997), pág 71): Sea $g(t)$, para $t \geq 0$, una función real tal que $g(s + t) = g(s)g(t)$ para todo s y t positivos, y tal que g es acotada en intervalos finitos. Entonces se cumple que g es idénticamente cero, o bien que existe una constante λ tal que $g(t) = e^{-\lambda t}$ para todo $t > 0$. Aplicando este resultado a la función $g(t) = \mathrm{P}(Y > t)$ se obtiene la caracterización (*) de la distribución exponencial.

17 Proceso markoviano de saltos (PMS)

Contenido: Proceso markoviano de saltos (alias: cadena de Markov a tiempo continuo), cadena encajada, distribuciones finito–dimensionales, proceso estándar, proceso regular (o no explosivo).

Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un proceso de Markov a tiempo continuo, con espacio de estados S. Si S es un conjunto *discreto* — es decir, finito o infinito numerable — se dice que $\mathrm{X}(\cdot)$ es un **proceso markoviano de saltos** (abreviado: PMS), también conocido como **cadena de Markov** (CM) **a tiempo continuo**.

El nombre proceso de "saltos" se debe a que el proceso evoluciona de la siguiente manera. En el tiempo $\tau_0 := 0$ el PMS inicia en algún estado $\mathrm{X}(0) = \mathrm{X}_0$, en el cual permanece hasta un tiempo aleatorio $\tau_1 \geq \tau_0$. En el tiempo τ_1 el PMS "salta" a un estado $\mathrm{X}_1 \neq \mathrm{X}_0$, con $\mathrm{X}_1 := \mathrm{X}(\tau_1)$, en el cual permanece hasta un tiempo $\tau_2 \geq \tau_1$. En τ_2, el estado es $\mathrm{X}_2 := \mathrm{X}(\tau_2) \neq \mathrm{X}_1$, y este comportamiento se repite en tiempos $\tau_3 \leq \tau_4 \leq \cdots$. Los tiempos $0 = \tau_0 \leq \tau_1 \leq \cdots$ se llaman los **tiempos de salto** del PMS $\mathrm{X}(\cdot)$ y se pueden definir recursivamente como

$$\tau_{n+1} := \inf\{t \geq \tau_n \,|\, \mathrm{X}(t) \neq \mathrm{X}(\tau_n)\} \quad \forall\, n = 0, 1, \ldots, \tag{17.1}$$

con $\tau_0 := 0$ y, además, $\inf \phi := +\infty$. Los tiempos

$$\Delta_n := \begin{cases} \tau_n - \tau_{n-1} & \text{si} \quad \tau_{n-1} < \infty, \\ \infty & \text{en o.c.,} \end{cases} \tag{17.2}$$

para $n = 1, 2, \ldots$, se llaman los **tiempos entre saltos** o **tiempos de permanencia**, y el PE en tiempo discreto $\mathrm{X}_\bullet = \{\mathrm{X}_n, n = 0, 1, \ldots\}$, con $\mathrm{X}_n := \mathrm{X}(\tau_n)$, es la **CM encajada** en $\mathrm{X}(\cdot)$, o simplemente la **cadena de saltos** de $\mathrm{X}(\cdot)$. (Después veremos que $\mathrm{X}_\bullet$ es, en efecto, una CM.) Observe que

$$\mathrm{X}(t) = \mathrm{X}_n \quad \forall\, \tau_n \leq t < \tau_{n+1} \quad \forall\, n = 0, 1, \ldots \tag{17.3}$$

y que

$$\tau_n = \tau_{n-1} + \Delta_n = \Delta_1 + \cdots + \Delta_n \quad \forall\, n = 1, 2, \ldots \; (\tau_0 := 0). \tag{17.4}$$

La v.a.

$$\tau_\infty := \lim_{n\to\infty} \tau_n = \sum_{k=1}^{\infty} \Delta_k \tag{17.5}$$

se llama el **tiempo de explosión** del PMS. Si $\tau_\infty = +\infty$, entonces el estado $X(t)$ está dado por (3) para todo tiempo $t \geq 0$, pero si $\tau_\infty < \infty$ entonces puede pensarse que el PMS salta a un estado *ficticio*, digamos ∞, que *no* está en S. En este último caso definimos $X(t) := \infty$ para todo $t \geq \tau_\infty$.

Por el Teorema 16.15 en las Notas § 16 podemos suponer que nuestro PMS es *separable*.

Para un PMS la condición de Markov en la Observación 16.5, ecuación (16.16.6), se puede escribir como

$$P[X(t) = j \,|\, X(r)\ \forall\, 0 \leq r < s,\ \ X(s) = i] = P[X(t) = j \,|\, X(s) = i] \tag{17.6}$$

para todo $0 \leq s \leq t$ y estados i, j en S. Alternativamente, podríamos escribir la condición de Markov como en (16.16.7). El lado derecho de (6) se denota por $P_{ij}(s,t)$, es decir

$$P_{ij}(s,t) := P[X(t) = j \,|\, X(s) = i] \quad \forall\, 0 \leq s \leq t;\ i, j \in S, \tag{17.7}$$

y se le llama la **probabilidad de transición** del estado i en el tiempo s al estado j en el tiempo t. Aquí consideraremos — a menos que se diga lo contrario — únicamente el caso en el que el PMS es **homogéneo** (en el tiempo), lo cual significa que la probabilidad de transición en (7) sólo depende de la diferencia $t-s$ de los tiempos de transición. En otras palabras, consideraremos probabilidades de transición de la forma

$$P_{ij}(t) := P[X(s+t) = j \,|\, X(s) = i] \quad \forall\, s \geq 0,\ t \geq 0. \tag{17.8}$$

En particular, para $s = 0$ tenemos

$$P_{ij}(t) = P[X(t) = j \mid X(0) = i],\ t \geq 0. \tag{17.9}$$

La matriz $P(t) := [P_{ij}(t)]$ se llama la **matriz de transición** del PMS. Esta matriz caracteriza la propiedad de Markov (6) en el siguiente sentido.

Teorema 17.1. *Las siguientes proposiciones son equivalentes para un PE* $X(\cdot) = \{X(t),\, t \geq 0\}$.

(a) $X(\cdot)$ *es un PMS con matriz de transición* $P(t)$.

(b) *Para cualquier tiempo* $s \geq 0$, *y condicionado a que* $\mathrm{X}(s) = i$, *el PE* $Y(\cdot) = \{Y(t),\, t \geq 0\}$, *definido como* $Y(t) := \mathrm{X}(t+s)$, *es un PMS con estado inicial* $Y(0) = i$, *matriz de transición* $\mathrm{P}(t)$, *y es independiente de* $\{\mathrm{X}(r),\, r \leq s\}$.

En el Ejercicio 1 se pide demostrar el Teorema 17.1.

Las probabilidades de transición $\mathrm{P}_{ij}(t)$ son números no–negativos que satisfacen

$$\sum_{j \in S} \mathrm{P}_{ij}(t) = 1 \quad \forall\, 0 \leq t < \tau_\infty \tag{17.10}$$

y la **ecuación de Chapman–Kolmogorov** (Ejercicio 2)

$$\mathrm{P}_{ij}(s+t) = \sum_{k \in S} \mathrm{P}_{ik}(s)\mathrm{P}_{kj}(t) = \sum_{k \in S} \mathrm{P}_{ik}(t)\mathrm{P}_{kj}(s), \tag{17.11}$$

o en forma matricial $\mathrm{P}(s+t) = \mathrm{P}(s)\mathrm{P}(t) = \mathrm{P}(t)\mathrm{P}(s)$. Además de estas condiciones *supondremos que el PMS es* **estándar** lo cual significa que

$$\lim_{t \downarrow 0} \mathrm{P}_{ij}(t) = \mathrm{P}_{ij}(0) = \delta_{ij}, \tag{17.12}$$

en donde δ_{ij} es la delta de Kronecker. Entre otras cosas, (12) implica que las probabilidades de transición $\mathrm{P}_{ij}(t)$ son *uniformemente continuas* en $t \geq 0$. (Vea el Ejercicio 17.3.)

Proposición 17.2. Un PMS es estocásticamente continuo.

Demostración. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\ t \geq 0\}$ un PMS, y sean $\varepsilon > 0$, $t > s \geq 0$. Entonces

$$\begin{aligned}
\mathrm{P}(|\mathrm{X}(t) - \mathrm{X}(s)| \geq \varepsilon) &= \sum_{i \in S} \mathrm{P}(\mathrm{X}(s) = i)\mathrm{P}(|\mathrm{X}(t) - i| \geq \varepsilon | \mathrm{X}(s) = i) \\
&\leq \sum_{i \in S} \mathrm{P}(\mathrm{X}(s) = i) \sum_{j \neq i} \mathrm{P}_{ij}(t-s) \\
&= \sum_{i \in S} \mathrm{P}(\mathrm{X}(s) = i)[1 - \mathrm{P}_{ii}(t-s)].
\end{aligned}$$

Por lo tanto, por convergencia dominada,

$$\mathrm{P}(|\mathrm{X}(t) - \mathrm{X}(s)| \geq \varepsilon) \to 0 \quad \text{cuando} \quad t \to s.$$

□

Observación 17.3. Del Lema 3 en Gikhman y Skorokhod (1969), p.348, se obtiene lo siguiente (ver también (12)).

(a) Un PMS no tiene discontinuidades de segundo orden y, además,

(b) tiene una versión cuyas trayectorias son, c.s., continuas por la derecha.

En vista de la Observación 17.3(b), en todo lo que sigue supondremos que nuestro PMS tiene trayectorias continuas por la derecha.

Definición 17.4. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ un PE con espacio de estados $(S, \mathcal{S})$.

(a) Una v.a. T con valores en $[0, \infty]$ se dice que es un **tiempo de paro** de $\mathrm{X}(\cdot)$ si para cualquier $t \geq 0$ el evento $\{T \leq t\}$ pertenece a la σ–álgebra generada por $\{\mathrm{X}(s),\, 0 \leq s \leq t\}$.

(b) Las probabilidades de la forma

$$\mathrm{P}[\mathrm{X}(t_0) \in B_0,\ \mathrm{X}(t_1) \in B_1, \ldots, \mathrm{X}(t_n) \in B_n] \tag{17.13}$$

para $n \geq 0$, tiempos $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$, y conjuntos $B_0, \ldots, B_n$ en $\mathcal{S}$ se llaman las **distribuciones finito–dimensionales** de $\mathrm{X}(\cdot)$. Si $\mathrm{X}(\cdot)$ es un PMS, los eventos $\{\mathrm{X}(t_k) \in B_k\}$ en (13) se pueden sustituir por $\{\mathrm{X}(t_k) = i_k\}$ con $i_k \in S$, para $k = 0, 1, \ldots, n$.

En relación con el siguiente teorema (que no demostraremos) vea las Notas 1 y 2 en el final de esta sección.

Teorema 17.5. *Sea* $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ *un PMS continuo por la derecha. Entonces:*

(a) *La probabilidad de cualquier evento A relativo a* $\mathrm{X}(\cdot)$ *(es decir, cualquier evento A en la σ–álgebra $\sigma\{\mathrm{X}(t),\, t \geq 0\}$) se puede determinar por las distribuciones finito–dimensionales en (13).*

(b) $\mathrm{X}(\cdot)$ *tiene la* **propiedad fuerte de Markov***; es decir, si T es un tiempo de paro de* $\mathrm{X}(\cdot)$ *con* $\mathrm{P}(T < \infty) = 1$*, entonces*

$$\begin{aligned} \mathrm{P}[\mathrm{X}(T+t) &= j \,|\, \mathrm{X}(r)\ \forall\, 0 \leq r < T,\ \mathrm{X}(T) = i] \\ &= \mathrm{P}[\mathrm{X}(t) = j \,|\, \mathrm{X}(0) = i]. \end{aligned} \tag{17.14}$$

para todo $t \geq 0$. [Compare (14) con (6).] Equivalentemente, condicionado a que $T < \infty$ y $\mathrm{X}(T) = i$, el PE $Y(\cdot) = \{\mathrm{X}(T+t),\, t \geq 0\}$ es un PMS con estado inicial $Y(0) = i$, y $Y(\cdot)$ es independiente de $\{\mathrm{X}(s),\, s \leq T\}$. [Compare este enunciado con el Teorema 17.1(b).]

La ecuación (14) la usaremos repetidas veces con el tiempo de paro $T = \tau_n$, suponiendo que $\tau_n < \infty$. Para ver que efectivamente τ_n es un tiempo de paro basta observar que, para cualquier $t \geq 0$, el evento $\{\tau_n \leq t\}$ es equivalente a que $\mathrm{X}(\cdot)$ haya saltado al menos n veces en el periodo $[0,t]$. Esto se puede precisar aún más introduciendo el proceso de conteo $N(\cdot) = \{N(t),\, t \geq 0\}$, en donde $N(t)$ es el **número de transiciones** o **saltos** del PMS hasta el tiempo t; es decir,

$$N(t) := \sup\{k \mid \tau_k \leq t\} \quad \text{para} \quad t \geq 0. \tag{17.15}$$

Entonces $\tau_n \leq t$ si y sólo si $N(t) \geq n$.

En el resto de esta sección deseamos probar que la cadena de saltos $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ del PMS [con $\mathrm{X}_n := \mathrm{X}(\tau_n)$] es en efecto una CM y que, dados los estados $\mathrm{X}_0, \ldots, \mathrm{X}_{n-1}$, los tiempos entre saltos $\Delta_1, \ldots, \Delta_n$ son vv.aa. independientes distribuidas exponencialmente. Empezaremos por demostrar lo siguiente.

Proposición 17.6. Para cada $i \in S$, $n \geq 1$ y $t \geq 0$:

$$\mathrm{P}[\Delta_n > t \mid \mathrm{X}(\tau_{n-1}) = i] = e^{-\alpha_i t} \tag{17.16}$$

para algún número $0 \leq \alpha_i < \infty$.

Nota 17.7. Se dice que un estado $i \in S$ es *instantáneo* si el parámetro en (16) es $\alpha_i = +\infty$. Bajo las hipótesis en esta sección, se sigue de la Observación 17.3(b) que el PMS no tiene estados instantáneos.

Demostración. Por la homogeneidad del PMS podemos tomar $n = 1$, de modo que $\Delta_1 = \tau_1$ y el lado izquierdo de (16) se puede escribir como $g(t) := \mathrm{P}[\tau_1 > t \mid \mathrm{X}(0) = i]$. Por la observación en el Ejercicio 16.31, para demostrar (16) basta verificar que g satisface la ecuación

$$g(t+s) = g(s)g(t). \tag{17.17}$$

Ahora observe que mediante un cálculo directo se obtiene

$$\mathrm{P}[\tau_1 > t+s \mid \mathrm{X}(0) = i,\ \tau_1 > t] = \frac{g(t+s)}{g(t)}. \tag{17.18}$$

Por otra parte, por la propiedad fuerte de Markov y la homogeneidad de $X(\cdot)$, el lado izquierdo de (18) resulta

$$\begin{aligned} P[\tau_1 > t+s \,|\, X(0)=i,\, \tau_1 > t] &= P[\tau_1 > t+s \,|\, X(t)=i] \\ &= P[\tau_1 > s \,|\, X(0)=i] \\ &= g(s). \end{aligned}$$

Este resultado junto con (18) da (17). □

Definición 17.8. Sea α_i el número en (16). Si $\alpha_i = 0$, decimos que i es un estado **absorbente**.

Si $i \in S$ es absorbente, entonces (16) resulta

$$P[\Delta_n > t \mid X(\tau_{n-1}) = i] = 1 \quad \forall\, t \geq 0$$

y, por lo tanto, si el PMS alcanza el estado i, entonces el proceso se queda en i permanentemente. En otras palabras, si $X(t) = i$ para algún t, entonces $X(s) = i$ para todo $s \geq t$. Por otra parte, si $\alpha_i > 0$ entonces la distribución condicional de Δ_n dado que $X_{n-1} := X(\tau_{n-1}) = i$ es $\mathrm{Exp}(\alpha_i)$.

A continuación presentamos el resultado anunciado en el párrafo anterior a la Proposición 17.6. Además de que el PMS es continuo por la derecha (Observación 17.3(b)) supondremos que el tiempo de explosión τ_∞ definido en (5) es infinito, en cuyo caso se dice que el PMS es **regular** (o **no explosivo**).

Hipótesis 17.9. El PMS es regular, es decir $P(\tau_\infty = \infty) = 1$.

En la Proposición 17.11 se dan algunas condiciones bajo las que $X(\cdot)$ es regular.

Teorema 17.10. *Suponga que el PMS es continuo por la derecha y regular, y que $\alpha_i > 0$ para todo $i \in S$. Entonces para cualesquiera dos estados i y j, y $t \geq 0$, se tiene:*

$$P[X_{n+1} = j,\, \Delta_{n+1} > t \,|\, X_0, \ldots, X_{n-1}, X_n = i,\, \Delta_1, \ldots, \Delta_n] \tag{17.19}$$

$$\begin{aligned} &= P[X_{n+1} = j,\ \Delta_{n+1} > t \,|\, X_n = i] \qquad (17.20) \\ &= p_{ij} e^{-\alpha_i t} \end{aligned}$$

en donde los $p_{ij} = \mathrm{P}[\mathrm{X}_{n+1} = j | \mathrm{X}_n = i]$ *satisfacen que*

$$p_{ij} \geq 0 \quad si \quad i \neq j, \; p_{ii} = 0, \quad y \quad \sum_{j \in S} p_{ij} = 1 \quad \forall\, i \in S. \tag{17.21}$$

La matriz $\mathrm{P} = [p_{ij}]$ *es la matriz de transición de la cadena de saltos* $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$*.*

Demostración. Sean $i \neq j$ estados arbitrarios. Por (3) y (4), el evento condicionante en (19) determina el PMS $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ para $0 \leq t \leq \tau_n$. Por lo tanto, la probabilidad condicional en (19) se puede escribir como

$$\begin{aligned} &\mathrm{P}[\mathrm{X}_{n+1} = j,\; \Delta_{n+1} > t \,|\, \mathrm{X}(s) \;\text{ para }\; s < \tau_n,\, \mathrm{X}(\tau_n) = i] \\ &= \mathrm{P}[X(\tau_n + \Delta_{n+1}) = j, \Delta_{n+1} > t | X(s), s < \tau_n, X(\tau_n) = i] \\ &= \mathrm{P}[\mathrm{X}_{n+1} = j,\; \Delta_{n+1} > t \,|\, \mathrm{X}(\tau_n) = i], \end{aligned} \tag{17.22}$$

por la propiedad fuerte de Markov, y se obtiene la primera igualdad en (20). Por otra parte, usando la homogeneidad del PMS la probabilidad en (22) resulta

$$\begin{aligned} \mathrm{P}[\mathrm{X}_1 &= j,\, \Delta_1 > t \,|\, \mathrm{X}(0) = i] \\ &= \mathrm{P}[\Delta_1 > t \,|\, \mathrm{X}_0 = i]\mathrm{P}[\mathrm{X}_1 = j \,|\, \mathrm{X}_0 = i,\, \Delta_1 > t] \\ &= e^{-\alpha_i t}\mathrm{P}[\mathrm{X}_1 = j \,|\, \mathrm{X}_0 = i,\, \Delta_1 > t]. \end{aligned} \tag{17.23}$$

Ahora deseamos demostrar que

$$\mathrm{P}[\mathrm{X}_1 = j \,|\, \mathrm{X}_0 = i,\, \Delta_1 > t] = \mathrm{P}[\mathrm{X}_1 = j \,|\, \mathrm{X}_0 = i] = p_{ij} \tag{17.24}$$

para lo cual conviene introducir el **tiempo residual**

$$Y(t) := \tau_{N(t)+1} - t \quad \forall\, t \geq 0, \tag{17.25}$$

en donde $N(t)$ es el número de transiciones hasta el tiempo t, definido en (15). Observe que $Y(t)$ es el tiempo hasta la primera transición *después* de t, y que

$$\mathrm{X}_1 = \mathrm{X}(t + Y(t)) \quad \text{para} \quad t < \tau_1 = \Delta_1.$$

Además, el evento $\{\mathrm{X}_0 = i,\, \Delta_1 > t\}$ es equivalente al evento $\{\mathrm{X}(s) = i \;\forall\, 0 \leq s \leq t\}$. Usando estas observaciones junto con la propiedad de Markov y la homogeneidad del PMS, el lado izquierdo de (24) resulta

$$\mathrm{P}[\mathrm{X}_1 = j \,|\, \mathrm{X}_0 = i,\, \Delta_1 > t] = \mathrm{P}[\mathrm{X}(t + Y(t)) = j \,|\, \mathrm{X}(s) = i \;\forall\, 0 \leq s \leq t]$$

$$
\begin{aligned}
&= \mathrm{P}[\mathrm{X}(t+Y(t)) = j \,|\, \mathrm{X}(t) = i] \\
&= \mathrm{P}[\mathrm{X}(Y(0)) = j \,|\, \mathrm{X}(0) = i] \\
&= \mathrm{P}[\mathrm{X}_1 = j \,|\, \mathrm{X}_0 = i].
\end{aligned}
$$

Esto demuestra (24) que junto con (22) y (23) da la última igualdad en (19)–(20). A su vez, esta última igualdad implica (21). $\square$

Concluiremos esta sección con algunas observaciones sobre la Hipótesis 17.9 de regularidad (o no–explosividad) del PMS y con un par de ejemplos.

Sea $\mathrm{X}_\bullet = \{\mathrm{X}_0, \mathrm{X}_1, \ldots\}$ la cadena de saltos del PMS y $\{\Delta_1, \Delta_2, \ldots\}$ la sucesión de tiempos entre saltos. Dada la CM $\mathrm{X}_\bullet$, los tiempos entre saltos son vv.aa. distribuidas exponencialmente, digamos $\Delta_n \sim \mathrm{Exp}(\lambda_n)$. (Por el Teorema 17.10, $\lambda_n = \alpha_i$ si $\mathrm{X}_{n-1} = i$.) Por la definición (5) del tiempo de explosión y el Teorema de Convergencia Monótona

$$
E(\tau_\infty) = \sum_{n=1}^{\infty} E(\Delta_n) = \sum_{n=1}^{\infty} (1/\lambda_n). \tag{17.26}
$$

De aquí se sigue que si $\sum_{n=1}^{\infty}(1/\lambda_n) < \infty$, entonces el PMS es *explosivo* porque $E(\tau_\infty) < \infty$ implica que $\mathrm{P}(\tau_\infty < \infty) = 1$. En otras palabras, si los parámetros λ_n son “muy grandes”, de tal forma que $\sum(1/\lambda_n) < \infty$, entonces el PMS *no* es regular. (Esto se cumple, por ejemplo, si $\lambda_n = n^r$ con $r > 1$.) Esto sugiere que, como en la siguiente proposición, el PMS es regular si los λ_n no son “demasiado grandes”.

Proposición 17.11. El PMS $\mathrm{X}(\cdot)$ es regular si se cumple alguna de las siguientes condiciones:

(a) El espacio de estados S es finito.

(b) $\sup_{i \in S} \alpha_i < \infty$.

(c) $\sum_{i \in S}(1/\alpha_i) = \infty$.

(d) $\alpha_i < \infty$ para todo $i \in S$ y la CM encajada es recurrente.

Demostración. Para evitar casos triviales supondremos que $\mathrm{X}(\cdot)$ no tiene estados absorbentes, de modo que $0 < \alpha_i < \infty$ para todo $i \in S$.

Para demostrar (a), (b) y (c) primero veremos que si $\Delta_1, \Delta_2, \ldots$ son vv.aa. independientes distribuidas exponencialmente con parámetros

$\lambda_1, \lambda_2, \ldots$ y tales que

$$\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty, \tag{17.27}$$

entonces

$$\mathrm{P}(\tau_\infty = \infty) = 1, \tag{17.28}$$

donde $\tau_\infty := \sum_{n=1}^{\infty} \Delta_n$. Nótese que

$$\prod_{n=1}^{\infty}(1+\lambda_n^{-1}) = \lim_{m\to\infty}\prod_{n=1}^{m}(1+\lambda_n^{-1}) \geq \lim_{m\to\infty}(1+\sum_{n=1}^{m}\lambda_n^{-1}) = \infty.$$

Por lo tanto, por independencia de las Δ_n y el Teorema de Convergencia Acotada,

$$E(e^{-\tau_\infty}) = \prod_{n=1}^{\infty} E(e^{-\Delta_n}) = \prod_{n=1}^{\infty}(1+\lambda_n^{-1})^{-1} = 0.$$

Esto implica (28).

Ahora, por el Teorema 17.10, y escribiendo $\alpha(i) := \alpha_i$, dada la cadena de saltos $\mathrm{X}_n = i_n$ para $n = 0, 1, \ldots$, los tiempos entre saltos Δ_n son independientes con distribución $\mathrm{Exp}(\lambda_n)$ y parámetros $\lambda_n := \alpha(i_{n-1})$ para $n = 1, 2 \ldots$. En los incisos (a) y (b), tomando $M := \sup_i \alpha(i)$ vemos que $1/\lambda_n = 1/\alpha(i_{n-1}) \geq 1/M$, lo cual implica (27) y, por lo tanto, (28). Asimismo, por (27), en el inciso (c) se obtiene de nuevo (28). Finalmente, la conclusión deseada también se obtiene en el caso (d), porque si la cadena de saltos es recurrente entonces visita cualquier estado inicial $\mathrm{X}_0 = i$ una infinidad de veces con probabilidad 1. $\square$

Ejemplo 17.12. Suponga que el PMS $\mathrm{X}(\cdot)$ es un proceso de Poisson con parámetro λ. Entonces por el Teorema 17.10(b), la correspondiente cadena de saltos $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$ toma los valores $\mathrm{X}_n := \mathrm{X}(\tau_n) = n$, y las probabilidades de transición en (21) son

$$p_{n,m} = 1 \quad \text{si} \quad m = n+1$$

y $p_{n,m} = 0$ si $m \neq n+1$. Es decir, la matriz de transición de $\mathrm{X}_\bullet$ es de la forma

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

De nuevo por el Teorema 16.11(b), los tiempos entre saltos Δ_n son i.i.d. con distribución Exp(λ). En este caso se satisfacen tanto la condición (b) como la (c) de la Proposición 17.11, de modo que el proceso de Poisson es un PMS regular.

Ejemplo 17.13. Sea $N(\cdot)$ un proceso de Poisson con parámetro λ, y sea X_0 una v.a. independiente de $N(\cdot)$ y con distribución $\mathrm{P}(\mathrm{X}_0 = i) = 1/2$ para $i = 1, -1$. Entonces el PE

$$\mathrm{X}(t) := \mathrm{X}_0 \cdot (-1)^{N(t)} \quad \text{para} \quad t \geq 0$$

se llama una **señal telegráfica aleatoria**. (Vea el Ejercicio 16.7.) Este es un PMS regular, cuya cadena de saltos tiene espacio de estados $S = \{-1, 1\}$ y probabilidades de transición $p_{ij} = 1 - \delta_{ij}$, es decir la matriz de transición es de la forma

$$\begin{array}{cc} & \begin{array}{cc} -1 & 1 \end{array} \\ \begin{array}{c} -1 \\ 1 \end{array} & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}.$$

Los tiempos entre saltos son i.i.d. con distribución Exp(λ).

En la siguiente sección veremos otros ejemplos.

Notas § 17

1. Sea Ω un conjunto arbitrario. Se dice que una familia $\mathcal{A}$ de subconjuntos de Ω es un $\boldsymbol{\pi}$**–sistema** si $\mathcal{A}$ es cerrada bajo intersecciones finitas, es decir, si A_1, A_2 están en $\mathcal{A}$, entonces $A_1 \cap A_2$ está en $\mathcal{A}$. La demostración del siguiente teorema se puede ver, por ejemplo, en Blumenthal y Getoor (1968), pág. 7.

Teorema. Sea $(\Omega, \mathcal{F})$ un espacio medible. Si P_1 y P_2 son dos medidas de probabilidad sobre $(\Omega, \mathcal{F})$ que coinciden sobre un π–sistema $\mathcal{A}$ que genera a $\mathcal{F}$ [i.e. $\sigma\{\mathcal{A}\} = \mathcal{F}$], entonces $\mathrm{P}_1 = \mathrm{P}_2$.

De este teorema se puede obtener que las distribuciones finito–dimensionales (13) de un PMS $\mathrm{X}(\cdot)$ continuo por la derecha determinan la probabilidad de cualquier evento relativo a $\mathrm{X}(\cdot)$. Esto se debe a que para este tipo de procesos los eventos de la forma $\{\mathrm{X}(t_0) = i_0, \ldots, \mathrm{X}(t_n) = i_n\}$ forman un π–sistema que genera a la σ–álgebra $\sigma\{\mathrm{X}(t),\, t \geq 0\}$. El papel de la continuidad por la derecha es que sin ella algunos eventos (por ejemplo, $\{\mathrm{X}(t) = i$ para algún $t > 0\}$) pueden *no* ser medibles con respecto a la σ–álgebra $\sigma\{\mathrm{X}(t),\, t \geq 0\}$.

2. La demostración de la propiedad fuerte de Markov en el Teorema 17.5(b) se puede ver, por ejemplo, en Gikhman y Skorokhod (1969), página 359, Corolario 2. Dicho teorema es un caso especial de un resultado mucho más general de Dynkin y Yushkevich (1956), Teorema 2. El resultado de Dynkin y Yushkevich afirma que un proceso de Markov $\mathrm{X}(\cdot)$, con espacio de estados S (no necesariamente discreto) tiene la propiedad fuerte de Markov si $\mathrm{X}(\cdot)$ es continuo por la derecha y sus probabilidades de transición satisfacen la "condición de Feller" — esta última condición se cumple trivialmente cuando el espacio S es discreto, es decir, cuando $\mathrm{X}(\cdot)$ es un PMS.

3. Para cada $t \geq 0$, sea $\mathrm{P}(t) = [p_{ij}(t)]$ la matriz de transición de un PMS. Entonces $\mathrm{P}(t)$ es una matriz estocástica que satisface

(a) $\mathrm{P}(0) = I$, es decir $p_{ij}(0) = \delta_{ij}$, y

(b) la ecuación de Chapman–Kolmogorov $\mathrm{P}(s+t) = \mathrm{P}(s)\mathrm{P}(t)$.

Debido a estas propiedades se dice que la familia $\{\mathrm{P}(t),\, t \geq 0\}$ es un **semigrupo** (estocástico). Si además el PMS es estándar, como en (12), entonces las probabilidades de transición $p_{ij}(t)$ son continuas (Ejercicio 3), y por lo tanto se dice que el semigrupo $\{\mathrm{P}(t)\, t \geq 0\}$ es **estándar** o **continuo**.

Ejercicios § 17

17.1. Demuestre el Teorema 17.2

17.2. Demuestre la ecuación de Chapman–Kolmogorov (11).

17.3. Use la ecuación de Chapman–Kolmogorov para demostrar que si $h > 0$, entonces

$$-[1 - \mathrm{P}_{ii}(h)] \leq \mathrm{P}_{ij}(t+h) - \mathrm{P}_{ij}(t) \leq 1 - \mathrm{P}_{ii}(h),$$

es decir, $|\mathrm{P}_{ij}(t+h) - \mathrm{P}_{ij}(t)| \leq 1 - \mathrm{P}_{ii}(h)$. Análogamente, para $h < 0$

$$|\mathrm{P}_{ij}(t+h) - \mathrm{P}_{ij}(t)| \leq 1 - \mathrm{P}_{ii}(|h|).$$

Usando estos resultados concluya que, por (12), las probabilidades de transición $\mathrm{P}_{ij}(t)$ son uniformemente continuas en $t \geq 0$.

17.4. Demuestre: si $Y_1, Y_2, \ldots$ son vv.aa. independientes distribuidas exponencialmente con parámetros $\lambda_1, \lambda_2, \ldots$, entonces las vv.aa. $\lambda_1 Y_1, \lambda_2 Y_2, \ldots$ son i.i.d. con distribución Exp(1).

17.5. Sean $Y_1, Y_2, \ldots$ vv.aa. i.i.d. con distribución exponencial $\mathrm{Exp}(\lambda)$, y sea N una v.a. independiente de $\{Y_k\}$ con distribución geométrica

$$\mathrm{P}(N = n) = \alpha(1-\alpha)^{n-1} \quad \text{para} \quad n = 1, 2, \ldots,\ 0 < \alpha < 1.$$

Demuestre que la suma aleatoria $Y_1 + \cdots + Y_N$ tiene distribución $\mathrm{Exp}(\alpha\lambda)$.

17.6. Demuestre que si un PMS es estándard (vea la ecuación (12)), entonces

$$\mathrm{P}_{ii}(t) > 0 \quad \forall\, t \geq 0.$$

(*Nota*: Un resultado más difícil es que bajo la misma hipótesis, para $i \neq j$ se satisface la llamada *dicotomía de Lévy:* (a) $\mathrm{P}_{ij}(t) = 0 \ \ \forall\ t > 0$; ó (b) $\mathrm{P}_{ij}(t) > 0 \ \ \forall\, t > 0$.)

17.7. Considere un PMS con matriz de transición $\mathrm{P}(t)$. Se dice que el PMS es **uniforme** si $\sup_i |\mathrm{P}_{ii}(t) - 1| \to 0$ cuando $t \downarrow 0$. Demuestre que si un PMS es uniforme, entonces es estándar. Recíprocamente, si el PMS es estándar y el espacio de estados es *finito*, entonces el PMS es uniforme. (En general, un PMS estándar no necesariamente es uniforme; vea el Ejercicio 18.9.)

18 La matriz generadora de un PMS

Contenido: Tasas de transición, matriz generadora (alias: matriz infinitesimal o Q–matriz), proceso estable, proceso conservante, ecuaciones de Kolmogorov.

En esta sección introducimos la matriz generadora de un PMS. Dicha matriz, que también es conocida como *matriz infinitesimal* o *Q-matriz*, es fundamental para el análisis de un PMS. En particular, nos permite obtener las ecuaciones de Kolmogorov y determinar, bajo ciertas condiciones, la existencia de medidas estacionarias.

En toda esta sección supondremos que $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ es un PMS *estándar* [vea la ecuación (17.17.12)], *continuo por la derecha* (Observación 17.3) y *regular* o no–explosivo (Hipótesis 17.9). El espacio de estados de $\mathrm{X}(\cdot)$ se denota por S, como en secciones anteriores.

Para cualesquiera dos estados i, j definimos la **intensidad** (o **tasa**) **de transición** q_{ij} de ir de i a j como la derivada de $\mathrm{P}_{ij}(t)$ en $t = 0$ por la derecha, es decir,

$$q_{ij} := \mathrm{P}'_{ij}(t)|_{t=0+} = \lim_{t \downarrow 0} \frac{1}{t}[\mathrm{P}_{ij}(t) - \delta_{ij}]. \tag{18.1}$$

Equivalentemente [usando la notación $o(t)$ introducida en la Observación 1 que aparece después del Ejercicio 16.22], si el límite en (1) existe entonces podemos escribir

$$\mathrm{P}_{ij}(t) = q_{ij}t + o(t) \quad \text{para} \quad i \neq j, \tag{18.2}$$

$$\mathrm{P}_{ii}(t) = 1 - q_i t + o(t), \quad \text{con} \quad q_i := -q_{ii}, \tag{18.3}$$

en donde los términos $o(t)$ dependen, en general, de los estados i y j. En otras palabras, cuando $t \downarrow 0$, las probabilidades $\mathrm{P}_{ij}(t)$ y $1-\mathrm{P}_{ii}(t)$ son proporcionales a t, excepto por términos $o(t)$, con constantes de proporcionalidad q_{ij} y q_i, respectivamente. En particular, como

$$1 - \mathrm{P}_{ii}(t) = \sum_{j \neq i} \mathrm{P}_{ij}(t), \tag{18.4}$$

$q_i = -q_{ii}$ es la intensidad de transición de i a un estado distinto de i. Por supuesto, si i es un estado *absorbente*, entonces $q_i = 0$.

La matriz $Q = [q_{ij}]$ se llama la **matriz generadora** del PMS, y también se le conoce como **la matriz infinitesimal** o la **Q-matriz** del PMS.

El resultado fundamental sobre las intensidades de transición es el siguiente teorema cuya demostración se puede ver, por ejemplo, en el libro de Gikhman y Skorokhod (1969), páginas 304-308.

Teorema 18.1. (a) *Para $i \neq j$, los límites $q_{ij} \geq 0$ en (1) existen y son finitos.*

(b) *El límite $q_i := -q_{ii} \geq 0$ existe, pero puede ser infinito.*

(c) *Si $q_i < \infty$, entonces las derivadas $\mathrm{P}'_{ij}(t)$ existen para todo $t \geq 0$ y $j \in S$. Además, dichas derivadas son continuas y satisfacen que:*

(c.1) $\mathrm{P}'_{ij}(t+s) = \sum_{k \in S} \mathrm{P}'_{ik}(t)\mathrm{P}_{kj}(s) \quad \forall\, t, s > 0,$

(c.2) $\sum_{j \in S} \mathrm{P}'_{ij}(t) = 0 \quad \forall\, t > 0,$ *y*

(c.3) $\sum_{j \in S} |\mathrm{P}'_{ij}(t)| \leq 2q_i \;\forall\, t \geq 0.$

Nótese que si $q_i < \infty$, entonces la condición (c.1) equivale a derivar (con respecto a t) la ecuación de Chapman–Kolmogorov. Asimismo, (c.2) equivale a derivar la ecuación

$$\sum_{j \in S} \mathrm{P}_{ij}(t) = 1$$

que obviamente es la misma que (4). En particular, si (c.2) se cumple en $t = 0$ se obtiene que

$$\sum_{j \in S} q_{ij} = 0, \quad \text{i.e} \quad q_i = \sum_{j \neq i} q_{ij}. \tag{18.5}$$

Estas observaciones son de gran utilidad y, por tal motivo, supondremos lo siguiente.

Hipótesis 18.2. El PMS (o la matriz generadora Q) es:

(a) **estable**, es decir, $q_i = -q_{ii} < \infty$ para cada $i \in S$, y

(b) **conservante**, es decir, satisface (5) para cada $i \in S$.

Bajo estas hipótesis podemos relacionar la matriz generadora $Q = [q_{ij}]$ con la matriz de transición $\mathrm{P} = [p_{ij}]$ y los parámetros α_i en el Teorema 17.10 de la siguiente manera.

Teorema 18.3. *Supóngase que* $\mathrm{X}(\cdot)$ *es un PMS estándar, continuo por la derecha y regular, y además se cumple la Hipótesis 18.2. Entonces:*

(a) $q_i = \alpha_i \; \forall \, i \in S$.

(b) *Si* i *no es absorbente (es decir,* $\alpha_i = q_i > 0$*), entonces* $p_{ii} = 0$ *y* $p_{ij} = q_{ij}/q_i$ *para* $j \neq i$*, o sea*

$$p_{ij} = (1 - \delta_{ij}) q_{ij}/q_i \qquad \forall \, j \in S. \tag{18.6}$$

(c) *Si* i *es absorbente (es decir,* $\alpha_i = q_i = 0$*), entonces* $p_{ij} = \delta_{ij}$ *para todo* $j \in S$.

Demostración. Observe que el inciso (c) es trivial, de manera que sólo demostraremos (a) y (b). Con este fin, considere el número de saltos $N(t) := \max\{k | \tau_k \leq t\}$ del PMS hasta el tiempo $t \geq 0$ [que ya habíamos definido en la ecuación (17.17.15)]. Primero veremos que cuando $t \downarrow 0$

$$\mathrm{P}_i[N(t) = 0] = 1 - \alpha_i t + o(t) \tag{18.7}$$

y

$$\mathrm{P}_i[N(t) = 1] = \alpha_i t + o(t), \tag{18.8}$$

de modo que

$$\mathrm{P}_i[N(t) \geq 2] = 1 - \mathrm{P}_i[N(t) \leq 1] = o(t). \tag{18.9}$$

(En estas ecuaciones el subindice i en P_i represanta la condición $\mathrm{X}(0) = i$.) La relación (7) se obtiene de (17.17.16) (o de (17.17.21)) porque

$$\mathrm{P}_i[N(t) = 0] = \mathrm{P}[\Delta_1 > t | \mathrm{X}(0) = i] = e^{-\alpha_i t},$$

en donde $\Delta_1 = \tau_1$ es el tiempo del primer salto de $\mathrm{X}(\cdot)$. Para verificar (8), note que los eventos $\{N(t) = 1\}$ y $\{\tau_1 \leq t < \tau_2\}$ son equivalentes. Por lo tanto,

$$\begin{aligned}
\mathrm{P}_i[N(t) = 1] &= \mathrm{P}_i[\tau_1 \leq t < \tau_2] \\
&= \sum_{j \neq i} p_{ij} \int_0^t \alpha_i e^{-\alpha_i s} P_j(\Delta_2 > t - s) ds \\
&= \alpha_i t + o(t).
\end{aligned}$$

Esto demuestra (8). Para ver la última igualdad se puede usar el siguiente cálculo:

$$\sum_{j\neq i} p_{ij}\int_0^t \alpha_i e^{-\alpha_i s}P_j(\Delta_2 > t-s)ds = \sum_{j\neq i} p_{ij}\alpha_i\int_0^t e^{-\alpha_i s}e^{-\alpha_j(t-s)}ds$$
$$= \sum_{j\neq i} p_{ij}\alpha_i e^{-\alpha_j t}\int_0^t e^{-\alpha_i s}e^{\alpha_j s}ds = \sum_{j\neq i} p_{ij}\alpha_i(1-\alpha_j t+o(t))\int_0^t e^{(\alpha_j-\alpha_i)s}ds$$
$$= \sum_{j\neq i} p_{ij}\alpha_i\int_0^t e^{(\alpha_j-\alpha_i)s}ds + \sum_{j\neq i} p_{ij}\alpha_i(-\alpha_j t+o(t))\int_0^t e^{(\alpha_j-\alpha_i)s}ds \pm \alpha_i t.$$

Regresando a la demostración de (a) y (b), de (7) y (9), como

$$P_i(N(t)\geq 2, X(t)=i) \leq P_i(N(t)\geq 2),$$

entonces $P_i(N(t)\geq 2, X(t)=i)=o(t)$. Así

$$\mathrm{P}_{ii}(t) = \mathrm{P}_i[N(t)=0] + \mathrm{P}_i[N(t)\geq 2, X(t)=i] = 1-\alpha_i t+o(t).$$

Utilizando (1) y la hipótesis de que es estable se obtiene (a). Por otra parte, para $j\neq i$,

$$\begin{aligned}
\mathrm{P}_{ij}(t) &= \mathrm{P}_i[\mathrm{X}(t)=j] = \mathrm{P}_i[N(t)\geq 1,\ \mathrm{X}(t)=j]\\
&= \mathrm{P}_i[N(t)=1,\ \mathrm{X}(t)=j] + \mathrm{P}_i[N(t)\geq 2,\ \mathrm{X}(t)=j]\\
&= [\alpha_i t+o(t)]p_{ij}+o(t) \quad \text{[por (8) y (9)];}
\end{aligned}$$

En la última igualdad hemos usando el hecho de que el tiempo de permanencia es independiente del estado a donde salte (ver Teorema 17.17.10). Podemos decir pues que

$$\mathrm{P}_{ij}(t) = \alpha_i p_{ij} t + o(t).$$

Esta ecuación y (1) dan $q_{ij}=\alpha_i\, p_{ij} = q_i\, p_{ij}$, y se obtiene (6) para $j\neq i$. Para $j=i$, el resultado $p_{ii}=0$ viene del Teorema 17.10. $\square$

Ecuaciones de Kolmogorov. De acuerdo con la definición (1) de las intensidades de transición, si conocemos la matriz de transición $\mathrm{P}(t)=[\mathrm{P}_{ij}(t)]$ entonces podemos calcular la matriz generadora $Q=[q_{ij}]$. Ahora consideraremos el problema recíproco: si conocemos Q, ¿cómo se calcula $P(t)$?.

Esto se puede hacer mediante la **ecuación de Kolmogorov hacia atrás** (también conocida como **primera ecuación de Kolmogorov**)

$$\mathrm{P}'(t) = Q\mathrm{P}(t) \quad \forall\, t > 0, \quad \text{con} \quad \mathrm{P}(0) = I, \tag{18.10}$$

o bien mediante la **ecuación de Kolmogorov hacia adelante** (o **segunda ecuación de Kolmogorov**)

$$\mathrm{P}'(t) = \mathrm{P}(t)Q \quad \forall\, t > 0, \quad \text{con} \quad \mathrm{P}(0) = I. \tag{18.11}$$

La principal diferencia entre estas ecuaciones es que para ser válidas requieren hipótesis diferentes. Antes de ver estas hipótesis, nótese que término a término la ecuación (10) se puede escribir como

$$\mathrm{P}'_{ij}(t) = -q_i\mathrm{P}_{ij}(t) + \sum_{k\neq i} q_{ik}\mathrm{P}_{kj}(t) \quad \forall\, t > 0, \tag{18.12}$$

en donde $q_i := -q_{ii}$, y, análogamente, (11) resulta

$$\mathrm{P}'_{ij}(t) = -\mathrm{P}_{ij}(t)q_j + \sum_{k\neq j} \mathrm{P}_{ik}(t)q_{kj} \quad \forall\, t > 0. \tag{18.13}$$

En ambos casos la condición inicial es $\mathrm{P}_{ij}(0) = \delta_{ij}$.

Teorema 18.4. *Sea* $\mathrm{X}(\cdot)$ *un PMS estándar, continuo por la derecha y regular.*

(a) *La ecuación de Kolmogorov hacia atrás, (12), se cumple para todo* i, j *si y sólo si se cumple la Hipótesis 18.2.*

(b) *Si* $\mathrm{X}(\cdot)$ *es estable y los límites*

$$q_{kj} = \lim_{h\downarrow 0} \mathrm{P}_{kj}(h)/h \quad \forall\, k \neq j \tag{18.14}$$

se satisfacen uniformemente en k *(para cada* j*), entonces se cumple la ecuación de Kolmogorov hacia adelante, (13), para todo* i, j*.*

Idea de la demostración (para detalles vea, por ejemplo, Gikhman y Skorokhod (1969), páginas 309-310).

(a) Suponga que (12) se cumple para todo i, j. Entonces necesariamente el PMS es estable (porque la ecuación (12) no tiene sentido si $q_i = \infty$) y, además, sumando en (12) sobre todo $j \in S$ se obtiene

$$0 = -q_i + \sum_{k\neq i} q_{ik} = \sum_{k\in S} q_{ik}$$

es decir, el PMS es conservante. Recíprocamente, suponga que el PMS es estable y conservante. Para calcular (12), primero use la ecuación de Chapman–Kolmogorov $\mathrm{P}(t+s) = \mathrm{P}(s)\mathrm{P}(t)$ para obtener

$$\mathrm{P}_{ij}(t+s) - \mathrm{P}_{ij}(t) = [\mathrm{P}_{ii}(s) - 1]\mathrm{P}_{ij}(t) + \sum_{k \neq i} \mathrm{P}_{ik}(s)\mathrm{P}_{kj}(t). \tag{18.15}$$

Multiplicando ambos miembros de esta ecuación por $1/s$ y tomando el límite cuando $s \downarrow 0$, se obtiene (12). Este argumento es válido si el espacio de estados, S, es *finito*, porque entonces la sumatoria que aparece en (15) es finita. Sin embargo, si S es infinito, entonces se debe verificar que efectivamente se pueden intercambiar el límite (cuando $s \downarrow 0$) y la sumatoria en (15).

(b) Análogamente, usando la ecuación de Chapman–Kolmogorov $\mathrm{P}(t+s) = \mathrm{P}(t)\mathrm{P}(s)$, podemos reescribir (15) en la forma

$$\mathrm{P}_{ij}(t+s) - \mathrm{P}_{ij}(t) = \mathrm{P}_{ij}(t)[\mathrm{P}_{jj}(s) - 1] + \sum_{k \neq j} \mathrm{P}_{ik}(t)\mathrm{P}_{kj}(s).$$

Después, multiplicando por $1/s$ y tomando el límite cuando $s \downarrow 0$, de la condición de uniformidad de los límites en (14) se obtiene la ecuación hacia adelante (13). $\square$

Ejemplo 18.5. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ un proceso de Poisson con parámetro λ. Como $\mathrm{X}(\cdot)$ tiene trayectorias no–decrecientes, entonces

$$\mathrm{P}_{ij}(t) = \mathrm{P}[\mathrm{X}(t) = j | \mathrm{X}(0) = i] = 0 \quad \text{si} \quad j < i. \tag{18.16}$$

Por otra parte, para $j \geq i$, las propiedades de $\mathrm{X}(\cdot)$ da que

$$\begin{aligned} \mathrm{P}_{ij}(t) &= \mathrm{P}[\mathrm{X}(t) - X(0) = j - i | \mathrm{X}(0) = i] \\ &= \mathrm{P}[\mathrm{X}(t) = j - i | \mathrm{X}(0) = 0], \end{aligned}$$

de modo que

$$\mathrm{P}_{ij}(t) = e^{-\lambda t}(\lambda t)^{j-i}/(j-i)! \quad \forall\, j \geq i,\ t > 0. \tag{18.17}$$

De (16) y (17), un cálculo directo da que para cualquier estado $i \geq 0$ las intensidades de transición de $\mathrm{X}(\cdot)$ están dadas por

$$q_{ij} = \begin{cases} -\lambda & \text{si} \quad j = i, \\ \lambda & \text{si} \quad j = i+1, \\ 0 & \text{si} \quad j < i \quad \text{ó} \quad j \geq i+2. \end{cases} \tag{18.18}$$

Por lo tanto, la matriz generadora $Q = [q_{ij}]$ de $\mathrm{X}(\cdot)$ es de la forma

$$\begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} & \left[\begin{array}{ccccc} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right] \end{array} \tag{18.19}$$

Es evidente que se satisfacen las hipótesis de ambos incisos, (a) y (b), del Teorema 18.4, así que las dos ecuaciones de Kolmogorov (12) y (13) son válidas. La ecuación hacia atrás resulta

$$\mathrm{P}'_{ij}(t) = -\lambda \mathrm{P}_{ij}(t) + \lambda \mathrm{P}_{i+1,j}(t) \quad \forall\, i, j, \quad \text{con} \quad \mathrm{P}_{ij}(0) = \delta_{ij}.$$

De hecho, como $\mathrm{P}_{ij}(t) = 0$ si $i > j$, se tiene que

$$\mathrm{P}'_{ii}(t) = -\lambda \mathrm{P}_{ii}(t) \quad \forall\, i \geq 0, \tag{18.20}$$

$$\mathrm{P}'_{ij}(t) = -\lambda \mathrm{P}_{ij}(t) + \lambda \mathrm{P}_{i+1,j}(t) \quad \forall\, j-1 \geq i \geq 0 \tag{18.21}$$

con la condición inicial $\mathrm{P}_{ij}(0) = \delta_{ij}$. Análogamente, la ecuación hacia adelante, (13), resulta

$$\mathrm{P}'_{ii}(t) = -\lambda \mathrm{P}_{ii}(t) \quad \forall\, i \geq 0, \tag{18.22}$$

$$\mathrm{P}'_{ij}(t) = -\lambda \mathrm{P}_{ij}(t) + \lambda \mathrm{P}_{i,j-1}(t) \quad \forall\, j-1 \geq i \geq 0, \tag{18.23}$$

con $\mathrm{P}_{ij}(0) = \delta_{ij}$.

En resumen, dadas las probabilidades de transición (16)–(17) de $\mathrm{X}(\cdot)$, se puede calcular la matriz generadora y de ahí las ecuaciones de Kolmogorov hacia atrás y hacia adelante. Recíprocamente, dada la matriz generadora Q en (19), resolviendo la ecuación hacia atrás (20)–(21) ó la ecuación hacia adelante (22)–(23), se obtienen las probabilidades de transición en (17). (Vea el Ejercicio 18.1.) $\square$

Ejemplo 18.6. Sea $\mathrm{X}(t) := \mathrm{X}_0 \cdot (-1)^{N(t)}$ la señal telegráfica aleatoria del Ejemplo 17.13, en donde $N(\cdot)$ es un proceso de Poisson con parámetro λ, y X_0 es una v.a. independiente de $N(\cdot)$, con distribución $\mathrm{P}(\mathrm{X}_0 = i) = 1/2$ para $i = 1, -1$. Para cualquier estado inicial $\mathrm{X}_0 = i$ en $S = \{-1, 1\}$ y cualquier tiempo $t \geq 0$, $\mathrm{X}(t) = i$ si y sólo si $N(t)$ es un número par, es decir, usando la independencia de $\mathrm{X}(0) = \mathrm{X}_0$ y $N(\cdot)$,

$$\mathrm{P}_{ii}(t) = \mathrm{P}_i[\mathrm{X}(t) = i] = \mathrm{P}[N(t) = 2n \quad \text{para algún} \quad n = 0, 1, \ldots].$$

Por lo tanto, como $\cosh(u) = \sum_{n=0}^{\infty} u^{2n}/(2n)! = (e^u + e^{-u})/2$,

$$\mathrm{P}_{ii}(t) = \sum_{n=0}^{\infty} \mathrm{P}[N(t) = 2n] = e^{-\lambda t} \sum_{n=0}^{\infty} (\lambda t)^{2n}/(2n)! = \frac{1}{2}(1 + e^{-2\lambda t}) \quad (18.24)$$

y para $j \neq i$,

$$\mathrm{P}_{ij}(t) = 1 - \mathrm{P}_{ii}(t) = \frac{1}{2}(1 - e^{-2\lambda t}). \quad (18.25)$$

De aquí se sigue que cualquier estado inicial $\mathrm{X}_0 = i$

$$q_{ij} := \mathrm{P}'_{ij}(t)|_{t=0+} = \begin{cases} -\lambda & \text{si} \quad j = i, \\ \lambda & \text{si} \quad j \neq i, \end{cases}$$

es decir la matriz generadora de $\mathrm{X}(\bullet)$ es

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{bmatrix}. \quad (18.26)$$

En el Ejercicio 2 se pide calcular las ecuaciones de Kolmogorov usando Q. $\square$

Ejemplo 18.7. Procesos de nacimiento y muerte. (Compare con el Ejemplo 12.18.) Sea $\mathrm{X}(\bullet) = \{\mathrm{X}(t), t \geq 0\}$ un PMS con espacio de estados $S = \{0, 1, 2, \ldots\}$. Decimos que $\mathrm{X}(\bullet)$ es un proceso de nacimiento y muerte (abreviado: proceso de NM) si estando en el estado i, en caso de que el proceso salte, sólo puede hacerlo a $i+1$ (en caso de un "nacimiento") o a $i-1$ (en caso de una "muerte"); si $i = 0$, el proceso sólo puede saltar al estado 1. Más precisamente, la matriz generadora $Q = [q_{ij}]$ del proceso de NM tiene componentes

$$q_{ij} := \begin{cases} \lambda_i & \text{si} \quad j = i+1 \quad (\text{para} \quad i \geq 0), \\ \mu_i & \text{si} \quad j = i-1 \quad (\text{para} \quad i \geq 1, \quad \text{con} \quad \mu_0 := 0), \\ -(\lambda_i + \mu_i) & \text{si} \quad j = i \quad (\text{para} \quad i \geq 0), \\ 0 & \text{si} \quad |j - i| \geq 2. \end{cases}$$

Las intensidades de transición $q_{i,i+1} = \lambda_i$ se llaman las *intensidades* (o *tasas*) *de nacimiento*, y las $q_{i,i-1} = \mu_i$ son las *intensidades* (o *tasas*) *de muerte*. Si el espacio de estados es *finito*, digamos $S = \{0, 1, \ldots, n\}$, entonces en (27) tomamos $\lambda_i = 0$ para $i \geq n$, y $\mu_i = 0$ para $i \geq n+1$. El PMS en el

Ejercicio 3 es un proceso de NM con espacio de estados $S = \{0, 1\}$. Otros casos especiales de procesos de un NM son los siguientes.

Procesos de puros nacimientos. Se dice que un proceso de NM es de puros nacimientos si $\mu_i = 0$ para todo estado i, lo cual significa que no ocurren muertes. Por ejemplo, un *proceso de Poisson* con parámetro λ es un proceso de puros nacimientos con $\lambda_i = \lambda$ para todo $i \geq 0$.

Otro ejemplo de proceso de puros nacimientos es el *proceso de Yule*, que aparece en biología y física. Este proceso se caracteriza por el hecho de que las tasas de nacimiento λ_i son proporcionales al número de "individuos en la población", es decir,

$$\lambda_i := i\lambda \quad \forall\, i = 0, 1, \ldots, \tag{18.27}$$

en donde λ es una constante positiva.

Por supuesto, también hay *procesos de puras muertes*, en los que $\lambda_i = 0$ para todo i.

Crecimiento lineal con inmigración. Un proceso de NM se dice que es de crecimiento lineal con inmigración si las tasas de nacimiento y muerte son de la forma

$$\lambda_i := i\lambda + a \quad \text{y} \quad \mu_i := i\mu \quad \forall\, i = 0, 1, \ldots, \tag{18.28}$$

donde λ, μ y a son números positivos. El término $a > 0$ en (29) se interpreta como una tasa instantánea de crecimiento de la población debida a alguna causa externa como es la inmigración. Si $a = 0$, se dice que el proceso de NM es **lineal**.

Procesos de ramificación. Dados dos números $q > 0$ y $0 < p < 1$, un proceso de NM con intensidades

$$\lambda_i := iqp \quad \text{y} \quad \mu_i := iq(1-p) \quad \forall\, i = 0, 1, \ldots \tag{18.29}$$

se dice que es un proceso de ramificación. En este caso, $\mathrm{X}(t)$ representa el número de "partículas" o "individuos" de una cierta población presentes en el tiempo t. Dichas partículas actúan independientemente para dar origen a nuevas generaciones de la siguiente manera: Cada partícula tiene un tiempo de vida aleatorio $T \sim \mathrm{Exp}(q)$ al final del cual da origen a dos nuevas partículas con probabilidad p, o bien, la partícula muere (desaparece de la población) con probabilidad $1 - p$. Las intensidades en (30) se obtienen de la fórmula (6), $q_{ij} = q_i p_{ij}$ con $q_i = iq$ y $p_{ij} = p$ si $j = i + 1, p_{ij} = 1 - p$ si $j = i - 1$.

Considérese de nuevo el proceso de NM con intensidades (27). Entonces, por el Teorema 18.3, vemos que la matriz de transición $\mathrm{P} = [p_{ij}]$ de la cadena de saltos tiene componentes

$$p_{ii} = 0,\ p_{i,i+1} = \lambda_i/q_i,\ p_{i,i-1} = \mu_i/q_i \quad \forall\, i \geq 0, \tag{18.30}$$

en donde $q_i = -q_{ii} = \lambda_i + \mu_i$, con $\mu_0 = 0$. Es decir, P es de la forma

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ p_{10} & 0 & p_{12} & 0 & 0 & \cdots \\ 0 & p_{21} & 0 & p_{23} & 0 & \cdots \\ 0 & 0 & p_{32} & 0 & p_{34} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

En el Ejercicio 4 se pide escribir las ecuaciones hacia atrás y hacia adelante de un proceso de NM. $\square$

Ejercicios § 18

18.1. Sea $\mathrm{X}(\cdot)$ el proceso de Poisson del Ejemplo 18.5.

(a) Verifique que las intensidades de transición están dadas por (18).

(b) Resuelva las ecuaciones (20)–(21) y (22)–(23), y en cada caso verifique que $\mathrm{P}_{ij}(t)$ satisface (17). (*Sugerencia*: use inducción y la Observación 16.20)

18.2. En el Ejemplo 18.6:

(a) use la matriz generadora en (26) para obtener las ecuaciones de Kolmogorov hacia atrás y hacia adelante de $\mathrm{X}(\cdot)$;

(a) resuelva alguna de las dos ecuaciones obtenidas en (a) y demuestre que las probabilidades de transición $\mathrm{P}_{ij}(t)$ de $\mathrm{X}(\cdot)$ están dadas por (24) y (25).

18.3. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \geq 0\}$ un PMS con espacio de estados $S = \{0, 1\}$ y cuya matriz generadora Q es

$$\begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

(a) Obtenga la ecuación de Kolmogorov hacia adelante de $\mathrm{X}(\cdot)$.

(b) Usando el resultado en (a) y el hecho de que $\mathrm{P}_{i0}(t) + \mathrm{P}_{i1}(t) = 1$ para $i = 0, 1$, demuestre que para todo $t \geq 0$

$$\mathrm{P}_{ii}(t) = \begin{cases} Ae^{-(\lambda+\mu)t} + B & \text{si} \quad i = 0, \\ Be^{-(\lambda+\mu)t} + A & \text{si} \quad i = 1, \end{cases}$$

en donde $A := \lambda/(\lambda + \mu)$ y $B := 1 - A = \mu/(\lambda + \mu)$.

(c) Demuestre que

$$\pi_0^* := \lim_{t\to\infty} \mathrm{P}_{i0}(t) = B \quad \text{y} \quad \pi_1^* := \lim_{t\to\infty} \mathrm{P}_{i1}(t) = A$$

para $i = 0, 1$.

18.4. Considere el proceso de NM con intensidades de transición q_{ij} en (27). Demuestre que las ecuaciones de Kolmogorov hacia atrás y hacia adelante son, respectivamente,

$$\mathrm{P}'_{ij}(t) = -(\lambda_i + \mu_i)\mathrm{P}_{ij}(t) + \lambda_i \mathrm{P}_{i+1,j}(t) + \mu_i \mathrm{P}_{i-1,j}(t)$$

y

$$\mathrm{P}'_{ij}(t) = -(\lambda_j + \mu_j)\mathrm{P}_{ij}(t) + \lambda_{j-1}\mathrm{P}_{i,j-1}(t) + \mu_{j+1}\mathrm{P}_{i,j+1}(t),$$

para toda $t > 0$, con la condición inicial $\mathrm{P}_{ij}(0) = \delta_{ij}$.

18.5. Demuestre que la solución de la ecuación de Kolmogorov hacia atrás (12) es

$$\mathrm{P}_{ij}(t) = \delta_{ij}e^{-q_i t} + \int_0^t e^{-q_i s} \left[\sum_{k\neq i} q_{ik}\mathrm{P}_{kj}(t-s)\right] ds, \ t \geq 0.$$

A esta ecuación se le llama la **forma integral de la ecuación hacia atrás**. (*Sugerencia*: use la Observación 2 en el párrafo anterior al Ejercicio 16.23.)

18.6. En forma análoga al Ejercicio 18.5, demuestre que la **forma integral de la ecuación hacia adelante** (13) es

$$\mathrm{P}_{ij}(t) = \delta_{ij}e^{-q_j t} + \int_0^t e^{-q_j s} \left[\sum_{k\neq j} \mathrm{P}_{ik}(t-s)q_{kj}\right] ds.$$

18.7. Considere un proceso de puros nacimientos con intensidades λ_i ($i = 0, 1, \ldots$); vea el Ejemplo 18.7. (Observe que $\mathrm{P}_{ij}(t) = 0$ si $j < i$. Explique por qué.) Demuestre:

(a) la ecuación hacia adelante es

$$\mathrm{P}'_{ij}(t) = -\lambda_j \mathrm{P}_{ij}(t) + \lambda_{j-1}\mathrm{P}_{i,j-1}(t);$$

(b) $\mathrm{P}_{ii}(t) = e^{-\lambda_i t} \quad \forall\, t \geq 0$;

(c) $\mathrm{P}_{ij}(t) = \lambda_{j-1} \int_0^t e^{-\lambda_j(t-s)} \mathrm{P}_{i,j-1}(s) ds \quad \forall\, j > i,\ t \geq 0.$

18.8. Considere el proceso de Yule en el Ejemplo 18.7; vea (28). Use el Ejercicio 7(b),(c) para demostrar que

$$\mathrm{P}_{ij}(t) = \begin{pmatrix} j-1 \\ j-i \end{pmatrix} e^{-i\lambda t}(1 - e^{-\lambda t})^{j-i} \quad \forall\, j \geq i,\ t \geq 0.$$

(*Sugerencia*: use inducción en $j \geq i$.)

18.9. (a) Demuestre que un PMS es uniforme si y sólo si $\sup_i q_i < \infty$. (Vea la definición de "uniforme" en el Ejercicio 17.7.)

(b) Dé un ejemplo de un PMS uniforme.

(c) Dé un ejemplo de un PMS estándar que *no* es uniforme. (Vea el Ejercicio 17.7.)

18.10 Considere el proceso de puros nacimientos del Ejemplo 18.7, con parámetros $\lambda_i > 0$. Demuestre que este proceso es uniforme si y sólo si $\sup_i \lambda_i < \infty$. Deduzca que un proceso de Poisson es uniforme.

18.10. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\ t \geq 0\}$ un PMS. Sean i, j dos estados. Se dice que *i se comunica con j* si $\mathrm{P}_{ij}(t) > 0$ para algún $t \geq 0$. Si cualquier estado se comunica con cualquier otro estado, se dice que el PMS es *irreducible.* Sea $\mathrm{X}_\bullet = \{\mathrm{X}_n,\ n = 0, 1, \ldots\}$ la CM encajada en $\mathrm{X}(\cdot)$. Demuestre que $\mathrm{X}(\cdot)$ es irreducible ssi $\mathrm{X}_\bullet$ es irreducible.

18.11. Sean $\mathrm{X}(\cdot)$ y $\mathrm{X}_\bullet$ como en el Ejercicio 18.10, y sea τ_1 el tiempo del primer salto de $\mathrm{X}(\cdot)$. La v.a.

$$T_i := \inf\{t > \tau_1,\ \mathrm{X}(t) = i | \mathrm{X}(0) = i\}$$

se llama el *tiempo del primer retorno* al estado i. Se dice que i es un estado *recurrente* (*transitorio*) para el PMS $\mathrm{X}(\cdot)$ si

$$\mathrm{P}_i(T_i < \infty) = 1 \ (< 1).$$

Demuestre que i es recurrente (transitorio) ssi i es recurrente (transitorio) para la CM encajada $\mathrm{X}_\bullet$.

18.12. En cada uno de los Ejemplos 18.5, 18.6 y 18.7 diga si el correspondiente PMS es (a) irreducible. (b) Si el PMS es irreducible, diga si es recurrente o transitorio. **Nota.** A diferencia de los conceptos definidos en los Ejercicios 18.10 y 18.11, la recurrencia positiva y la recurrencia nula en un PMS no se pueden definir en términos de la CM encajada. Esto se debe a que la recurrencia positiva en un PMS depende de los tiempos de saltos $\{\tau_n\}$, así que la CM encajada no es suficiente para definir recurrencia positiva. (Hay ejemplos en los que un PMS es recurrente positivo pero la CM encajada es recurrente nula.) Sea T_i el tiempo del primer retorno, definido en el Ejercicio 18.11. Un estado i es *recurrente positivo* (*recurrente nulo*) en el PMS si

$$E_i(T_i) < \infty \ (= \infty).$$

Nótese la analogía entre esta definición y la definición de recurrencia positiva y recurrencia nula en la Sección 13.

19 Comportamiento asintótico de un PMS

Contenido: Distribución límite, distribución invariante (alias: distribución estacionaria), recurrencia.

En las secciones 12, 13 y 14 estudiamos varios conceptos importantes sobre el comportamiento de una cadena de Markov (CM) a tiempo discreto, por ejemplo, distribuciones invariantes, recurrencia y transitoriedad. Ahora veremos como estos conceptos se pueden extender a un PMS $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$.

Para cada $t \geq 0$, sea $\pi(t) = \{\pi_j(t),\, j \in S\}$ la **distribución** de $\mathrm{X}(t)$, es decir,

$$\pi_j(t) = \mathrm{P}[\mathrm{X}(t) = j] \quad \forall\, j \in S. \tag{19.1}$$

Equivalentemente, como $\mathrm{P}[\mathrm{X}(t) = j] = \sum_{i \in S} \mathrm{P}[\mathrm{X}(0) = i,\, \mathrm{X}(t) = j]$, tenemos que

$$\pi_j(t) = \sum_{i \in S} \pi_i(0) \mathrm{P}_{ij}(t) \quad \forall\, j \in S. \tag{19.2}$$

Además, interpretando $\pi(t)$ como un *vector fila*, podemos expresar (2) en forma abreviada, matricial, como

$$\pi(t) = \pi(0)\mathrm{P}(t), \tag{19.3}$$

en donde $\mathrm{P}(t) = [\mathrm{P}_{ij}(t)]$ es la matriz de transición del PMS. La distribución $\pi(t)$ también se puede escribir como solución de una cierta ecuación diferencial si suponemos que se satisface la ecuación hacia adelante (18.18.11), que aquí repetimos:

$$\mathrm{P}'(t) = \mathrm{P}(t)Q \quad \forall\, t > 0, \quad \text{con} \quad \mathrm{P}(0) = I. \tag{19.4}$$

En particular, *bajo las hipótesis del Teorema 18.4(b)*, de (3) y (4) vemos que

$$\pi'(t) = \pi(0)\mathrm{P}'(t) = \pi(0)\mathrm{P}(t)Q = \pi(t)Q,$$

i.e.

$$\pi'(t) = \pi(t)Q. \tag{19.5}$$

Término a término, la ecuación (5) resulta

$$\pi_j'(t) = \sum_{i\in S} \pi_i(t) q_{ij} = -\pi_j(t) q_j + \sum_{i\neq j} \pi_i(t) q_{ij} \quad \forall\, j \in S. \tag{19.6}$$

En resumen, las ecuaciones (1), (2), (3), (5) y (6) son distintas representaciones de $\pi(t)$.

Definición 19.1. Sea $\pi^* = \{\pi_i^*,\, i \in S\}$ una distribución de probabilidad sobre S. Se dice que π^* es

(a) la **distribución límite** del PMS si se cumple que

$$\lim_{t\to\infty} \mathrm{P}_{ij}(t) = \pi_j^* \quad \forall\, i, j \in S; \tag{19.7}$$

(b) una **distribución invariante** (o **distribución estacionaria**) del PMS si

$$\pi^* = \pi^* \mathrm{P}(t) \quad \forall\, t \geq 0, \tag{19.8}$$

i.e.

$$\pi_j^* = \sum_{i\in S} \pi_i^* \mathrm{P}_{ij}(t) \quad \forall\, t \geq 0,\ j \in S. \tag{19.9}$$

Nótese que estos conceptos son análogos a los correspondientes para CMs. Por ejemplo, *si π^* es una distribución invariante y $\pi(0) = \pi^*$, entonces*

$$\pi(t) = \pi^* \quad \forall\, t \geq 0. \tag{19.10}$$

Esto se sigue de (8) y (3). Asimismo, se puede demostrar que *si la distribución límite π^* existe, entonces π^* es la única distribución invariante del PMS.* (Vea el Ejercicio 19.1.)

Como ejemplo, la distribución $\pi^* = (\pi_0^*, \pi_1^*)$ en el Ejercicio 18.3(c) es la distribución límite (y por lo tanto, la única distribución invariante) del PMS con espacio de estados $S = \{0, 1\}$ y matriz generadora

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}. \tag{19.11}$$

En el resto de esta sección veremos que la matriz generadora de un PMS está muy relacionada con la existencia de distribuciones invariantes y de la distribución límite, como en el siguiente resultado.

Proposición 19.2. Supóngase que X(•) es un PMS estándar, continuo por la derecha y regular, y además, X(•) satisface la Hipótesis 18.2 (el PMS es estable y conservante). Sea $\pi^* = \{\pi_i^*,\, i \in S\}$ una distribución de probabilidad sobre S, y $\hat{\mu} = \{\hat{\mu}_i,\, i \in S\}$ el vector fila definido como $\hat{\mu}_i := \pi_i^* q_i$. Entonces las siguientes condiciones son equivalentes:

(a) π^* es una distribución estacionaria del PMS.

(b) π^* satisface que
$$\pi^* Q = 0 \tag{19.12}$$

(c) $\hat{\mu}\mathrm{P} = \hat{\mu}$, en donde $\mathrm{P} = [p_{ij}]$ es la matriz de transición de la CM encajada en X(·) y se cumple $\sum_{i \in S} \hat{\mu}_i < \infty$.

Nota 19.3. Si se cumple la ecuación hacia adelante (4), entonces la implicación "(a)⇒(b)" es trivial; vea las ecuaciones (5) y (10).

Demostración. Bajo las presentes hipótesis se satisface la ecuación hacia atrás
$$\mathrm{P}'(t) = Q\mathrm{P}(t) \quad \forall\, t > 0,\ \mathrm{P}(0) = I. \tag{19.13}$$
Para demostrar la equivalencia de (a) y (b), supóngase primero que π^* es una distribución estacionaria. Luego, por (8), $\pi^*\mathrm{P}(t)$ es el vector constante π^* cuya derivada, por supuesto, es cero. Por lo tanto, premultiplicando (13) por π^* obtenemos

$$0 = (\pi^*\mathrm{P}(t))' = \pi^*\mathrm{P}'(t) = \pi^* Q\mathrm{P}(t) \quad \forall\, t > 0$$

lo cual implica (12); es decir, (a) implica (b). Recíprocamente, si (12) se cumple, de nuevo vemos que la derivada de $\pi^*\mathrm{P}(t)$ es cero, así que $\pi^*\mathrm{P}(t)$ es una constante. Es decir, para cada estado $j \in S$ existe una constante c_j tal que
$$(\pi^*\mathrm{P}(t))_j = \sum_{i \in S} \pi_i^* \mathrm{P}_{ij}(t) = c_j \quad \forall\, t \geq 0.$$
En particular, en $t = 0$ se tiene $\mathrm{P}_{ij}(0) = \delta_{ij}$ y se sigue de $\pi_j^* = c_j$. Esto significa que (b) implica (a).

Para demostrar la equivalencia de (b) y (c), reescribiremos la ecuación (18.18.6) en la forma

$$q_{ij} = q_i(p_{ij} - \delta_{ij}) \quad \forall\, i, j \in S.$$

De esta ecuación y de la definición de $\hat{\mu}$, vemos que la j–ésima componente de $\hat{\mu}(\mathrm{P}-I)$ es

$$\sum_i \hat{\mu}_i(p_{ij}-\delta_{ij}) = \sum_i \pi_i^* q_i(p_{ij}-\delta_{ij}) = \sum_i \pi_i^* q_{ij} \quad \forall\, j,$$

es decir $\hat{\mu}(\mathrm{P}-I) = \pi^* Q$. Esto da que $\hat{\mu}(\mathrm{P}-I) = 0$ si y sólo si $\pi^* Q = 0$. $\square$

Observación 19.4. La Proposición 19.2 da *dos* formas para tratar de encontrar una distribución estacionaria π^* de un PMS. Una de ellas es encontrar una distribución de probabilidad π^* que satisface la ecuación (12), en donde Q es la matriz generadora del PMS. La otra forma usa la matriz de transición P de la cadena de saltos del PMS; se encuentra una solución no–nula $\hat{\mu}$, con componentes no–negativas, de la ecuación $\hat{\mu} = \hat{\mu}\mathrm{P}$, y se toma

$$\pi_i^* := \hat{\mu}_i / q_i \quad \forall\, i \in S. \tag{19.14}$$

Esto define una distribución estacionaria del PMS si se normaliza de tal forma que $\sum_i \pi_i^* = 1$. Asimismo, si $\sum_i \hat{\mu}_i = 1$, entonces $\hat{\mu}$ es una distribución estacionaria para la CM encajada en $\mathrm{X}(\cdot)$. En otras palabras, (14) establece una relación entre una distribución invariante π^* para el PMS y una distribución invariante $\hat{\mu}$ para la cadena encajada.

Ejemplo 19.5. Considérese un PMS con espacio de estados $S = \{0, 1\}$ y la matriz generadora Q en (11). Deseamos encontrar una distribución estacionaria $\pi^* = (\pi_0^*, \pi_1^*)$. Si intentamos resolver la ecuación $\pi^* Q = 0$, vemos que

$$-\lambda\pi_0^* + \mu\pi_1^* = 0 \quad \text{y} \quad \lambda\pi_0^* - \mu\pi_1^* = 0.$$

Por lo tanto, usando que $\pi_0^* + \pi_1^* = 1$, obtenemos la distribución estacionaria

$$\pi_0^* = \mu/(\lambda+\mu), \quad \pi_1^* = \lambda/(\lambda+\mu). \tag{19.15}$$

(Compare con el resultado en el Ejercicio 18.3(c).) Por otra parte, si deseamos encontrar una distribución estacionaria como en (14), primero vemos que la matriz de transición de la cadena encajada es (por la ecuación (18.18.6))

$$\mathrm{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

De aquí se sigue que cualquier vector $\hat{\mu} = (\hat{\mu}_0, \hat{\mu}_1)$ con componentes $\hat{\mu}_0 = \hat{\mu}_1$ satisface la ecuación $\hat{\mu}\mathrm{P} = \hat{\mu}$. En particular, si $\hat{\mu}_0 > 0$, vemos de (14) que

$$\pi_0^* = \hat{\mu}_0/\lambda, \quad \pi_1^* = \hat{\mu}_0/\mu. \tag{19.16}$$

Además, para tener la igualdad $\pi_0^* + \pi_1^* = 1$, necesariamente $\hat{\mu}_0 = \lambda\mu/(\lambda+\mu)$. Sustituyendo este valor de $\hat{\mu}_0$ en (16), se obtiene la distribución estacionaria (15). $\Box$

Ejemplo 19.6. Considérese el proceso de Poisson $\mathrm{X}(\cdot)$ del Ejemplo 18.5, cuya matriz generadora Q está dada por (18.18.19), i.e.

$$\begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

En este caso, la ecuación $\pi^* Q = 0$ da que

$$-\lambda\pi_0^* = 0 \quad \text{y} \quad \lambda\pi_{i-1}^* - \lambda\pi_i^* = 0 \quad \forall\, i = 1, 2, \ldots,$$

es decir, $\pi_0^* = 0$ y $\pi_i^* = \pi_{i-1}^*$ para todo $i = 0, 1, \ldots$. De aquí se sigue que la única solución de la ecuación $\pi^* Q = 0$ es el vector "cero" $\pi^* = (0, 0, \ldots)$, así que $\mathrm{X}(\cdot)$ *no* tiene una distribución estacionaria. (Este resultado es obvio porque un proceso de Poisson tiene trayectorias no–decrecientes.) Análogamente, como la matriz de transición P de la cadena encajada es

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

la única solución de la ecuación $\hat{\mu}\mathrm{P} = \hat{\mu}$ es el vector cero, $\hat{\mu} = (0, 0, \ldots)$. Luego, de (14) obtenemos de nuevo $\pi_i^* = \hat{\mu}_i/q_i = 0$ para todo estado i. $\Box$

Ejemplo 19.7. Considérese el proceso de NM con intensidades de nacimiento y muerte en (18.18.27), de modo que la matriz generadora $Q = [q_{ij}]$ es de la forma

$$\begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -q_1 & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -q_2 & \lambda_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{19.17}$$

con $q_i := \lambda_i + \mu_i$ para todo $i = 0, 1, \ldots, y\, \mu_0 := 0$. Deseamos encontrar condiciones sobre λ_i, μ_i para la existencia de distribuciones invariantes π. Con este fin, consideramos la ecuación $\pi Q = 0$ que más explícitamente resulta

$$\pi_1 \mu_1 = \pi_0 \lambda_0,$$

$$\pi_j(\lambda_j + \mu_j) = \pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} \quad \forall\, j = 1, 2, \ldots$$

De estas ecuaciones se obtiene (por inducción) que $\pi_j \mu_j = \pi_{j-1}\lambda_{j-1}$, o equivalentemente

$$\pi_j = \pi_{j-1}(\lambda_{j-1}/\mu_j) \quad \forall\, j = 1, 2, \ldots$$

Iterando esta relación se obtiene que

$$\pi_j = \beta(j)\pi_0 \quad \forall\, j = 0, 1, \ldots \tag{19.18}$$

en donde $\beta(0) := 1$ y

$$\beta(j) := (\lambda_0 \cdots \lambda_{j-1})/(\mu_1 \cdots \mu_j) \quad \text{para} \quad j = 1, 2, \ldots \tag{19.19}$$

Además, definiendo

$$\overline{\beta} := \sum_{j=0}^{\infty} \beta(j) = 1 + \sum_{j=1}^{\infty} \beta(j),$$

vemos que

$$\sum_{j=0}^{\infty} \pi_j = \overline{\beta}\pi_0.$$

De aquí se sigue que (18) *define una distribución estacionaria si y sólo si* $\overline{\beta} < \infty$ *y, además* $\pi_0 = 1/\overline{\beta}$. En este caso, por (18), la distribución estacionaria está dada por

$$\pi_j = \beta(j)/\overline{\beta}. \tag{19.20}$$

En el Ejercicio 2 se pide demostrar un resultado similar pero usando la ecuación $\hat{\mu}\mathrm{P} = \hat{\mu}$ en la Proposición 19.2. (En el Ejercicio 19.3 se ve el caso de intensidades constantes $\lambda_i = \lambda$, $\mu_i = \mu$.) $\square$

La Proposición 19.2 da condiciones necesarias y suficientes para que un PMS tenga una distribución estacionaria. Sin embargo, la existencia de distribuciones límites es más complicada en el sentido de que prácticamente sólo existen condiciones suficientes. Este es el caso de la siguiente proposición (que se demuestra, por ejemplo, en el libro de Heyman y Sobel (1982), página 303, Teorema 8–5).

Proposición 19.8. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ un PMS estándar, continuo por la derecha y regular, y suponga que $\mathrm{X}_\bullet = \{\mathrm{X}_n,\, n = 0, 1, \ldots\}$ la CM encajada en $\mathrm{X}(\cdot)$ es recurrente positiva. Además supóngase que

(a) $\mathrm{X}(\cdot)$ es estable (es decir, $q_j < \infty$ para todo $j \in S$);

(b) $\mathrm{X}_\bullet$ es irreducible (lo cual implica que $q_j > 0$ para todo $j \in S$);

(c) $\mathrm{X}_\bullet$ tiene una única distribución invariante $\hat{\pi} = \{\hat{\pi}_j,\, j \in S\}$; y

(d) $M := \sum_{j \in S} (\hat{\pi}_j / q_j) < \infty$.

Entonces la distribución límite $\pi^* = \{\pi_j^*,\, j \in S\}$ en (7) existe y está dada por

$$\pi_j^* = \hat{\pi}_j / M q_j \quad \forall\, j \in S. \tag{19.21}$$

Por el Teorema 14.5, las condiciones (b) y (c) en la proposición anterior equivalen a pedir que la CM $\mathrm{X}_\bullet$ sea recurrente positiva. Nótese, por otra parte, que la distribución dada por la ecuación (21) equivale a normalizar los números en (14) para que su suma total sea 1.

Como ya mencionamos, si la distribución límite π^* existe, entonces π^* es la única distribución invariante. El recíproco de ésto se cumple bajo hipótesis adecuadas, como en la siguiente proposición.

Proposición 19.9. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ un PMS estándar, continuo por la derecha y regular, y sea $\mathrm{X}_\bullet = \{\mathrm{X}_n,\, n = 0, 1, \ldots\}$ la CM encajada en $\mathrm{X}(\cdot)$. Supóngase que

(a) $\mathrm{X}(\cdot)$ es irreducible (lo cual significa que la CM $\mathrm{X}_\bullet$ es irreducible); y

(b) $\mathrm{X}(\cdot)$ tiene una distribución estacionaria $\pi = \{\pi_j,\, j \in S\}$ tal que

$$C := \sum_{j \in S} \pi_j q_j < \infty. \tag{19.22}$$

Entonces π es la *única* distribución estacionaria de $\mathrm{X}(\cdot)$ y, además, π es la distribución límite de $\mathrm{X}(\cdot)$.

Demostración. Para cada $j \in S$, sean

$$\mu_j := \pi_j q_j \quad \text{y} \quad \hat{\pi}_j := \mu_j / C = \pi_j q_j / C.$$

con C como en (22). Nótese que $C = \sum_{j\in S} \mu_j$ y por lo tanto $\hat{\pi} = \{\hat{\pi}_j,\, j \in S\}$ es una distribución de probabilidad sobre S. De hecho, como π es una distribución estacionaria para $\mathrm{X}(\cdot)$, se sigue de la Proposición 19.2 que $\hat{\pi}$ es una distribución estacionaria para la CM $\mathrm{X}_\bullet$. Además,

$$M := \sum_j (\hat{\pi}_j/q_j) = \frac{1}{C}\sum_j \pi_j = \frac{1}{C} < \infty,$$

y de aquí se sigue que $\mathrm{X}(\cdot)$ y $\mathrm{X}_\bullet$ satisfacen las hipótesis de la Proposición 19.8. Por lo tanto, la ecuación (21) resulta

$$\pi_j^* = \hat{\pi}_j/Mq_j = \pi_j \quad \forall\, j \in S;$$

es decir, $\pi^* = \pi$ es la distribución límite de $\mathrm{X}(\cdot)$ y, en consecuencia, es la única distribución estacionaria. □

Ejemplo 19.10. Sea $\mathrm{X}(\cdot)$ el proceso de NM en el Ejemplo 19.7, con parámetros λ_i, μ_i positivos (excepto $\mu_0 := 0$), y sea $\beta(j)$ como en (19). Si

$$\overline{\beta} := 1 + \sum_{j=1}^{\infty} \beta(j) < \infty,$$

sabemos que $\mathrm{X}(\cdot)$ tiene la distribución estacionaria $\pi_j = \beta(j)\pi_0 = \beta(j)/\overline{\beta}$ en (20). Supongamos ahora que

$$\sum_{j=1}^{\infty} \beta(j)\lambda_j < \infty. \tag{19.23}$$

Entonces, como $q_j = \lambda_j + \mu_j$, se puede ver que

$$C := \sum_{j=0}^{\infty} \pi_j q_j = \frac{1}{\overline{\beta}}\left[2\lambda_0 + 2\sum_{j=1}^{\infty} \beta(j)\lambda_j\right] < \infty$$

y, por lo tanto, de la Proposición 19.9 concluimos que, bajo la hipótesis (23), $\pi = \{\pi_j\}$ es la distribución límite de $\mathrm{X}(\cdot)$. Esta misma conclusión se obtiene si en lugar de la Proposición 19.9 usamos la 19.8 y el Ejercicio 19.2. □

Ejercicios § 19

19.1. Demuestre que (3) y la ecuación de Chapman–Kolmogorov implican que

$$\pi(s+t) = \pi(s)\mathrm{P}(t) \quad \forall\, s,t \geq 0.$$

Tomando el límite cuando $s \to \infty$, concluya que si π^* es la distribución límite del PMS (suponiendo que tal distribución existe), entonces π^* es la única distribución invariante.

19.2. Considere el proceso de NM $\mathrm{X}(\cdot)$ en el Ejemplo 19.7.

(a) Calcule la matriz de transición $\mathrm{P} = [\mathrm{P}_{ij}]$ de la cadena encajada $\mathrm{X}_\bullet = \{\mathrm{X}_n\}$.

(b) Demuestre que $\mathrm{X}_\bullet$ tiene una distribución estacionaria $\hat{\pi}$ si y sólo si

$$\alpha := \sum_{j=0}^{\infty} \alpha(j) < \infty,$$

en donde

$$\alpha(j) := \lambda_0^{-1}\beta(j)q_j, \quad \text{con} \quad \beta(j) \quad \text{como en (19);}$$

además, $\hat{\pi} = \{\hat{\pi}_j,\, j \in S\}$ está dada por

$$\hat{\pi}_j = \alpha(j)/\alpha \quad \forall\, j = 0,1,\ldots.$$

(Compare este resultado con el obtenido en el Ejemplo 12.18.)

(c) Del inciso (b) (junto con (14) o con la Proposición 19.2) concluya que si $\alpha < \infty$, entonces el proceso $\mathrm{X}(\cdot)$ tiene la distribución estacionaria $\pi = \{\pi_j\}$ en (20).

19.3. Proceso de NM con intensidades constantes. Suponga que el proceso de NM $\mathrm{X}(\cdot)$ en el Ejemplo 19.7 tiene intensidades de nacimiento y muerte *constantes*

$$\lambda_i \equiv \lambda > 0 \quad \text{y} \quad \mu_i \equiv \mu > 0$$

para todo $i = 0,1,\ldots$, excepto $\mu_0 := 0$. Demuestre que $\mathrm{X}(\cdot)$ tiene una distribución invariante $\pi = \{\pi_j\}$ si y sólo si $\lambda < \mu$, en cuyo caso

$$\pi_j = (1-\rho)\rho^j \quad \forall\, j = 0,1,\ldots, \quad \text{con} \quad \rho := \lambda/\mu.$$

19.4. Considere el PMS con dos estados, 0 y 1, del Ejemplo 19.5 y el Ejercicio 18.3.

(a) Calcule la distribución $\pi(t) = (\pi_0(t), \pi_1(t))$ de $\mathrm{X}(t)$ para $t \geq 0$.

(b) Calcule $E\mathrm{X}(t)$ y demuestre que $\lim_{t\to\infty} E\mathrm{X}(t) = \pi_1^*$.

(c) Sea $b(t) := t^{-1} \int_0^t E[\mathrm{X}(s)]ds$ y demuestre que $\lim_{t\to\infty} b(t) = \pi_1^*$. (Observe que $b(t)$ es la fracción esperada de tiempo durante el intervalo $[0, t]$ en la que $\mathrm{X}(\cdot)$ toma el valor 1).

19.5. En el Ejercicio 19.4 supóngase que $\lambda = \mu$, y sea $N(t)$ el número de transiciones de $\mathrm{X}(\bullet)$ hasta el tiempo t; vea la ecuación (17.17.15). Demuestre que $N(t) \sim \mathrm{Poi}(\lambda t)$, i.e.

$$\mathrm{P}[N(t) = n] = e^{-\lambda t}(\lambda t)^n/n! \quad \forall\, n = 0, 1, \ldots.$$

19.6. Considere un sistema telegráfico que consiste de dos cables que funcionan independientemente, y cada uno de los cuales sólo puede procesar un mensaje a la vez. El tiempo que cada cable permanece operando hasta que se descompone es una v.a. con distribución exponencial $\mathrm{Exp}(\lambda)$. El tiempo de reparación de cada cable tiene distribución $\mathrm{Exp}(\mu)$. Para cada $t \geq 0$, sea $\mathrm{X}(t)$ el número de cables operando al tiempo t, de modo que $\mathrm{X}(\bullet) = \{\mathrm{X}(t),\, t \geq 0\}$ es un PE con espacio de estados $S = \{0, 1, 2, \}$.

(a) Demuestre que $\mathrm{X}(\bullet)$ es un PMS cuya matriz generadora Q es de la forma

$$\begin{bmatrix} -2\mu & 2\mu & 0 \\ \lambda & -(\lambda+\mu) & \mu \\ 0 & 2\lambda & -2\lambda \end{bmatrix}$$

(b) Calcule la distribución estacionaria π^* para $\mathrm{X}(\bullet)$, si es que existe.

(c) Describa la CM $\mathrm{X}_\bullet$ encajada en $\mathrm{X}(\bullet)$ y encuentre su distribución estacionaria $\hat{\pi}$, si es que existe. Diga cual es la relación entre $\hat{\pi}$ y π^*.

19.7. Considérese un proceso de NM con intensidades

$$\lambda_i := \alpha/(i+1) \quad \text{y} \quad \mu_i := \mu \quad \forall\, i = 0, 1, \ldots \quad (\text{excepto } \ \mu_0 := 0),$$

en donde α y μ son constantes positivas. Encuentre condiciones bajo las cuales el proceso tiene una distribución estacionaria π, y en tal caso calcule

π. (En teoría de colas, a un proceso de esta forma se le llama un sistema de "arribos desanimados" debido a que la tasa de arribos λ_i disminuye cuando i crece.)

19.8. Sea $X(\cdot) = \{X(t),\, t \geq 0\}$ un proceso de puros nacimientos en el que

$$P\,[\text{un evento ocurre en } (t, t+h)|X(t) = \text{ impar}] = \lambda_1 h + o(h),$$

$$P\,[\text{un evento ocurre en } (t, t+h)|X(t) = \text{ par}] = \lambda_2 h + o(h).$$

Suponga que $X(0) = 0$. Calcule las probabilidades

$$p_1(t) := P[X(t) = \text{ impar}], \quad p_2(t) := P[X(t) = \text{ par}].$$

[*Sugerencia:* Deduzca las ecuaciones diferenciales

$$p_1'(t) = -\lambda_1 p_1(t) + \lambda_2 p_2(t), \quad p_2'(t) = \lambda_1 p_1(t) - \lambda_2 p_2(t)$$

y resuélvalas.]

19.9. Sea $X(\cdot) = \{X(t),\, t \geq 0\}$ un proceso de puras muertes con $\mu_i = i\mu$ para $i = 0, 1, \ldots$. Supóngase que $X(0) = i > 0$. Calcule $\pi_j(t) := P[X(t) = j]$ y demuestre que

$$EX(t) = ie^{-\mu t} \quad \text{y} \quad \text{Var}[X(t)] = ie^{-\mu t}(1 - e^{-\mu t}).$$

19.10. Repita el Ejercicio 9 en el caso en el que $X(\cdot)$ es el proceso de ramificación con intensidades λ_i, μ_i en la ecuación (18.18.28).

19.11. Calcule mediante dos métodos distintos la distribución del proceso de NM con parámetros $\lambda_i := \lambda p^i$ y $\mu_i := \mu$, en donde λ, μ y p son números positivos, con $p < 1$.

19.12. Sea $X(\cdot) = \{X(t),\, t \geq 0\}$ el proceso de crecimiento lineal con inmigración con intensidades en (18.18.28). Para cada $t \geq 0$, sea $\pi(t) = \{\pi_j(t),\, j = 0, 1, \ldots\}$ la distribución de $X(t)$, y sea $m(t) := EX(t)$.

(a) Demuestre que $m'(t) = a + (\lambda - \mu)m(t)$. (*Sugerencia:* use (5)–(6).)

(b) Resuelva la ecuación diferencial en (a) y describa qué ocurre con $m(t)$ cuando $t \to \infty$ en cada uno de los casos $\lambda = \mu$, $\lambda > \mu$, y $\lambda < \mu$.

19.13. Del Ejercicio 19.2 y la Proposición 19.8 obtenga condiciones suficientes para que la distribución definida por (20) sea la distribución límite del PMS en el Ejemplo 19.7. (Compare con el Ejemplo 19.10.)

19.14. Considere un PMS con espacio de estados S y cuya matriz generadora Q es estable y conservante (vea la Hipótesis 18.2). Sea π una distribución de probabilidad sobre S. Se dice que Q y π **están en balance** si

$$\pi_i q_{ij} = \pi_j q_{ji} \quad \forall\, i,\, j \in S.$$

Demuestre que si Q y π están en balance, entonces π es una distribución invariante para el PMS. (Compare con el Ejercicio 12.5.)

20 Procesos estacionarios

Contenido: Proceso de segundo orden, proceso estacionario, teoremas espectrales, teoremas ergódicos.

A. Conceptos básicos

Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\}$ un PE sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$. De acuerdo con la Definición 16.7(a), $\mathrm{X}(\cdot)$ es un **proceso de segundo orden** si $\mathrm{X}(t) \in L_2$ para todo $t \in T$. En este caso también se dice que $\mathrm{X}(\cdot)$ es un PE en L_2 y escribimos $\mathrm{X}(\cdot) \in L_2$.

En esta sección introduciremos algunos conceptos básicos de procesos en L_2 que nos permitirán definir un cálculo diferencial e integral en L_2.

Para empezar, nótese que la función de covarianza

$$K_{\mathrm{X}}(s,t) := \mathrm{Cov}\,(\mathrm{X}(s), \mathrm{X}(t))$$

es una función **simétrica**, es decir, $K_{\mathrm{X}}(s,t) = K_{\mathrm{X}}(t,s)$. Asimismo, del Ejercicio 1(b) se puede ver que K_{X} es una función **definida no–negativa**, es decir

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k K_{\mathrm{X}}(t_j, t_k) \geq 0$$

para todo $n \geq 1$, $a_1, \ldots, a_n$ en $\mathbb{R}$, y $t_1, \ldots, t_n$ en T.

Definición 20.1. Se dice que dos PEs $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\})$ y $Y(\cdot) = \{Y(t),\, t \in T\}$ **tienen las mismas distribuciones finito–dimensionales** si para cualquier conjunto finito de índices $t_1, \ldots, t_n$ en T, las vv.aa. $\mathrm{X}(t_1), \ldots, \mathrm{X}(t_n)$ tienen la misma distribución conjunta que las vv.aa. $Y(t_1), \ldots, Y(t_n)$.

Por ejemplo, si $\mathrm{X}(\cdot)$ y $Y(\cdot)$ son dos procesos *gaussianos* y tienen la misma función media y la misma función de covarianza, entonces $\mathrm{X}(\cdot)$ y $Y(\cdot)$ tienen las mismas distribuciones finito–dimensionales. (Esto no necesariamente se cumple si al menos uno de los PEs $\mathrm{X}(\cdot)$ ó $Y(\cdot)$ no es gaussiano.)

Definición 20.2. (a) Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in \mathbb{R}\}$ un PE (no necesariamente en L_2), y para cada $h \in \mathbb{R}$ fijo sea $Y_h(\cdot) = \{Y_h(t),\, t \in \mathbb{R}\}$ el PE dado por

$$Y_h(t) := \mathrm{X}(t+h). \tag{20.1}$$

Decimos que $\mathrm{X}(\cdot)$ es **estrictamente estacionario** si para cualquier $h \in \mathbb{R}$ los PEs $\mathrm{X}(\cdot)$ y $Y_h(\cdot)$ tienen las mismas distribuciones finito–dimensionales.

(b) Se dice que un PE $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in \mathbb{R}\}$ en L_2 es L_2**-estacionario** o **débilmente estacionario** si satisface que:

(b$_1$) su función media $m_{\mathrm{X}}(t) := E\mathrm{X}(t)$ es constante, i.e.

$$m_{\mathrm{X}}(t) \equiv m_{\mathrm{X}} \quad \forall\, t \in \mathbb{R}; \tag{20.2}$$

(b$_2$) su función de covarianza $K_{\mathrm{X}}(s,t)$ depende sólo de $t-s$, es decir

$$K_{\mathrm{X}}(s,t) = K_{\mathrm{X}}(0, t-s) \quad \forall\, s,t \in \mathbb{R}. \tag{20.3}$$

Denotaremos con $r_{\mathrm{X}}(t-s)$ el lado derecho de (3), i.e., para toda $s,t \in \mathbb{R}$,

$$r_{\mathrm{X}}(t) := K_{\mathrm{X}}(0,t) = K_{\mathrm{X}}(s, s+t). \tag{20.4}$$

si $\mathrm{X}(\cdot)$ es L_2–estacionario.

De las Definiciones 16.3 y 16.4 (vea también el Ejemplo 16.8) es evidente que el proceso de Poisson y el proceso de Wiener no son ni estrictamente ni débilmente estacionarios. Tenemos, por otra parte, el siguiente resultado general.

Proposición 20.3. Si $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in \mathbb{R}\}$ está en L_2 y es estrictamente estacionario, entonces es L_2–estacionario.

Demostración. Por (1), $\mathrm{X}(t)$ y $\mathrm{X}(t+h)$ tienen la misma distribución para todo t y h en $\mathbb{R}$. Esto implica que, en particular, tales vv.aa. tienen la misma media, i.e.

$$m_{\mathrm{X}}(t) = m_{\mathrm{X}}(t+h) \quad \forall\, t,h \in \mathbb{R}.$$

De aquí se sigue que $m_{\mathrm{X}}(t)$ es constante. Análogamente, para cualquier s,t,h

$$\begin{aligned} E[\mathrm{X}(s)\mathrm{X}(t)] &= \int\int xy\ dF_{\mathrm{X}(s),\mathrm{X}(t)}(x,y) \\ &= \int\int xy\ dF_{\mathrm{X}(s+h),\mathrm{X}(t+h)}(x,y) \\ &= E[\mathrm{X}(s+h)\mathrm{X}(t+h)]. \end{aligned}$$

En particular, tomando $h = -s$ se obtiene $E[\mathrm{X}(s)\mathrm{X}(t)] = E[\mathrm{X}(0)\mathrm{X}(t-s)]$. $\square$

Observación 20.4. (a) El recíproco de la Proposición 20.3 es **falso**, en general. Por ejemplo, sean $\mathrm{X}(t)$ vv.aa. independientes tales que

$$\mathrm{X}(t) \sim \begin{cases} N(0,1) & \text{si } t \in \mathbb{Z} \text{ y } t \geq 0, \\ \mathrm{Uni}[-a,a] & \text{si } t \in \mathbb{Z} \text{ y } t < 0, \end{cases}$$

donde $a := \frac{1}{2}\sqrt{12}$. Nótese que $\mathrm{X}(\cdot)$ está en L_2 y tiene media $m_{\mathrm{X}}(t) = 0$ para todo $t \in \mathbb{Z}$, y además,

$$K_{\mathrm{X}}(s,t) = E[\mathrm{X}(s)\mathrm{X}(t)] = \begin{cases} 0 & \text{si } s \neq t, \\ 1 & \text{si } s = t. \end{cases}$$

Es decir, la función de covariancia es de la forma $K_{\mathrm{X}}(s,t) = K_{\mathrm{X}}(0, t-s)$ lo cual significa que $\mathrm{X}(\cdot)$ es un proceso L_2–estacionario. Sin embargo, es obvio que $\mathrm{X}(\cdot)$ no es estrictamente estacionario porque la distribución de $\mathrm{X}(t)$ depende de t. (En el Ejercicio 3 se da otro ejemplo semejante.)

(b) Una excepción: el recíproco de la Proposición 20.3 sí se cumple si $\mathrm{X}(\cdot)$ es gaussiano. Mas explícitamente, si $\mathrm{X}(\cdot)$ es un PE **gaussiano** y L_2–estacionario, entonces es estrictamente estacionario.

(c) Usando la desigualdad de Cauchy-Schwarz

$$[E(\mathrm{X}Y)]^2 \leq E(\mathrm{X}^2) \cdot E(Y^2) \tag{20.5}$$

para vv.aa. X y Y en L_2, se obtiene que

$$|K_{\mathrm{X}}(s,t)|^2 \leq K_{\mathrm{X}}(s,s)\ K_{\mathrm{X}}(t,t) = \mathrm{Var}[\mathrm{X}(s)] \cdot \mathrm{Var}[\mathrm{X}(t)]. \tag{20.6}$$

(Vea el Ejercicio 4). $\square$

Proposición 20.5. Si $\mathrm{X}(\cdot)$ es L_2–estacionario, entonces para todo $t \in \mathbb{R}$:

(a) r_{X} es una función par, i.e. $r_{\mathrm{X}}(t) = r_{\mathrm{X}}(-t)$;

(b) $\mathrm{Var}[\mathrm{X}(t)] = r_{\mathrm{X}}(0)$; y

(c) $|r_{\mathrm{X}}(t)| \leq r_{\mathrm{X}}(0)$.

Demostración. (a). Por la simetría de K_{X}, o sea $K_{\mathrm{X}}(s,t) = K_{\mathrm{X}}(t,s)$, se obtiene $K_{\mathrm{X}}(0, t-s) = K_{\mathrm{X}}(0, s-t)$, i.e. $r_{\mathrm{X}}(t-s) = r_{\mathrm{X}}(s-t)$. Tomando $s = 0$ se obtiene (a).

(b) $\mathrm{Var}[\mathrm{X}(t)] = K_{\mathrm{X}}(t,t) = K_{\mathrm{X}}(0,0) = r_{\mathrm{X}}(0)$.

(c) Tomando $s = 0$ en la desigualdad (6) se obtiene

$$|r_{\mathrm{X}}(t)|^2 \leq \mathrm{Var}[\mathrm{X}(0)] \cdot \mathrm{Var}[\mathrm{X}(t)] = r_{\mathrm{X}}(0)^2, \quad \text{por (b)}.$$

$\square$

Ejemplo 20.6. Sean Y_1, Y_2 i.i.d. con distribución $N(0, \sigma^2)$, y sea λ una constante. Demuestre que el PE $\mathrm{X}(\cdot)$ definido como

$$\mathrm{X}(t) := Y_1 \cos \lambda t + Y_2 \operatorname{sen} \lambda t \quad \forall\, t \in \mathbb{R}$$

es (a) gaussiano, y (b) L_2–estacionario.

Solución. (a) Para cualquier $n = 1, 2, \ldots$, y números $t_1, \ldots, t_n, a_1, \ldots, a_n$ en $\mathbb{R}$,

$$\sum_{j=1}^{n} a_j \mathrm{X}(t_j) = Y_1 \cdot \sum_{j=1}^{n} a_j \cos \lambda t_j + Y_2 \cdot \sum_{j=1}^{n} a_j \operatorname{sen} \lambda t_j,$$

que es de la forma $\alpha Y_1 + \beta Y_2 \sim N(0, (\alpha^2 + \beta^2)\sigma^2)$.

(b) Para todo s, t en $\mathbb{R}$

$$\begin{aligned}
m_{\mathrm{X}}(t) &:= E\mathrm{X}(t) = (EY_1) \cos \lambda t + (EY_2) \operatorname{sen} \lambda t = 0; \\
K_{\mathrm{X}}(s,t) &= E[\mathrm{X}(s)\mathrm{X}(t)] \\
&= E[Y_1^2 \cos \lambda s \cdot \cos \lambda t + Y_2^2 \operatorname{sen} \lambda s \cdot \operatorname{sen} \lambda t \\
&\quad + Y_1 Y_2 \cos \lambda s \cdot \operatorname{sen} \lambda t + Y_1 Y_2 \operatorname{sen} \lambda s \cdot \cos \lambda t] \\
&= \sigma^2 (\cos \lambda s \cdot \cos \lambda t + \operatorname{sen} \lambda s \cdot \operatorname{sen} \lambda t) \\
&= \sigma^2 \cos \lambda (t - s).
\end{aligned}$$

Por lo tanto, $\mathrm{X}(\cdot)$ es débilmente estacionario con media $m_{\mathrm{X}} = 0$ y función de covarianza $r_{\mathrm{X}}(t) = K_{\mathrm{X}}(0,t) = \sigma^2 \cos \lambda t$. $\square$

Ejemplo 20.7. Proceso de Poisson con $T = \mathbb{R}$. (Compare con la Definición 16.3 en la que $T = [0, \infty)$.) Decimos que $N(\cdot) = \{N(t),\, t \in \mathbb{R}\}$ es un **proceso de Poisson** con parámetro $\lambda > 0$ si:

(a) $N(0) = 0$,

(b) $N(\cdot)$ tiene incrementos independientes, y

(c) $N(\cdot)$ tiene incrementos estacionarios tales que

$$N(t) - N(s) \sim \text{Poi}\,(\lambda(t-s)) \quad \forall\, t \geq s. \tag{20.7}$$

Demuestre que $N(\cdot)$ tiene función media y función de covarianza

$$m_N(t) = \lambda t \tag{20.8}$$

$$K_N(s,t) = \begin{cases} \lambda \cdot \min\{|s|, |t|\} & \text{si} \quad s \cdot t \geq 0, \\ 0 & \text{si} \quad s \cdot t < 0. \end{cases} \tag{20.9}$$

En particular, $\text{Var}\,[N(t)] = \lambda|t|$ para todo $t \in \mathbb{R}$.

Solución. Si $t \geq 0$, se sigue de (7) que $N(t) \sim \text{Poi}(\lambda t)$, y se obtiene (8). Análogamente, si $t < 0$, se tiene que $-N(t) \sim \text{Poi}(-\lambda t)$, de donde

$$m_N(t) = -E[-N(t)] = -(-\lambda t) = \lambda t.$$

Esto completa la demostración de (8).

Si $s, t \geq 0$, (9) se obtiene del Ejemplo 16.8(a):

$$K_N(s,t) = \lambda \cdot \min\{s,t\} \quad \forall\, s,t \geq 0. \tag{20.10}$$

Ahora supóngase que $s < t \leq 0$. Entonces escribiendo $N(t) = N(t) - N(0)$ y

$$N(s) = N(s) - N(t) + N(t) = -[N(t) - N(s)] + N(t)$$

y usando (b) y (c), vemos que

$$\begin{aligned} E[N(s)N(t)] &= -E[(N(t) - N(s)) \cdot N(t)] + EN^2(t) \\ &= -\lambda(t-s) \cdot \lambda t + \lambda|t| + \lambda^2 t^2 \\ &= \lambda^2 st - \lambda t \end{aligned}$$

Por lo tanto, como $EN(s) \cdot EN(t) = \lambda^2 st$, se sigue que

$$K_N(s,t) = -\lambda t = \lambda(-t) = \lambda|t| \quad \text{si} \;\; s < t \leq 0.$$

En forma similar se obtiene que

$$K_N(s,t) = \lambda|s| \quad \text{si} \quad t < s \leq 0.$$

Combinando estos resultados con (10), se obtiene (9) cuando $s \cdot t \geq 0$, es decir, cuando s y t tienen el mismo signo.

Por otra parte, si $s < 0 < t$, escribiendo

$$N(s) = N(s) - N(0) = -[N(0) - N(s)] \quad \text{y} \quad N(t) = N(t) - N(0),$$

usando (b) y (c) se obtiene

$$E[N(s)N(t)] = -(-\lambda s)(\lambda t) = \lambda^2 st.$$

Luego, como $EN(s) \cdot EN(t) = \lambda^2 st$, se sigue que $K_N(s,t) = 0$ si $s < 0 < t$. En forma análoga se obtiene $K_N(s,t) = 0$ si $t < 0 < s$. $\square$

Definición 20.8. La **función de covarianza conjunta** de dos PEs $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\ t \in T\}$ y $Y(\cdot) = \{Y(t),\ t \in T\}$ en L_2 se define como

$$K_{\mathrm{X}Y}(s,t) := \mathrm{Cov}(\mathrm{X}(s), Y(t)) \quad \forall\, s,t \in T$$

es decir,

$$\begin{aligned} K_{\mathrm{X}Y}(s,t) &= E[(\mathrm{X}(s) - m_{\mathrm{X}}(s)) \cdot (Y(t) - m_Y(t))] \\ &= E[\mathrm{X}(s)Y(t)] - m_{\mathrm{X}}(s)m_Y(t). \end{aligned}$$

Se dice que $\mathrm{X}(\cdot)$ y $Y(\cdot)$ **no están correlacionados** si $K_{\mathrm{X}Y}(s,t) = 0$ para todo $s,t \in T$.

En la sección de ejercicios se ven algunas propiedades de la función de covarianza conjunta.

B. Teoremas espectrales

Para esta sección puede primero consultar el Apéndice.

Sea $X := \{X_n,\ n \in \mathbb{Z}\}$ un proceso L_2-estacionario complejo valuado con función de covarianza

$$r(n) := Cov(X_n, X_0) := E[(X_n - E(X_n))(\overline{X_0 - E[X_0]})] = Cov(X_{n+k}, X_k).$$

Observación 20.9. Se puede ver que r es una función no-negativa definida, es decir, tal que para cualquier colección finita $z_1, \ldots, z_n \in \mathbb{C}$ y $k_1, \ldots, k_n \in \mathbb{Z}$ se tiene que $\sum_{i,j=1}^n z_i r(k_i - k_j)\overline{z_j} \geq 0$.

Se puede probar también que

i) $r(0) \geq 0$,
ii) $r(-n) = \overline{r(n)}$,
iii) $|r(n)| \leq r(0)$ y
iv) $|r(n) - r(m)|^2 \leq 2r(0)(r(0) - Re(r(n-m)))$.

Ejemplo 20.10. Compare con Ejemplo 20.6 Para N fijo, sea $X_n := \sum_{k=1}^N z_k e^{i\lambda_k n}$, donde $z_1, \ldots, z_N \in \mathbb{C}$ (amplitudes) son vv.aa. en L_2 con media cero y ortogonales, y $\lambda_1, \ldots, \lambda_N \in [-\pi, \pi]$ (frecuencias) son números fijos con $\lambda_i \neq \lambda_j$, $i \neq j$. Se dice que X es la suma de osciladores armónicos con ciertas amplitudes y frecuencias.

Se puede mostrar que X es L_2-estacionario y que $r(n) = \sum_{k=1}^N \sigma_k^2 e^{i\lambda_k n}$, donde $\sigma_k^2 := E|z_k|^2$. Esto se puede extender al caso $N = \infty$ cuando $\sum_{k=1}^\infty \sigma_k^2 < \infty$.

Definiendo $F(\lambda) := \sum_{\{k:\lambda_k \leq \lambda\}} \sigma_k^2$ vemos que

$$r(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda).$$

Teorema 20.11. ***(Herglotz)****. Una función $r : \mathbb{Z} \to \mathbb{C}$ es no-negativa definida y $r(-n) = \overline{r(n)}$ si y solo si hay una medida F positiva y finita sobre $[-\pi, \pi]$ con la σ-álgebra de Borel tal que*

$$r(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda). \tag{20.11}$$

A F le llamamos la ***distribución espectral****.*

Demostración. ($\Leftarrow$) Si se cumple (20.11), entonces $r(-n) = \overline{r(n)}$. También, para $z_1, \ldots, z_n \in \mathbb{C}$

$$\sum_{k,j=1}^{n} z_k r(k-j) \overline{z_j} = \int_{-\pi}^{\pi} \sum_{k,j=1}^{n} z_k \overline{z_j} e^{i(k-j)\lambda} F(d\lambda) = \int_{-\pi}^{\pi} \left| \sum_{k=1}^{n} z_k e^{ik\lambda} \right|^2 F(d\lambda),$$

lo cual es no-negativo.

($\Rightarrow$) Definamos

$$f_N(\lambda) := \frac{1}{2\pi N} \sum_{k,j=1}^{N} e^{-ik\lambda} r(k-j) e^{ij\lambda},$$

y vemos por la hipótesis que es no-negativo para toda λ. Reescribiendo tenemos

$$f_N(\lambda) = \frac{1}{2\pi N} \sum_{k=-(N-1)}^{N-1} (N - |k|) e^{-ik\lambda} r(k).$$

Definiendo $F_N(\lambda) := \int_{-\pi}^{\lambda} f_N(x) dx$ con $\lambda \in [-\pi, \pi]$ llegamos a

$$\begin{aligned} \int_{-\pi}^{\pi} e^{in\lambda} F_N(d\lambda) &= \frac{1}{2\pi} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) r(k) \int_{-\pi}^{\pi} e^{i(n-k)\lambda} d\lambda \\ &= \begin{cases} \left(1 - \frac{|n|}{N}\right) r(n) & |n| < N \\ 0 & e.o.c. \end{cases} \end{aligned}$$

Ya vimos que $f_N \geq 0$, por que F_N define una medida no-negativa. Como $F_N(\pi) = r(0) < \infty$ se ve que la familia de medidas $\{F_N; N \geq 1\}$ es ajustada (o "tight" en inglés) y por el Teorema de Prokhorov para alguna subsucesión hay convergencia débil a alguna medida F sobre $[-\pi, \pi]$. Esto es, que para toda función $g : [-\pi, \pi] \to \mathbb{C}$ con $g(\pi) = g(-\pi)$ se tiene

$$\int_{-\pi}^{\pi} g(\lambda) F_{N_k}(d\lambda) \to \int_{-\pi}^{\pi} g(\lambda) F(d\lambda), \ k \to \infty.$$

Así vemos que se cumple (20.11). □

Ahora veremos como X tiene una representación del tipo (20.11) pero en terminos de un integral c.r.a un cierto proceso estocástico.

Definición 20.12. Un proceso estocástico $Z := \{Z(t), \ t \in [-\pi, \pi]\}$ complejo valuado es de **incrementos ortogonales** (PIO) si para toda $t \in [-\pi, \pi]$ $E[Z(t)] = 0$, $E|Z(t)|^2 < \infty$ y siempre que $-\pi \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq \pi$

$$E((Z(t_1) - Z(s_1))\overline{(Z(t_2) - Z(s_2))}) = 0.$$

Ejemplo 20.13. Ejemplos de PIOs son el MB y el proceso de Poisson, ambos sobre $[-\pi, \pi]$. Un ejemplo de un PIO en $\mathbb{C}$ es $Z(t) := W_1(t) + iW_2(t)$, donde W_1 y W_2 son MBs independientes.

Proposición 20.14. Si Z es un PIO, entonces existe una medida deBorel μ sobre $[-\pi, \pi]$ tal que $F(t) - F(s) = E((Z(t) - Z(s))\overline{(Z(t) - Z(s))})$, donde $F(t) := \mu([0, t])$.

Ahora construimos la integral estocástica c.r.a un PIO. Para ello considere los siguientes espacios de Hilbert. Sea $L^2(F)$ es el espacio de funciones $f : [-\pi, \pi] \to \mathbb{C}$ cuadrado integrables con respecto al producto interno $\int_{-\pi}^{\pi} f(x)\overline{g(x)} F(dx)$. Defina también $L_2 := L_2(\Omega, \mathcal{F}, P)$.

Una función $f \in L^2(F)$ es simple si es de la forma

$$f(x) = \sum_{i=0}^{n} c_i I_{(t_{i1}, t_{i+1}]}(x),$$

donde $c_1, \ldots, c_n \in \mathbb{C}$ y $-\pi = t_0 \leq t_1 \leq \ldots \leq t_n \leq t_{n+1} = \pi$. Notamos que cualquier función simple admite una única representación si pedimos que $c_i \neq c_j$ cuando $j = i + 1$. Llamemos pues a esta clase de funciones $\mathcal{D}$.

Definimos la integral estocástica de $f \in \mathcal{D}$ c.r.a a un PIO Z como el mapeo $I : L^2(F) \to L_2$ tal que

$$I(f) = \sum_{i=0}^{n} c_i (Z(t_{i+1}) - Z(t_i)).$$

Se puede verificar que dicho mapeo es consistente en $\mathcal{D}$, pues aunque f y g sean representaciones diferentes de una misma función en $\mathcal{D}$ se tiene que $I(f) = I(g)$ c.s.

Resulta que dicho mapeo, por lo pronto de $\mathcal{D} \subset L^2(F)$ a L_2, conserva el producto interno (y por lo tanto la norma), es decir

$$(\forall f, g \in \mathcal{D}) \int_{-\pi}^{\pi} f(x)\overline{g(x)}F(dx) = E[I(f)\overline{I(g)}].$$

Así, se dice que I es una isometría, la cual es de hecho también lineal.

Ahora bien, como $C[-\pi, \pi]$ es denso en $L^2(F)$ con la norma L^2 y $\mathcal{D}$ es denso en $C[-\pi, \pi]$ con la norma del supremo, entonces se puede verificar que $\mathcal{D}$ es denso en $L^2(F)$. Por lo tanto, si f es un elemento arbitrario en $L^2(F)$, entonces hay una sucesión $\{f_n\}$ en $\mathcal{D}$ con $f_n \to f,\ n \to \infty$. Pero esto implica que $\{f_n\}$ es de Cauchy, así que $\{I(f_n)\}$ es también de Cauchy (justo porque I preserva la norma). Pero si $\{I(f_n)\}$ es de Cauchy en L_2, hay v.a. $\xi \in L_2$ tal que $I(f_n) \to \xi,\ n \to \infty$. Definimos y denotamos la integral estocástica de f c.r.a Z como

$$\int_{-\pi}^{\pi} f(t)Z(dt) := I(f) := \xi = \lim_{n\to\infty} I(f_n).$$

Queda entonces extendido I sobre todo $L^2(F)$, en particular para la función e^{int}. Podemos ahora enunciar y probar la representación espectral de X.

Teorema 20.15. ***(de representación espectral)*** *Si $X := \{X_n,\ n \in \mathbb{Z}\}$ es un proceso L_2-estacionario con distribución espectral F, y tal que $E(X_n) = 0 y var(X_n) = 1$. Entonces existe un PIO Z tal que*

$$(\forall n \in \mathbb{Z})\ X_n = \int_{-\pi}^{\pi} e^{int} Z(dt),$$

además $E(|Z(t) - Z(s)|^2) = F(t) - F(s)$, siempre que $-\pi \le s \le t \le \pi$

Demostración. Defina los siguiente subespacios lineales de $L^2(F)$ y L_2, respectivamente:

$$L_0^2(F) := \left\{ f(t) := \sum_{k=-m}^{n} \alpha_k e^{ikt} \right\},\ L_0^2(X) := \left\{ \xi := \sum_{k=-m}^{n} \alpha_k X_k \right\},$$

con α_k's $\in \mathbb{C}$.

Definamos el mapeo

$$T\left(\sum_{k=-m}^{n} \alpha_k e^{ikt}\right) = \sum_{k=-m}^{n} \alpha_k X_k.$$

Se puede verificar que T es una isometría lineal. Ahora, por un procedimiento similar al definir la integral estocástica tenemos la extensión

$$T : \overline{L_0^2(F)} \to \overline{L_0^2(X)}.$$

Aunque hay que observar que de hecho $\overline{L_0^2(F)} = L^2(F)$. De tal manera que para $I_{(t,s]}(s)$, la cual está en $L^2(F)$, hay $\xi \in \overline{L_0^2(X)}$ tal que $T(I_{(t,s]}) = \xi$. Definamos pues el proceso $Z(t) := T(I_{(0,t]})$, el cual es un PIO tal que

$$E(|Z(t) - Z(s)|^2) = \|I_{(t,s]}\|_2^2 = F(t) - F(s).$$

Además, por propiedades de linealidad tenemos la siguiente igualdad para funciones simples con intervalos $(t_i, t_{i+1}]$ disjuntos:

$$T\left(\sum_{i=0}^{n} c_i I_{(t_{i+1},t_i]}\right) = \sum_{i=0}^{n} c_i(Z(t_{i+1}) - Z(t_i)).$$

O sea que T coincide con I para funciones simples. Tomando límites para cualquier $f \in L^2(F)$, también se tiene que $T(f) = I(f)$. En particular para la función $f(t) := e^{int}$, pero como por definición $T(f) = X_n$, concluimos que

$$X_n = \int_{-\pi}^{\pi} e^{int} Z(dt).$$

$\square$

Observación 20.16. Estas representaciones espectrales también se tienen para procesos L_2-estacionarios con tiempo $t \in \mathbb{R}$, aunque tanto la densidad espectral y el PIO correspondientes están definidos sobre todo $\mathbb{R}$.

C. Teoremas ergódicos

Definición 20.17. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $T : \Omega \to \Omega$ una transformación medible (i.e. que la imagen inversa de medibles son medibles). Se dice que T es un **morfismo** si

$$(\forall A \in \mathcal{F})\ P(T^{-1}(A)) = P(A).$$

Proposición 20.18. Sea ξ una v.a. sobre $(\Omega, \mathcal{F}, P)$ y T un morfismo. Entonces $X_n := \xi \circ T^n$ $n = 0, 1, \ldots$ es un proceso estacionario.

Proposición 20.19. Dado un proceso estacionario $\{X_n\}_{n\in\mathbb{N}}$, existe un espacio de probabilidad que soporta una v.a. ξ y un morfismo T tal que $\{X_n\}_{n\geq 1}$ es igual en distribución a $\{\xi \circ T^n\}_{n\geq 1}$.

Demostración. Sea $\Omega' := \prod_{i=1}^{\infty} \mathbb{R}$ dotado con su $\sigma-$álgebra natural $\mathcal{F}'$ y con una medida de probabilidad dada por

$$P'(B) := P(\{\omega \in \Omega : X(\omega) \in B\})$$

para cualquier $B \in \mathcal{F}'$. Ahora, para $\omega' := (\omega'_1, \omega'_2, \ldots) \in \Omega'$ definamos

$$T(\omega') = T((\omega'_1, \omega'_2, \ldots)) := (\omega'_2, \omega'_3, \ldots)$$

y $\xi(\omega') := \omega'_1$. Esta construcción satisface lo deseado. $\square$

Teorema 20.20. ***(de recurrencia de Poincaré).*** *Sea T un morfismo sobre $(\Omega, \mathcal{F}, P)$. Entonces, para cualquier $A \in \mathcal{F}$ y casi todo $\omega \in A$, se tiene que $T^n(\omega) \in A$ para una cantidad infinita de valores n.*

Demostración. Sea $D := \{\omega \in A : (\forall n \geq 1)\ T^n(\omega) \notin A\}$. Notemos que $D \bigcap T^{-n}(D) = \emptyset$ para toda $n \geq 1$. Luego $T^{-m}(D) \bigcap T^{-(m+n)}(D) = T^{-m}(D \bigcap T^{-n} D) = \emptyset$. Así podemos ver que $D, T^{-1}D, T^{-2}D, \ldots$ son disjuntos, y todos tiene la misma medida pues T es morfismo. Con esto, tenemos que

$$\sum_{n=0}^{\infty} P(D) = \sum_{n=0}^{\infty} P(T^{-n}D) \leq P(\Omega) = 1,$$

entonces $P(T^{-n}D) = 0$ para $n = 0, 1, \ldots$. Definamos $D_n := T^{-n}D$, entonces justamente $A - \bigcup_{n=0}^{\infty} D_n$ son todos los ω's que "regresan" a A una cantidad infinita de veces, y como vemos, son casi todos.

$\square$

Corolario 20.21. *Con las condiciones anteriores, sea $X \geq 0$ v.a. Entonces $(\forall \omega \in \{X > 0\})$*

$$\sum_{k=0}^{\infty} X(T^k(\omega)) = \infty \ c.s.$$

Definición 20.22. Sea T un morfismo sobre $(\Omega, \mathcal{F}, P)$.
i) $A \in \mathcal{F}$ es **T-invariante** si $T^{-1}A = A$, y X v.a. es T-invariante si $X = X \circ T$ simpre.
ii) T es **ergódica** si para todo conjunto A que es T-invariante, se tiene $P(A) = 0$ o 1.

Se puede mostrar que la colección

$$\mathcal{I} := \{A \in \mathcal{F} :\ T\text{-invariante}\}$$

es una σ-álgebra.

Proposición 20.23. Las siguientes son equivalentes:
i) T es ergódica.
ii) Para $A \in \mathcal{F}$ con $P(A \triangle T^{-1}A) = 0$, se tiene $P(A) = 0$ o 1.
iii) Para $A \in \mathcal{F}$ con $P(A) > 0$, $P(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$.
iv) Para $A, B \in \mathcal{F}$ con $P(A)P(B) > 0$, existe $n \geq 1$ tal que $P(B \bigcap T^{-n}A) > 0$.

Demostración. (i$\Rightarrow$ ii) El conjunto

$$C := \bigcap_{m=0}^{\infty} \bigcup_{n=m}^{\infty} T^{-n}A$$

es T-invariante y además es casi todo A.

Primero vemos que $(\forall m \geq 0)$

$$A \triangle T^{-m}(A) \subset \bigcup_{n=m}^{\infty} A \triangle T^{-n}A.$$

Ahora vemos que $(\forall n \geq 1)$ $P\left(A \triangle T^{-n}A\right) = 0$, pues $A \triangle T^{-n}A$ está contenido en

$$\bigcup_{i=0}^{n-1} T^{-i}A \triangle T^{-(i+1)}A,$$

el cual tiene probabilidad cero. Sea pues $C_m := \bigcup_{n=m}^{\infty} T^{-n}A$. Como $\{C_m\}_{m=1}^{\infty}$ es monótona decreciente y $P(C_m \triangle A) = 0$, entonces $P(C \triangle A) = 0$, i.e. $P(C) = P(A)$.

Finalmente, vemos que

$$T^{-1}C = \bigcap_{m=0}^{\infty} \bigcup_{n=m+1}^{\infty} T^{n}A = C.$$

Así, como C es $T-$invariante y T es ergódico, $P(C) = 0$ o 1, i.e. $P(A) = 0$ o $P(A) = 1$.
(ii$\Rightarrow$ iii) Sea $B := \bigcup_{n=1}^{\infty} T^{-n}A$, entonces $T^{-1}B \subset B$. Pero como $P(T^{-1}B) = P(B)$, $P(B) = 0$ o 1. Sin embargo la única posibilidad es que $P(B) = 1$, pues $T^{-1}A \subset B$ y $P(T^{-1}A) = P(A) > 0$.
(iii$\Rightarrow$ iv) Esto ocurre de la desigualdad

$$0 < P(B) = P(\bigcup_{n=1}^{\infty}(B \bigcap T^{-n}A)) \leq \sum_{n=1}^{\infty} P(B \bigcap T^{-n}A).$$

(iv$\Rightarrow$ i) De $0 = P(A^c \bigcap A) = P(A^c \bigcap T^{-n}A)$, se tiene que $P(A) = 0$ o $P(A^c) = 0$. $\square$

Observación 20.24. Cuando T es ergódica, el resultado anterior nos dice en particular como al dividir Ω en una colección $\mathcal{D}$ de cojuntos de medida estrictamente positiva, casi todo punto visitará una cantidad infinita de veces cada conjunto de la colección $\mathcal{D}$. Si T no es ergódica entonces esto NO ocurre. Se podrá apreciar el efecto de esto en los teoremas ergódicos que posteriormente veremos.

Proposición 20.25. Para una v.a. X y T un morfismo:
i) $E[X] = E[X \circ T]$ y $E[X|\mathcal{I}] \circ T = E[X|\mathcal{I}]$ c.s.
ii) Cuando T es ergódica, toda v.a. T-invariante es constante c.s.

Lema 20.26. (**Teorema ergódico maximal**). Sea T un morfismo sobre un espacio de probabilidad y X v.a. con media finita. Defina

$$S_k(\omega) := X(\omega) + X(T(\omega)) + X(T^2(\omega)) + \ldots + X(T^{k-1}(\omega)) \tag{20.12}$$

y

$$M_k(\omega) := \max\{0, S_1(\omega), \ldots, S_k(\omega)\}. \tag{20.13}$$

Entonces

$$E\left[XI_{\{M_n>0\}}\right] \geq 0 \text{ para toda } n \geq 1. \tag{20.14}$$

Demostración. Si $n \geq k$, $M_n \circ T \geq S_k \circ T$, por lo que $X + M_n \circ T \geq X + S_k \circ T$. Pero $X + S_k \circ T = S_{k+1}$, luego $X \geq S_{k+1} - M_n \circ T$ para $k = 1, 2, \ldots, n$. Como también $X \geq S_1 - M_n \circ T$, entonces $X \geq \max\{S_1, \ldots, S_n\} - M_n \circ T$. Así,

$$E\left[XI_{\{M_n>0\}}\right] \geq E\left[(\max\{S_1, \ldots, S_n\} - M_n \circ T)I_{\{M_n>0\}}\right].$$

Sin embargo, en $\{M_n > 0\}$, $\max(S_1, \ldots, S_n) = M_n$, luego

$$E\left[XI_{\{M_n>0\}}\right] \geq E[(M_n - M_n \circ T)I_{\{M_n>0\}}] \geq E[M_n - M_n \circ T],$$

pues $E[M_n] = E[M_n I_{\{M_n>0\}}]$ y $E[M_n \circ T] \geq E[M_n \circ T I_{\{M_n>0\}}]$. Por la Proposición 20.25 $E[M_n] = E[M_n \circ T]$, de donde se sigue el resultado. $\square$

El siguiente teorema dice como el promedio en el tiempo converge a la media en el espacio, concepto que está plasmado en otro teorema ergódico (por von Neumann) que precede al siguiente resultado.

Teorema 20.27. ***(Birkoff)**. Sea T un morfismo y ξ una v.a. con media finita. Entonces*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \xi(T^k(\omega)) = E[\xi|\mathcal{I}](\omega) \tag{20.15}$$

para casi toda $\omega \in \Omega$. Además, si T es ergódica, $E[\xi|\mathcal{I}]$ es constante.

Demostración. Trabajemos mejor con $X := \xi - E[\xi|\mathcal{I}]$, y usando la notación en (20.12) queremos mostrar que

$$0 \leq \xi_* := \liminf_n \frac{1}{n} S_n \leq \xi^* := \limsup_n \frac{1}{n} S_n \leq 0 \ c.s., \tag{20.16}$$

pues usando la Proposición 20.25 se tendría el resultado.

De la igualdad

$$\frac{n+1}{n}\left(\frac{1}{n+1}\sum_{i=0}^{n} X \circ T^i\right) = \frac{1}{n}\sum_{i=0}^{n-1} X \circ T^i \circ T + \frac{1}{n}X, \tag{20.17}$$

y al tomar $\liminf$ y $\limsup$ obtenemos que $\xi_* \circ T = \xi_*$ y $\xi^* \circ T = \xi^*$, respectivamente, o sea que ξ_* y ξ^* son T-invariantes. En particular para ϵ arbitraria $A_\epsilon := \{\xi^* > \epsilon\}$ es T-invariante.

Definamos $X^* := (X - \epsilon)I_{A_\epsilon}$ y usando la notación en (20.12) y (20.13) con X^* en lugar de X definimos respectivamente S_k^* y M_k^*. Así, usando el Lema 20.26 tenemos que

$$E\left[X^* I_{\{M_n^*>0\}}\right] \geq 0 \text{ para toda } n \geq 1.$$

Al tomar $n \to \infty$ vemos que

$$\{M_n^* > 0\} = \left\{\max_{k=1,\ldots,n} S_k^* > 0\right\} = \left\{\max_{k=1,\ldots,n} \frac{S_k^*}{k} > 0\right\} \uparrow \left\{\sup_{k\geq 1} \frac{S_k^*}{k} > 0\right\}. \tag{20.18}$$

Usando la T-invarianza de A_ϵ, podemos mostrar lo siguiente:

$$\left\{\sup_{k\geq 1}\frac{S_k^*}{k}>0\right\}=\left\{\sup_{k\geq 1}\frac{S_k}{k}>\epsilon\right\}\bigcap A_\epsilon=A_\epsilon. \tag{20.19}$$

La primera igualdad surge de expandir completamente el lado izquierdo, y la segunda igualdad se obtiene a partir del hecho

$$\sup_k \frac{1}{k}S_k \geq \limsup_k \frac{1}{k}S_k.$$

Ahora, como $E[|X^*|]\leq E[|X|]+\epsilon$, por el teorema de convergencia dominada

$$0\leq E\left[X^* I_{\{M_n^*>0\}}\right]\to E\left[X^* I_{A_\epsilon}\right]. \tag{20.20}$$

Sin embargo

$$\begin{aligned} 0 &\leq E\left[X^* I_{A_\epsilon}\right]=E\left[X I_{A_\epsilon}\right]-\epsilon P(A_\epsilon)\\ &= E\left[E\left[X I_{A_\epsilon}|\mathcal{I}\right]\right]-\epsilon P(A_\epsilon)=-\epsilon P(A_\epsilon), \end{aligned}$$

por lo que $P(A_\epsilon)=0$, pero como ϵ es arbitrariamente pequeño concluimos que $P(\xi^*\leq 0)=1-P(\xi^*>0)=1$.

Efectuando el mismo analisis para $-X$, como

$$\limsup -\frac{S_k}{k}=-\liminf\frac{S_k}{k}=-\xi_*,$$

podemos decir que $P(-\xi_*\leq 0)=P(\xi_*\geq 0)=1$. Lo cual termina por mostrar (20.16).

Como ya tenemos el límite (20.15), podemos decir que $E[\xi|\mathcal{I}]$ es T-invariante, pero por la Proposición 20.25 sabemos que es de hecho constante. □

De la Proposition 20.19 podemos obtener lo siguiente.

Corolario 20.28. ***(Teorema Ergódico)***. *Sea $X:=\{X_n\}_{n=0,1,\ldots}$ un proceso estacionario. Si $E|X_0|<\infty$, entonces existe Y v.a. con $E[Y]=E[X_0]$ tal que*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}X_k=Y \ c.s.$$

Ejercicios § 20

20.1. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\}$ un PE en L_2 con función de covarianza K_{X}. Demuestre:

(a) $\operatorname{Cov}\left(\sum_{j=1}^{n} a_j \mathrm{X}(t_j), \sum_{k=1}^{m} b_k \mathrm{X}(s_k)\right) = \sum_{j=1}^{n}\sum_{k=1}^{m} a_j b_k K_{\mathrm{X}}(t_j, s_k)$; en particular,

(b) $\operatorname{Var}\left[\sum_{j=1}^{n} a_j \mathrm{X}(t_j)\right] = \sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \mathrm{K}_X(t_j, t_k)$.

20.2. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in \mathbb{R}\}$. Demuestre que las siguientes condiciones son equivalentes.

(a) $\mathrm{X}(\cdot)$ es L_2–estacionario.

(b) Para cada $h \in \mathbb{R}$, el PE $Y(\cdot)$ definido por

$$Y(t) := \mathrm{X}(t+h) \quad \forall\, t \in \mathbb{R}$$

tiene la misma función media y la misma función de covarianza que $\mathrm{X}(\cdot)$.

(c) Las funciones $E\mathrm{X}(s)$ y $E[\mathrm{X}(s)\mathrm{X}(s+t)]$ son independientes de s.

20.3. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in \mathbb{R}\}$ el PE definido por

$$\mathrm{X}(t) := Z_1 \cos \lambda t + Z_2 \operatorname{sen} \lambda t,$$

en donde $\lambda \in \mathbb{R}$ es una constante y Z_1, Z_2 son i.i.d. con distribución

$$\mathrm{P}(Z_i = 1) = \mathrm{P}(Z_i = -1) = 1/2 \quad \text{para} \quad i = 1, 2.$$

Demuestre que $\mathrm{X}(\cdot) \in L_2$ es débilmente estacionario, pero **no** estrictamente estacionario.

20.4. Demuestre la desigualdad (6).

20.5. Sean $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \in T\}$ y $Y(\cdot) = \{Y(t), t \in T\}$ dos PEs en L_2. Demuestre que la función de covarianza conjunta satisface las siguientes propiedades.

(a) $K_{\mathrm{X}Y}(s,t) = K_{Y\mathrm{X}}(t,s)$,

(b) $K_{\mathrm{XX}}(s,t) = K_{\mathrm{X}}(s,t)$,

(c) $K_{\mathrm{X}+Y}(s,t) = K_{\mathrm{X}}(s,t) + K_Y(s,t) + K_{\mathrm{X}Y}(s,t) + K_{Y\mathrm{X}}(s,t)$.

(d) $K_{\mathrm{X}+Y}(s,t) = K_{\mathrm{X}}(s,t) + K_Y(s,t)$ si $\mathrm{X}(\cdot)$ y $Y(\cdot)$ *no están correlacionados.*

20.6. Sean $\mathrm{X}_j(\cdot) = \{\mathrm{X}_j(t), t \in T\}$, para $j = 1, \cdots, n$, PEs en L_2 que *no están correlacionados*, i.e.

$$K_{\mathrm{X}_j\mathrm{X}_k}(s,t) = 0 \quad \forall\, s,t \in T \quad \text{si } \; j \neq k.$$

Sea $\mathrm{X}(t) := \mathrm{X}_1(t) + \ldots + \mathrm{X}_n(t)$ para $t \in T$. Demuestre que la función media y la función de covarianza de $\mathrm{X}(\cdot)$ son

$$m_{\mathrm{X}}(t) = m_{\mathrm{X}_1}(t) + \cdots + m_{\mathrm{X}_n}(t),$$

$$K_{\mathrm{X}}(s,t) = K_{\mathrm{X}_1}(s,t) + \cdots + K_{\mathrm{X}_n}(s,t).$$

20.7. Para cada $j = 1, \ldots, n$ sea $\mathrm{X}_j(\cdot)$ el PE dado por

$$\mathrm{X}_j(t) := Z_{j1} \cos \lambda_j t + Z_{j2} \operatorname{sen} \lambda_j t \quad \forall\, t \in \mathbb{R},$$

donde los λ_j son números reales, y los Z_{ji} ($j = 1, \ldots, n$; $i = 1, 2$) son vv.aa., independientes con distribución $N(0, \sigma_j^2)$. Demuestre que

$$\mathrm{X}(t) := \mathrm{X}_1(t) + \cdots + \mathrm{X}_n(t), \;\; t \in \mathbb{R},$$

es un PE gaussiano y L_2–estacionario con función media 0 y función de covarianza

$$r_{\mathrm{X}}(t) = \sum_{j=1}^{n} \sigma_j^2 \cos \lambda_j t \quad \forall\, t \in \mathbb{R}.$$

(*Sugerencia*: compare con el Ejercicio 6.)

20.8. Sea $N(\cdot) = \{N(t), t \in \mathbb{R}\}$ el proceso de Poisson con parámetro λ del Ejemplo 20.7, y sea $\mathrm{X}(\cdot)$ el PE definido por

$$\mathrm{X}(t) := N(t+1) - N(t) \quad \forall\, t \in \mathbb{R}.$$

Demuestre que $\mathrm{X}(\cdot)$ es un PE débilmente estacionario con función media λ y función de covarianza

$$K_{\mathrm{X}}(s,t) = \begin{cases} \lambda(1 - |t-s|) & \text{si } \ |t-s| < 1, \\ 0 & \text{si } \ |t-s| \geq 1, \end{cases}$$

es decir,

$$r_{\mathrm{X}}(t) = \begin{cases} \lambda(1 - |t|) & \text{si } \ |t| < 1, \\ 0 & \text{si } \ |t| \geq 1. \end{cases}$$

20.9. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \in \mathbb{R}\}$ un PE L_2–estacionario, y sea

$$Y(t) := \mathrm{X}(t+1) - \mathrm{X}(t) \quad \forall\, t \in \mathbb{R}.$$

Demuestre que $Y(\cdot)$ es L_2–estacionario con función media cero y función de covarianza

$$r_Y(t) = 2r_{\mathrm{X}}(t) - r_{\mathrm{X}}(t-1) - r_{\mathrm{X}}(t+1) \quad \forall\, t \in \mathbb{R}.$$

20.10. Sean $Y_1, \ldots, Y_n$ i.i.d. distribuidas uniformemente sobre el intervalo $(0,1)$. Sea

$$\mathrm{X}(t) := \frac{1}{n} \sum_{j=1}^{n} I_{[0,t]}(Y_j) \quad \text{para} \quad 0 \leq t \leq 1.$$

Demuestre que para todo s, t en $[0,1]$,

$$m_{\mathrm{X}}(t) = t \quad \text{y} \quad K_{\mathrm{X}}(s,t) = \frac{1}{n}[\min(s,t) - st].$$

20.11. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \in \mathbb{R}\}$ un proceso gaussiano, y sean $f, g : \mathbb{R} \to \mathbb{R}$ funciones dadas. Sea

$$Y(t) := f(t)\mathrm{X}(g(t)) \quad \forall\, t \in \mathbb{R}.$$

Demuestre que $Y(\cdot)$ es un proceso gaussiano y calcule su función media y su función de covarianza.

20.12. Proceso de Wiener con $T = \mathbb{R}$. (Compare con la Definición 16.4 y el Ejemplo 20.7.) Decimos que $W(\cdot) = \{W(t), t \in \mathbb{R}\}$ es un **proceso de Wiener** con parámetro $\sigma^2 > 0$ si

(a) $W(0) = 0$,

(b) $W(\cdot)$ tiene incrementos independientes, y

(c) $W(\cdot)$ tiene incrementos estacionarios con $W(t) - W(s) \sim N(0, \sigma^2(t-s))$ para todo $t \geq s$.

Demuestre que $W(\cdot)$ es un PE en L_2 con función media cero y función de covarianza

$$K_W(s,t) = \begin{cases} \sigma^2 \cdot \min\{|s|, |t|\} & \text{si } \; s \cdot t \geq 0, \\ 0 & \text{si } \; s \cdot t < 0. \end{cases}$$

20.13. Sea $\mathrm{X}(\cdot)$ la señal telegráfica aleatoria del Ejercicio 16.7. Demuestre que $\mathrm{X}(\cdot)$ es L_2–estacionario.

20.14. Movimiento browniano con deriva μ. Sea $W(\cdot) = \{W(t), t \geq 0\}$ un proceso de Wiener (o movimiento browniano) estándar. Entonces el PE $\mathrm{X}(\cdot)$ definido por

$$\mathrm{X}(t) := \mu t + \sigma W(t) \quad \forall\, t \geq 0,$$

con $\mu \in \mathbb{R}$ y $\sigma > 0$ constantes, se llama movimiento browniano con deriva μ. Demuestre que $\mathrm{X}(\cdot)$ es un proceso gaussiano con función media $m_{\mathrm{X}}(t) = \mu t$ y función de covarianza $K_{\mathrm{X}}(s,t) = \sigma^2 \cdot \min(s,t)$.

20.15. Movimiento browniano geométrico. Sean $W(\cdot)$, μ y σ como en el ejercicio anterior, y sea

$$\mathrm{X}(t) := e^{\mu t + \sigma W(t)} \quad \forall\, t \geq 0.$$

Demuestre que $\mathrm{X}(\cdot)$ es un PE en L_2, que **no** es gaussiano, y con función media y de covarianza

$$m_{\mathrm{X}}(t) = e^{\mu t + \sigma^2 t/2} \quad \forall\, t \geq 0,$$

$$K_{\mathrm{X}}(s,t) = e^{(\mu + \sigma^2/2)(t+s)}(e^{\sigma^2 s} - 1) \quad \forall\, 0 \leq s \leq t.$$

21 Cálculo en L_2

Contenido: Continuidad en L_2, diferenciabilidad en L_2, integrabilidad en L_2.

Para esta sección puede primero consultar el Apéndice.

A. Continuidad en L_2

Decimos que el PE $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\ t \in T\} \in L_2$ es L_2–**continuo en** $t \in T$ si $\mathrm{X}(t+h) \to \mathrm{X}(t)$ en L_2 cuando $h \to 0$. Decimos simplemente que el PE $\mathrm{X}(\cdot)$ es L_2–**continuo** si es L_2–continuo en todo punto $t \in T$.

Recuérdese la notación $m_{\mathrm{X}}(t) := E\mathrm{X}(t)$ y $K_{\mathrm{X}}(s,t) := \mathrm{Cov}\,(\mathrm{X}(s), \mathrm{X}(t))$.

Puesto que

$$E[(\mathrm{X}(t+h) - \mathrm{X}(t))^2] \geq [m_{\mathrm{X}}(t+h) - m_{\mathrm{X}}(t)]^2,$$

si $\mathrm{X}(\cdot)$ es L_2–continuo entonces la función media $m_{\mathrm{X}}(\cdot)$ es continua.

Teorema 21.1. (**Criterio de continuidad**) *Supóngase que $m_{\mathrm{X}} : T \to \mathbb{R}$ es una función continua. Entonces $\mathrm{X}(\cdot)$ es L_2–continuo en t ssi K_{X} es continua en (t,t).*

Demostración. Nótese que $\mathrm{X}(\cdot)$ es L_2–continuo ssi $\mathrm{X}'(\cdot) = \{\mathrm{X}(t) - m_{\mathrm{X}}(t), t \in T\}$ es L_2–continuo, y, por otra parte, $\mathrm{X}'(\cdot)$ tiene función media 0 y la misma función de covariancia que $\mathrm{X}(\cdot)$. Luego, podemos suponer que

$$m_{\mathrm{X}}(t) = E\mathrm{X}(t) \equiv 0. \tag{21.1}$$

(⇒) Por definición, si $\mathrm{X}(\cdot)$ es L_2–continuo en t, entonces

$$\mathrm{X}(t+h) \to \mathrm{X}(t) \quad \text{y} \quad \mathrm{X}(t+h') \to \mathrm{X}(t) \quad \text{en} \quad L_2 \quad \text{cuando} \quad h, h' \to 0.$$

Por tanto, por el Lema 27.8.,

$$K_{\mathrm{X}}(t+h, t+h') = E[\mathrm{X}(t+h)\mathrm{X}(t+h')] \to E\mathrm{X}^2(t) = K_{\mathrm{X}}(t,t);$$

es decir, K_X es continua en (t,t).

(⇐) Si K_X es continua en (t,t), entonces

$$\begin{aligned} E[(X(t+h)-X(t))^2] &= K_X(t+h,t+h)+K_X(t,t)-2K_X(t+h,t) \\ &\to 0 \quad \text{cuando} \quad h\to 0; \end{aligned}$$

es decir, $X(t+h)\to X(t)$ en L_2 cuando $h\to 0$. □

Ejemplo 21.2. (a) Sea $X(\cdot)=\{X(t),\, t\in\mathbb{R}\}$ el proceso de Poisson con parámetro λ, del Ejemplo 20.7. Entonces $X(\cdot)$ tiene función media $m_X(t)=\lambda t$ continua, y función de covariancia

$$\begin{aligned} K_X(s,t) &= \lambda\cdot\min\{|s|,|t|\} \quad \text{si} \quad s\cdot t\geq 0 \\ &= 0 \quad \text{si} \quad s\cdot t<0. \end{aligned}$$

En particular, $K_X(t,t)=$Var$[X(t)]=\lambda|t|$ para todo $t\in\mathbb{R}$. Por lo tanto, por el Teorema 21.1, el **proceso de Poisson es L_2–continuo**.

(b) Análogamente, el proceso de Wiener $W(\cdot)=\{W(t),\, t\in\mathbb{R}\}$ del Ejercicio 20.12 es L_2–continuo. □

Corolario 21.3. *Supóngase que m_X es continua. Si K_X es continua en (t,t) para todo $t\in T$, entonces $K_X(s,t)$ es continua en (s,t) para todo $s,t\in T$.*

Demostración. De nuevo supondremos (21.1). Por el Teorema 21.1, para todo $s,t\in T$ tenemos

$$X(t+h)\to X(t) \quad \text{y} \quad X(s+h')\to X(s) \quad \text{en} \quad L_2 \quad \text{cuando} \quad h,h'\to 0.$$

Por lo tanto, por la continuidad del producto interno (Lema 27.8),

$$K_X(t+h,s+h')\to K_X(t,s) \quad \text{cuando} \quad h,h'\to 0.$$

□

Nota 21.4. En todo lo que sigue supondremos que (21.1) se cumple si m_X es continua.

Teorema 21.5. *Supóngase que $X(\cdot)\in L_2$ es débilmente estacionario, y sea $r_X(t):=K_X(0,t)$ su función de covarianza.*

(a) *Si* $X(\cdot)$ *es* L_2*–continuo en algún punto* $t \in T$*, entonces* r_X *es continua en* 0.

(b) *Si* r_X *es continua en* 0*, entonces es continua en todo* $t \in T$ *y, además,* $X(\cdot)$ *es* L_2*–continuo.*

Demostración. (a) Es obvio que $X(t) \to X(t)$ en L_2, y si $X(\cdot)$ es L_2–continuo, entonces $X(t+h) \to X(t)$ en L_2 cuando $h \to 0$. De aquí se sigue que (por la continuidad del producto interno)

$$E[X(t)X(t+h)] = r_X(h) \to EX(t)^2 = r_X(0) \quad \text{cuando} \quad h \to 0,$$

i.e. $r_X(h) \to r_X(0)$.

(b) Si r_X es continua en 0, entonces $X(\cdot)$ es L_2–continuo porque

$$E[(X(t+h) - X(t))^2] = r_X(0) + r_X(0) - 2r_X(h) \to 0 \; \forall \, t \in T, \text{ cuando } h \to 0.$$

Luego, cuando $h \to 0$ tenemos que $X(s+t+h) \to X(s+t)$ en L_2 (y también $X(s) \to X(s)$ en L_2), de modo que $r_X(t+h) \to r_X(t)$. □

Ejemplo 21.6. Sea $X(\cdot) \in L_2$ el PE definido como

$$X(t) := Y_1 \cos \lambda t + Y_2 \operatorname{sen} \lambda t \quad \forall \, t \in \mathbb{R},$$

donde λ es una constante, y Y_1, Y_2 son i.i.d $\sim N(0, \sigma^2)$. Por el Ejemplo 20.6, $X(\cdot)$ es L_2–estacionario con $m_X = 0$ y $r_X(t) = \sigma^2 \cos \lambda t \;\; \forall t \in \mathbb{R}$. Luego, como $r_X(t)$ es continua en $t = 0, X(\cdot)$ es L_2–continuo (por el Teorema 21.5). □

B. Diferenciabilidad en L_2

Definición 21.7. Sea $X(\cdot) = \{X(t),\, t \in T\}$ un PE en L_2. Decimos que $X(\cdot)$ es L_2–**diferenciable** en el punto $t \in T$ si existe una v.a. $X'(t) \in L_2$ tal que

$$\frac{X(t+h) - X(t)}{h} \to X'(t) \quad \text{en} \quad L_2 \quad \text{cuando} \quad h \to 0.$$

En este caso escribimos $X'(t) = \frac{d}{dt}X(t)$ (en L_2).

Por el Ejercicio 5, si $X(\cdot)$ es un proceso **gaussiano** y L_2–diferenciable, entonces también $X'(\cdot)$ es gaussiano.

Ejemplo 21.8. El proceso de Wiener $W(\cdot)$ **no** es L_2–diferenciable. En efecto, si fuera L_2–diferenciable, entonces por el Lema 27.9 se tendría una constante $\ell \in \mathbb{R}$ tal que

$$E\left[\frac{1}{h}(W(t+h)-W(t))\right]^2 \to \ell \quad \text{cuando} \quad h \to 0. \tag{21.2}$$

Sin embargo, como $W(t+h)-W(t) \sim N(0,\sigma^2|h|)$, vemos que el lado izquierdo de (2) es igual a

$$\frac{1}{h^2}\cdot\sigma^2|h| = \frac{\sigma^2}{|h|} \to \infty \quad \text{cuando} \quad h \to 0,$$

es decir

$$\lim_{h\to 0} E\left[\frac{1}{h}(W(t+h)-W(t))\right]^2 = \infty.$$

Asimismo, si existiera la derivada $W'(t)$ se tendría

$$\lim_{h\to 0} E\left[\frac{1}{h}(W(t+h)-W(t)) - W'(t)\right]^2 = 0,$$

de modo que

$$E(W'(t))^2 = \lim_{h\to 0} E\left[\frac{1}{h}(W(t+h)-W(t))\right]^2 = \infty.$$

Análogamente se puede demostrar que el proceso de Poisson no es L_2–diferenciable (Ejercicio 21.6). $\Box$

La siguiente proposición da, en particular, la función de covarianza de la derivada $\mathrm{X}'(\cdot)$ cuando $\mathrm{X}(\cdot)$ es L_2–estacionario. El caso general se estudia más adelante (vea Proposición 21.14).

Proposición 21.9. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\}$ un PE L_2–estacionario con función de covarianza $r(t) = K_{\mathrm{X}}(0,t)$.
i) Si $\mathrm{X}(\cdot)$ es L_2–diferenciable en todo punto $t \in T$, entonces r es dos veces diferenciable, y el PE $\mathrm{X}'(\cdot) = \{\mathrm{X}'(t), t \in T\}$ es L_2–estacionario con media cero y función de covarianza $-r''(t)$.
ii) Reciprocamente, si r es dos veces diferenciable, entonces X es L_2-diferenciable.

Demostración. i) Como $\mathrm{X}(s+t) \to \mathrm{X}(s+t)$ en L_2 y también

$$\frac{1}{h}[\mathrm{X}(s+h) - \mathrm{X}(s)] \to \mathrm{X}'(s) \text{ en } L_2,$$

del Lema 27.8 se sigue que

$$\frac{1}{h}[r(t-h) - r(t)] \to E[\mathrm{X}(s+t)\mathrm{X}'(s)].$$

Por lo tanto, r es diferenciable en todo t y

$$-r'(t) = E[\mathrm{X}(s+t)\mathrm{X}'(s)]. \tag{21.3}$$

Por otra parte

$$\frac{1}{h'}[\mathrm{X}(s+t+h') - \mathrm{X}(s+t)] \to \mathrm{X}'(s+t) \text{ en } L_2$$

y $\mathrm{X}'(s) \to \mathrm{X}'(s)$ en L_2, así que de (21.3) y de nuevo por el Lema 27.8 se obtiene

$$\frac{1}{h'}[-r'(t+h') + r'(t)] \to E[\mathrm{X}'(s+t)\mathrm{X}'(s)].$$

Es decir, $r''(t)$ existe para todo t y es igual a $-E[\mathrm{X}'(s+t)\mathrm{X}'(s)]$.

ii) Sea $Y(h) := \frac{X(t+h)-X(t)}{h}$. Entonces

$$\begin{aligned}
E[Y(h)Y(h')] &= E\left[\frac{X(t+h)-X(t)}{h}\frac{X(t+h')-X(t)}{h'}\right] \\
&= \frac{E[X(t+h)X(t+h')]}{hh'} - \frac{E[X(t+h)X(t)]}{hh'} \\
&\quad -\frac{E[X(t)X(t+h')]}{hh'} + \frac{E[X(t)^2]}{hh'} \\
&= \frac{r(h-h')}{hh'} - \frac{r(h)}{hh'} - \frac{r(h')}{hh'} + \frac{r(0)}{hh'} \\
&= \frac{1}{h}\frac{r(h-h')-r(h)}{h'} - \frac{1}{h}\frac{r(h')-r(0)}{h'} \\
&\overset{h'\to 0}{\to} \frac{1}{h}(-r'(h) + r'(0)) \overset{h\to 0}{\to} -r''(0).
\end{aligned}$$

Por lo tanto, por el Lema 27.9, $\lim_{h\to 0} Y(h)$ existe.

$\square$

C. Integrabilidad en L_2

Repaso: la integral de Riemann. Sea $f : [a,b] \to \mathbb{R}$ una función dada y $\pi_n : a = t_0 < t_1 < \cdots < t_n = b$ una **partición** del intervalo $[a,b]$. El conjunto

$$\pi_n^* := \{t_1^*, \ldots, t_n^* \mid t_{i-1} \leq t_i^* \leq t_i \quad \forall\, i = 1, \ldots, n\}$$

se llama una **partición intermedia** de π_n. En este caso definimos la **suma de Riemann**

$$S(\pi_n, \pi_n^*) := \sum_{i=1}^{n} f(t_i^*)(t_i - t_{i-1}).$$

Sea $|\pi_n| := \max_{1\leq i\leq n}(t_i - t_{i-1})$ la **malla** de la partición π_n. Si el límite

$$R(f) := \lim_{|\pi_n|\to 0} S(\pi_n, \pi_n^*)$$

existe y es independiente de la selección de las particiones π_n y π_n^*, se dice que $R(f)$ es la **integral de Riemann** de f sobre $[a,b]$, y escribimos

$$R(f) := \int_a^b f(t)dt.$$

Esta integral existe si, por ejemplo, f es continua o continua por partes. $\square$

Ahora sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\}$ un proceso en L_2 y $[a,b]$ un intervalo contenido en T. La media y la función de covariancia de $\mathrm{X}(\cdot)$ son $m_{\mathrm{X}}(t)$ y $K_{\mathrm{X}}(s,t)$, respectivamente. Además, $g : [a,b] \to \mathbb{R}$ es una función dada.

Dada una partición $\pi_n : a = t_0 < t_1 < \cdots < t_n = b$, definimos

$$I(\pi_n) := \sum_{i=1}^{n} g(t_i)\mathrm{X}(t_i)(t_i - t_{i-1}).$$

Si existe una v.a. $I \in L_2$ tal que $I(\pi_n) \to I$ en L_2 cuando $|\pi_n| \to 0$ y, además, el límite I es independiente de la selección de las particiones π_n, decimos entonces que $g(\cdot)\mathrm{X}(\cdot)$ es L_2–**integrable sobre** $[a,b]$ y escribimos

$$I := \int_a^b g(t)\mathrm{X}(t)dt.$$

Teorema 21.10. *Supóngase que m_{X} y g son continuas sobre $[a,b]$. (Basta pedir que g sea* **continua por partes***.)*

(a) *Si* $g(\cdot)\mathrm{X}(\cdot)$ *es* L_2*–integrable sobre* $[a,b]$, *entonces* $E(I)$ *satisface que*

$$E\left(\int_a^b g(t)\mathrm{X}(t)dt\right) = \int_a^b g(t)m_{\mathrm{X}}(t)dt. \tag{21.4}$$

(b) *Si además de la continuidad de* m_{X} *y* g *se tiene que* K_{X} *es continua sobre* $[a,b]\times[a,b]$, *entonces* $g(\cdot)\mathrm{X}(\cdot)$ *es* L_2*–integrable sobre* $[a,b]$ *y la integral* I *tiene la* **esperanza** $E(I)$ *en (21.4),* **segundo momento** $E(I^2)$ *dado por*

$$E(\int_a^b g(t)\mathrm{X}(t)dt)^2 = \int_a^b\int_a^b g(s)g(t)E[\mathrm{X}(s)\mathrm{X}(t)]ds\,dt, \tag{21.5}$$

y **varianza**

$$Var\,(I) = \int_a^b\int_a^b g(s)g(t)K_{\mathrm{X}}(s,t)ds\,dt. \tag{21.6}$$

Demostración. (a) Sea $\pi_n : a = t_0 < t_1 < \cdots < t_n = b$ una partición de $[a,b]$ con malla $|\pi_n| \to 0$. Entonces, por las propiedades de la integral de Riemann,

$$E[I(\pi_n)] = \sum_{i=1}^n g(t_i)m_{\mathrm{X}}(t_i)(t_i - t_{i-1}) \to \int_a^b g(t)m_{\mathrm{X}}(t)\,dt.$$

Por otra parte, como $I(\pi_n) \to I$ en L_2, de la desigualdad de Cauchy–Schwarz obtenemos

$$|EI(\pi_n) - EI|^2 = |E(I(\pi_n) - I)|^2 \le E|I(\pi_n) - I|^2 \to 0.$$

(b) Sea π_n como en la demostración de (a) y sea $\pi'_m : a = s_0 < s_1 < \cdots < s_m = b$ otra partición de $[a,b]$. Entonces

$$\begin{aligned} E[I(\pi_n)I(\pi'_m)] &= \sum_{i=1}^n\sum_{j=1}^m g(t_i)g(s_j)E[\mathrm{X}(t_i)\mathrm{X}(s_j)](t_i - t_{i-1})(s_j - s_{j-1}) \\ &\to \int_a^b\int_a^b g(t)g(s)E[\mathrm{X}(t)\mathrm{X}(s)]ds\,dt. \end{aligned} \tag{21.7}$$

Luego, por el Lema 27.9, $I(\pi_n)$ converge en L_2 a una v.a. I. Además, como $I(\pi_n) \to I$ en L_2 y $I(\pi'_m) \to I$ en L_2, por la continuidad del producto interno obtenemos

$$E[I(\pi_n)I(\pi'_m)] \to E(I^2)$$

y comparando con (21.7) se sigue (21.5). Finalmente, (21.6) se obtiene de (21.4) y (21.5). $\square$

El siguiente resultado se obtiene de manera análoga al Teorema 21.10.

Teorema 21.11. *Sean $f : [a,b] \to \mathbb{R}$ y $g : [c,d] \to \mathbb{R}$ funciones continuas por partes, y supóngase que m_{X} y K_{X} son ambas funciones continuas. Entonces*

$$Cov[\int_a^b f(t)\mathrm{X}(t)dt, \int_c^d g(s)\mathrm{X}(s)ds] = \int_a^b f(t)\left[\int_c^d g(s)K_{\mathrm{X}}(s,t)ds\right]dt$$

$$= \int_c^d g(s)\left[\int_a^b f(t)K_{\mathrm{X}}(s,t)dt\right]ds. \tag{21.8}$$

Además, el proceso $Y(t) := \int_a^t \mathrm{X}(s)ds$ satisface que para todo t, $s \in T$, con $s, t \geq a$:

$$m_Y(t) := \int_a^t m_{\mathrm{X}}(u)du, \tag{21.9}$$

$$K_Y(s,t) = \int_a^s \int_a^t K_{\mathrm{X}}(u,v)du\,dv = \int_a^t \int_a^s K_{\mathrm{X}}(u,v)dv\,du \tag{21.10}$$

Observación 21.12. Supóngase que $g : [a,b] \to \mathbb{R}$ es continua por partes y que m_{X} y K_{X} son continuas. Si $\mathrm{X}(\cdot)$ es un proceso **gaussiano**, entonces $I = \int_a^b g(t)\mathrm{X}(t)dt$ es una v.a. **gaussiana**. (Vea el Ejercicio 5.)

Ejemplo 21.13. Sea $W(\cdot) = \{W(t), t \geq 0\}$ un proceso de Wiener con parámetro $\sigma^2 > 0$. Entonces para cada $b \geq 0$, la integral $\int_0^b W(t)dt$ es una v.a. normal con media 0 y varianza (por (21.10) ó (21.6))

$$\begin{aligned}
\mathrm{Var}\,(\int_0^b W(t)dt) &= \int_0^b \int_0^b K_W(s,t)ds\,dt \\
&= 2\int_0^b \left(\int_0^s \sigma^2 t\,dt\right)ds \\
&= \sigma^2 \int_0^b s^2 ds = \sigma^2 b^3/3. \quad \square
\end{aligned}$$

Proposición 21.14. (**La función de covarianza de** $\mathrm{X}'(\cdot)$.) Supóngase que $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\}$ es L_2–diferenciable, y sea $Y(\cdot) = \{Y(t),\, t \in T\}$ un PE en L_2 con media m_Y y función de covariancia K_Y continuas. Entonces para todo $s,\, t \in T$:

(a$_1$) $K_{Y\mathrm{X}'}(s,t) = \frac{\partial}{\partial t} K_{Y\mathrm{X}}(s,t)$, (a$_2$) $K_{\mathrm{X}'Y}(s,t) = \frac{\partial}{\partial s} K_{\mathrm{X}Y}(s,t)$;

(b) $K_{\mathrm{X}\mathrm{X}'}(s,t) = \frac{\partial}{\partial t} K_{\mathrm{X}\mathrm{X}}(s,t) = \frac{\partial}{\partial t} K_{\mathrm{X}}(s,t) = K_{\mathrm{X}'\mathrm{X}}(t,s)$;

(c) $K_{\mathrm{X}'\mathrm{X}'}(s,t) = \frac{\partial}{\partial s} K_{\mathrm{X}\mathrm{X}'}(s,t)$, i.e. $K_{\mathrm{X}'}(s,t) = \frac{\partial^2}{\partial s \partial t} K_{\mathrm{X}}(s,t)$.

Demostración. Como (b) y (c) se siguen de (a), basta demostrar (a). Demostraremos (a$_1$) y por simetría se obtiene (a$_2$).

Fíjese $t > a$. Por el Ejercicio 8(b), $\mathrm{X}(t) - \mathrm{X}(a) = \int_a^t \mathrm{X}'(r)dr$ de modo que

$$Y(s)[\mathrm{X}(t) - \mathrm{X}(a)] = \int_a^t Y(s)\mathrm{X}'(r)dr.$$

Esto implica que

$$E\{Y(s)[\mathrm{X}(t) - \mathrm{X}(a)]\} = \int_a^t E[Y(s)\mathrm{X}'(r)]dr. \tag{21.11}$$

Por otra parte, como $m_{\mathrm{X}}(t) - m_{\mathrm{X}}(a) = \int_a^t m_{\mathrm{X}'}(r)dr$ (ver el Ejercicio 8(c)), tenemos

$$m_Y(s)[m_{\mathrm{X}}(t) - m_{\mathrm{X}}(a)] = \int_a^t m_Y(s) m_{\mathrm{X}'}(r)dr. \tag{21.12}$$

Luego, restando (12) de (11):

$$K_{Y\mathrm{X}}(s,t) - K_{Y\mathrm{X}}(s,a) = \int_a^t K_{Y\mathrm{X}'}(s,r)dr \tag{21.13}$$

Finalmente, como $K_{Y\mathrm{X}'}$ es necesariamente continua, para cada $s \in T$ fijo derivamos (13) con respecto a t y obtenemos (a$_1$). □

D. Expansión de procesos en L_2

Sea $\{X(t),\ t \in [a,b]\}$ un proceso L_2-continuo, con función covarianza $K(s,t)$ continua en $[a,b] \times [a,b]$ y media cero. Se sabe lo siguiente de la teoría de operadores compactos.

Teorema 21.15. *El operador*

$$[Tf](t) := \int_a^b K(s,t)f(s)ds$$

es compacto y simétrico, donde $f \in L_2[a,b]$. Luego, por el teorema espectral

$$K(s,t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t),$$

donde $\{(\lambda_n, e_n)\}_{n=1}^{\infty}$ son las parejas de eigenvalores y eigenfunciones de T; además las eigenfunciones forman un conjunto ortonormal.

Definamos pues

$$\xi_n := \int_a^b X(s)e_n(s)ds \tag{21.14}$$

Por el Teorema 21.11

$$E[\xi_n \xi_m] = \int_a^b e_m(s) \int_a^b e_n(t)K(s,t)dtds = \lambda_n \delta_{nm}.$$

También, usando la Proposición 21.14

$$E[X(t)\xi_n] = \int_a^b K(s,t)e_n(s)ds = \lambda_n e_n(t).$$

De tal manera que

$$\begin{aligned} E\left[\left|X(t) - \sum_{k=1}^{n} \xi_k e_k(t)\right|^2\right] &= K(t,t) - 2\sum_{k=1}^{n} e_k(t)E[\xi_k X(t)] + \sum_{k=1}^{n} \lambda_k e_k^2(t) \\ &= K(t,t) - \sum_{k=1}^{n} \lambda_k e_k^2(t) \to 0 \; n \to \infty. \end{aligned}$$

En resumen tenemos que

Teorema 21.16. *El proceso X L_2-continuo descrito anteriormente admite la siguiente expansión en series para cada $t \in [a,b]$ fija*

$$X(t) = \sum_{n=1}^{\infty} \xi_n e_n(t),$$

donde ξ_k son ortonormales dadas por (21.14) y $E[\xi_k^2] = \lambda_k$.

Nota 21.17. Para saber si

$$\{\{X(t), t \in [a,b]\}\} \stackrel{d}{=} \left\{\left\{\sum_{n=1}^{\infty} \xi_n e_n(t), t \in [a,b]\right\}\right\},$$

se puede intentar usar las distribuciones finito-dimensionales y el concepto de "tightness".

Ejemplo 21.18. Consideremos el MB $B(t),\ t \in [0,1]$. Sabemos que tiene media cero y que $K(s,t) = \min(s,t)$, entonces buscamos eigenvalores y eigenfunciones tales que

$$\lambda_n e_n(t) = \int_0^1 \min(s,t) e_n(s) ds = \int_0^t s e_n(s) ds + \int_t^1 t e_n(s) ds.$$

Usado la regla de Leibnitz derivamos con respecto a t:

$$\lambda_n e'(t) = t e_n(t) + \frac{d1}{dt} t e_n(t) - \frac{dt}{dt} t e_n(t) + \int_t^1 e_n(s) ds = \int_t^1 e_n(s) ds.$$

Volviendo a derivar tenemos $\lambda_n e_n''(t) = -e_n(t)$, y vemos que $e_n(0)$ y $e_n'(1)$ deben ser 0.

Y la solución de este problema de valor inicial es

$$e_n(t) = \sqrt{2}\sin((n+1/2)\pi t) \text{ y } \lambda_n = ((n+1/2)\pi)^{-2},\ n = 0, 1, \ldots.$$

Entonces

$$B(t) = \sqrt{2}\sum_{n=0}^{\infty} \xi_n \frac{\sin((n+1/2)\pi t)}{(n+1/2)\pi}, \tag{21.15}$$

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

Ejercicios § 21

21.1. Demuestre la Proposición 27.5(c).

21.2. Sean X y $\mathrm{X}_n (n = 1, 2, \ldots)$ vv.aa. en L_2. Demuestre que si $\mathrm{X}_n \to \mathrm{X}$ en L_2, entonces $E\mathrm{X}_n \to E\mathrm{X}, E(\mathrm{X}_n^2) \to E(\mathrm{X}^2)$ y $\mathrm{Var}(\mathrm{X}_n) \to \mathrm{Var}(\mathrm{X})$.

21.3. Dé un ejemplo de un PE L_2–estacionario que no es L_2–continuo.

21.4. Demuestre que un PE L_2–diferenciable es L_2–continuo. Dé un ejemplo mostrando que el recíproco es falso en general.

21.5. Demuestre: si $\mathrm{X}_n \to \mathrm{X}$ en L_2 y las vv.aa. X_n son gaussianas, entonces también X es gaussiana.

21.6. Demuestre que el proceso de Poisson del Ejemplo 21.2 no es L_2–diferenciable.

21.7. Si $\mathrm{X}(\cdot)$ es L_2–estacionario y además es L_2–diferenciable en un punto, demuestre que entonces es L_2–diferenciable en todo punto.

21.8. Demuestre: (a) Si $\mathrm{X}(\cdot)$ es L_2–continuo sobre $[a, b]$, entonces la integral

$$Y(t) := \int_a^t \mathrm{X}(s)ds$$

es L_2–diferenciable y

$$\frac{d}{dt}Y(t) = \frac{d}{dt}\int_a^t \mathrm{X}(s)ds = \mathrm{X}(t) \quad \text{(en } L_2\text{)} \quad \forall\, a \le t \le b.$$

(b) Si $\mathrm{X}(t)$ es L_2–diferenciable para todo $t \in [a, b]$, entonces

$$\int_a^t \mathrm{X}'(s)ds = \mathrm{X}(t) - \mathrm{X}(a) \quad \forall\, a \le t \le b.$$

(c) De (b), $\int_a^t E\mathrm{X}'(s)ds = E\mathrm{X}(t) - E\mathrm{X}(a)$, es decir,

$$\int_a^t m_{\mathrm{X}'}(s)ds = m_{\mathrm{X}}(t) - m_{\mathrm{X}}(a) \quad \forall\, a \le t \le b,$$

por lo que

$$\frac{d}{dt}m_{\mathrm{X}}(t) = m_{\mathrm{X}'}(t).$$

21.9. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in \mathbb{R}\}$ un PE L_2–estacionario y L_2–diferenciable. Demuestre que $\mathrm{X}'(\cdot) = \{\mathrm{X}'(t),\, t \in \mathbb{R}\}$ es L_2–estacionario con función media cero y función de covariancia

$$r_{\mathrm{X}'}(t) = -\frac{d^2}{dt^2}r_{\mathrm{X}}(t) = -r''_{\mathrm{X}}(t) \quad \forall\, t \in \mathbb{R}.$$

Además, $\mathrm{X}(\cdot)$ y $\mathrm{X}'(\cdot)$ no están correlacionados. (*Sugerencia:* use el Ejercicio 7(c) y la Proposición 21.14.)

21.10. Sea $W(\cdot) = \{W(t),\, t \geq 0\}$ un proceso de Wiener y sea

$$\mathrm{X}(t) := \int_0^t W(s)ds \quad \forall\, t \geq 0.$$

Calcule (a) la función media y la función de covarianza de $\mathrm{X}(\cdot)$, y (b) $E[W(r)\mathrm{X}(t)]$ para $0 \leq r \leq t$.

21.11. Sea $W(\cdot)$ el proceso de Wiener del Ejercicio 20.12, y sea

$$Y(t) := \int_t^{t+1} [W(s) - W(t)]ds \quad \forall\, t \in \mathbb{R}.$$

Demuestre que $Y(\cdot)$ es débilmente estacionario y calcule su media y su función de covarianza.

21.12. Sea $\mathrm{X}(\cdot) \in L_2$ un PE con media cero y función de covarianza $K(s,t)$ continua sobre $[a,b] \times [a,b]$. Demuestre que si $g : [a,b] \to \mathbb{R}$ es continua, entonces

$$E[\mathrm{X}(s) \int_a^b g(t)\mathrm{X}(t)dt] = \int_a^b g(t)K(s,t)dt.$$

21.13. Sea $\{a_n\}$ una sucesión de números reales, y $\{Y_n\}$ una sucesión de vv.aa. i.i.d. con media cero y varianza $\sigma^2 < \infty$. Demuestre que si $\sum_{n=0}^\infty a_n^2 < \infty$, entonces $\sum_{n=0}^\infty a_n Y_n$ converge en L_2. (*Sugerencia:* use el Lema 27.9.)

22 Ecuaciones diferenciales en L_2

Contenido: Integral con respecto al proceso de Wiener, ecuación diferencial estocástica en L_2, la ecuación de Langevin, la ecuación de Ornstein–Uhlenbeck.

En esta sección estudiamos una familia de ecuaciones diferenciales estocásticas que formalmente se pueden escribir en la forma

$$a\mathrm{X}'(t) + b\mathrm{X}(t) = \xi(t), \tag{22.1}$$

en donde $\xi(\cdot) = \{\xi(t), t \in \mathbb{R}\}$ es un "ruido blanco gaussiano". Empezaremos con una breve discusión de este proceso.

Observación 22.1. (a) Análogo a la sección B del Capítulo 20 tenemos lo siguiente en tiempo continuo. Si $\mathrm{X}(\cdot)$ es un PE L_2–estacionario cuya función de covarianza satisface que

$$\int_{-\infty}^{\infty} |r_{\mathrm{X}}(t)| dt < \infty,$$

entonces la transformada de Fourier de r_{X}, es decir,

$$h_{\mathrm{X}}(\lambda) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} r_{\mathrm{X}}(t) dt \quad \forall\, \lambda \in \mathbb{R},$$

se llama la **función de densidad espectral** de $\mathrm{X}(\cdot)$. Si h_{X} es tal que

$$\int_{-\infty}^{\infty} |h_{\mathrm{X}}(\lambda)| d\lambda < \infty, \tag{22.2}$$

entonces la función de covarianza r_{X} se puede calcular mediante la transformada de Fourier inversa

$$r_{\mathrm{X}}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} h_{\mathrm{X}}(\lambda) d\lambda \quad \forall\, t \in \mathbb{R}. \tag{22.3}$$

(b) Un PE $\mathrm{X}(\cdot)$ L_2–estacionario con media 0 y función de densidad espectral constante, digamos

$$h_{\mathrm{X}}(\lambda) \equiv c \quad \forall\, \lambda \in \mathbb{R}, \tag{22.4}$$

en donde c es una constante positiva, se dice que es un **ruido blanco**. (El nombre se debe a la analogía con la "luz blanca", que contiene uniformemente todas las frecuencias de luz.) Sin embargo, en este caso $h_{\mathrm{X}}(\lambda)$ **no** satisface la condición de integrabilidad (2), así que $\mathrm{X}(\cdot)$ es un proceso que no existe en el sentido usual. De hecho, por (3), $\mathrm{X}(\cdot)$ tendría varianza infinita:

$$r_{\mathrm{X}}(0) = \mathrm{Var}[\mathrm{X}(t)] = \infty \quad \forall\, t \in \mathbb{R}. \tag{22.5}$$

(c) Si el proceso de Wiener $W(\cdot) = \{W(t), t \in \mathbb{R}\}$ fuera L_2–diferenciable (¡que **no** lo es!), su derivada $W'(\cdot)$ sería un **ruido blanco gaussiano**. En efecto, por el Ejercicio 21.6, $W'(\cdot)$ sería un proceso gaussiano, mientras que por el Ejemplo 21.8, $W'(\cdot)$ sería L_2–estacionario con media cero, pues

$$m_{W'}(t) = \frac{d}{dt} m_W(t) = 0 \quad \forall\, t \in \mathbb{R},$$

y función de covarianza

$$r_{W'}(t) = \mathrm{Cov}[W'(s), W'(t+s)] = \begin{cases} 0 & \text{si } \ t \neq 0, \\ \infty & \text{si } \ t = 0. \end{cases} \tag{22.6}$$

En particular, por el hecho de ser gaussiano, se tendría de (6) que $W'(s)$ y $W'(t)$ serían vv.aa. *independientes* para $s \neq t$. $\Box$

La Observación 22.1 pone de manifiesto que la ecuación diferencial estocástica (EDE) en (1) se debe interpretar de alguna manera adecuada si es que $\xi(\cdot) = W'(\cdot)$ es un ruido blanco gaussiano. Con este fin, considérese el espacio vectorial $C^1[a,b]$ que consiste de todas las funciones $f : [a,b] \to \mathbb{R}$ cuya derivada f' es continua en el intervalo $[a,b]$. Interpretamos W' como un "funcional" que a cada función $f \in C^1[a,b]$ le asocia la v.a. gaussiana

$$W'(f) := f(b)W(b) - f(a)W(a) - \int_a^b f'(t)W(t)dt. \tag{22.7}$$

De hecho, en este caso se acostumbra escribir $W'(f)$ como la integral

$$W'(f) \equiv \int_a^b f(t)dW(t), \tag{22.8}$$

y vemos entonces que el lado derecho de (7) resultaría de la "fórmula de integración por partes", i.e.

$$\int_a^b f(t)dW(t) = f(t)W(t)|_{t=a}^{t=b} - \int_a^b f'(t)W(t)dt. \tag{22.9}$$

Algunas veces escribiremos $\int_a^b f(t)dW(t)$ simplemente como

$$\int_a^b f\,dW.$$

Nótese, por otra parte, que (7) ó (9) se pueden expresar como

$$\int_a^b f\,dW = f(b)[W(b) - W(a)] - \int_a^b f'(t)[W(t) - W(a)]dt. \tag{22.10}$$

Esta fórmula es útil para algunos cálculos.

Ejemplo 22.2. Tomando $f(t) \equiv 1$ en (9) ó (10) vemos que

$$\int_a^b dW = W(b) - W(a).$$

Análogamente,

$$\int_0^1 t\,dW(t) = W(1) - \int_0^1 W(t)dt$$

y

$$\int_0^1 e^{\alpha t}\,dW(t) = e^{\alpha}W(1) - \alpha \int_0^1 e^{\alpha t}W(t)dt. \quad \square$$

Las fórmulas (12) y (13) en la siguiente proposición son casos especiales de las igualdades (14) y (15) que demostraremos en la Sección 24, en donde a $\int_a^b f\ dW$ se le llama la "integral de Ito de f". Por otra parte, una demostración de (12) y (13), basada directamente en (7) ó (9) se puede ver, por ejemplo, en el libro de Hoel, Port y Stone (1972), pp. 142–145. (Nótese que (11) se sigue trivialmente de (7). Además, (14) es consecuencia de (11) y de la independencia de incrementos de $W(\cdot)$; esto se deduce fácilmente usando la ecuación (10).)

Proposición 22.3. Sean f y g funciones en $C^1[a,b]$. Entonces

(a) $\int_a^b f dW$ es una v.a. gaussiana con media cero, i.e.

$$E\left(\int_a^b f\ dW\right) = 0, \tag{22.11}$$

y varianza

$$\mathrm{Var}\left(\int_a^b f\ dW\right) = E\left(\int_a^b f\ dW\right)^2 = \sigma^2 \int_a^b f^2(t)dt. \tag{22.12}$$

(b) Si $a \leq b \leq c$,

$$E\left[\int_a^b f\ dW \cdot \int_a^c g\ dW\right] = \sigma^2 \int_a^b f(t)g(t)dt. \tag{22.13}$$

(c) Si $a \leq b \leq c \leq d$,

$$E\left[\int_a^b f\ dW \cdot \int_c^d g\ dW\right] = 0. \tag{22.14}$$

Ejemplo 22.4. Sean

$$\mathrm{X} := \int_0^1 t\ dW(t) \quad \text{y} \quad Y := \int_0^1 e^{\alpha t} dW(t)$$

las integrales en el Ejemplo 22.2. Por la Proposición 22.3(a), X y Y son vv.aa. gaussianas con media cero y varianzas

$$E(\mathrm{X}^2) = \sigma^2 \int_0^1 t^2 dt = \sigma^2 t^3/3|_{t=0}^{t=1} = \sigma^2/3,$$

$$E(Y^2) = \sigma^2 \int_0^1 e^{2\alpha t} dt = \frac{\sigma^2}{2\alpha}(e^{2\alpha} - 1).$$

Asimismo, por (13), la covarianza de X y Y es

$$\begin{aligned} E(\mathrm{X}Y) &= E\left[\int_0^1 t dW(t) \cdot \int_0^1 e^{\alpha t} dW(t)\right] \\ &= \sigma^2 \int_0^1 t e^{\alpha t} dt \\ &= (\sigma/\alpha)^2 [1 + (\alpha - 1)e^\alpha]. \quad \square \end{aligned}$$

Ejemplo 22.5. Sea $\mathcal{F}_t^W := \sigma\{W(s), 0 \leq s \leq t\}$, para $t \geq 0$, la filtración natural del proceso de Wiener W. Además, si $f \in C^1[a, b]$, entonces para cada $a \leq t \leq b$ definimos

$$\mathrm{X}(t) := \int_a^t f\ dW. \tag{22.15}$$

Demuestre que $\mathrm{X}(\cdot)$ es una martingala con respecto a $\{\mathcal{F}_t^W, a \leq t \leq b\}$.

Demostración. Obviamente $\mathrm{X}(t)$ está en L_1 y, además, por (9) ó (10), $\mathrm{X}(t)$ es medible con respecto a $\mathcal{F}_t^W$. Por lo tanto, para demostrar que $\mathrm{X}(\cdot)$ es una martingala sólo falta verificar que

$$E[\mathrm{X}(t)|\mathcal{F}_s^W] = \mathrm{X}(s) \quad \forall\, a \leq s \leq t \leq b. \tag{22.16}$$

Con este fin, nótese primero que para $a \leq s \leq t \leq b$

$$\mathrm{X}(t) = \int_a^s f dW + \int_s^t f\ dW = \mathrm{X}(s) + \int_s^t f\ dW$$

Además, por (10),

$$\int_s^t f\ dW = f(t)[W(t) - W(s)] + \int_s^t f'(r)[W(r) - W(s)]dr.$$

así que la independencia de incrementos de $W(\cdot)$ implica que

$$\int_s^t f\ dW = \mathrm{X}(t) - \mathrm{X}(s)$$

es independiente de $\mathcal{F}_s^W$. De aqí se sigue que

$$E[\mathrm{X}(t) - \mathrm{X}(s)|\mathcal{F}_s^W] = E[\mathrm{X}(t) - \mathrm{X}(s)] = 0 \quad \text{(por (11))}$$

que es equivalente a (16). $\square$

Ahora consideremos la EDE (1) con $\xi(\cdot) \equiv W'(\cdot)$, es decir,

$$a\mathrm{X}'(t) + b\mathrm{X}(t) = W'(t) \quad \text{para} \quad t \geq t_0, \tag{22.17}$$

con una condición inicial $\mathrm{X}(t_0)$ dada, la cual puede ser aleatoria en L_2, digamos $\mathrm{X}(t_0) = \mathrm{X}_0$, o un punto $\mathrm{X}(t_0) = x_0$ en $\mathbb{R}$. La EDE (17) algunas veces se escribe en "forma diferencial"

$$a\ d\mathrm{X}(t) + b\mathrm{X}(t)dt = dW(t),$$

pero de hecho debe interpretarse en la **forma integral**

$$a[\mathrm{X}(t) - \mathrm{X}(t_0)] + b\int_{t_0}^t \mathrm{X}(r)dr = W(t) - W(t_0), \quad t \geq t_0. \tag{22.18}$$

Definición 22.6. Decimos que un PE $\mathrm{X}(\cdot)$ en L_2 es **solución** de la EDE (17) con condición inicial $\mathrm{X}(t_0)$ si $\mathrm{X}(\cdot)$ es un PE con trayectorias continuas y satisface (18).

Proposición 22.7. Sean a y b las constantes en (17) y $\alpha := -b/a$. Entonces la solución de (17) es

$$\mathrm{X}(t) = e^{\alpha(t-t_0)}\mathrm{X}(t_0) + \frac{1}{a}\int_{t_0}^{t} e^{\alpha(t-s)}\, dW(s), \quad t \geq t_0. \tag{22.19}$$

Más explícitamente, usando (9),

$$\begin{aligned} \mathrm{X}(t) &= e^{\alpha(t-t_0)}\left[\mathrm{X}(t_0) - \frac{1}{a}W(t_0)\right] \\ &\quad + \frac{1}{a}W(t) + \frac{\alpha}{a}\int_{t_0}^{t} e^{\alpha(t-s)}W(s)ds. \end{aligned} \tag{22.20}$$

Además, si $\mathrm{X}(t_0)$ es una constante (no aleatoria) o si $\mathrm{X}(t_0)$ es v.a. gaussiana independiente de $\{W(t), t \geq t_0\}$, entonces $\mathrm{X}(t)$ es una v.a. gaussiana.

Demostración. La solución de (17) se obtiene siguiendo básicamente los mismos pasos que para una ecuación diferencial ordinaria (determinista). En efecto, reescribamos (17) en la forma

$$\mathrm{X}'(t) - \alpha\mathrm{X}(t) = \frac{1}{a}W'(t) \quad \text{con} \quad \alpha := -b/a.$$

Multiplicando ambos miembros de esta ecuación por $e^{-\alpha t}$ se obtiene

$$e^{-\alpha t}[\mathrm{X}'(t) - \alpha\mathrm{X}(t)] = \frac{1}{a}e^{-\alpha t}W'(t)$$

o equivalentemente, por el Ejercicio 2,

$$\frac{d}{dt}[e^{-\alpha t}\mathrm{X}(t)] = \frac{1}{a}e^{-\alpha t}W'(t).$$

Luego, integrando de t_0 a t, del Ejercicio 21.8 (b) y de (8) obtenemos

$$e^{-\alpha t}\mathrm{X}(t) = e^{-\alpha t_0}\mathrm{X}(t_0) + \frac{1}{a}\int_{t_0}^{t} e^{-\alpha s}\, dW(s),$$

que a su vez da (19). $\square$

Ejemplo 22.8. La solución de la EDE

$$aX'(t) + bX(t) = W'(t),\ t \geq 0, \quad \text{con} \quad X(0) = x_0, \tag{22.21}$$

es (con $\alpha := -b/a$)

$$X(t) = x_0 e^{\alpha t} + \frac{1}{a} \int_0^t e^{\alpha(t-s)}\, dW(s). \tag{22.22}$$

Este es un proceso gaussiano con media

$$EX(t) = x_0 e^{\alpha t} \quad \forall\, t \geq 0$$

y la función de covarianza se obtiene usando (22) y el Ejercicio 3:

$$K_X(s,t) = \frac{\sigma^2}{2ab} \cdot [e^{\alpha|t-s|} - e^{\alpha(s+t)}] \quad \forall\, s,t \geq 0. \tag{22.23}$$

Con $s = t$ se obtiene la variancia

$$\operatorname{Var}[X(t)] = \frac{\sigma^2}{2ab} \cdot (1 - e^{2\alpha t}). \quad \square$$

Ejemplo 22.9. (la ecuación de Langevin). La solución $V(\cdot)$ de la ecuación de Langevin

$$mV'(t) + fV(t) = W'(t),\ t \geq 0, \quad \text{con} \quad V(0) = v_0 \in \mathbb{R}, \tag{22.24}$$

se llama el proceso de "velocidad" de Langevin. Las constantes m (masa) y f (fuerza) son positivas. De (21)–(23), con $\alpha := -f/m$,

$$V(t) = v_0 e^{\alpha t} + \frac{1}{m} \int_0^t e^{\alpha(t-s)} dW(s),\ t \geq 0,$$

es un proceso gaussiano con media $m_V(t) = v_0 e^{\alpha t}$ y función de covarianza

$$K_V(s,t) = \frac{\sigma^2}{2mf} [e^{\alpha|t-s|} - e^{\alpha(t+s)}].$$

Si $V(t) = X'(t)$ es la velocidad de una partícula cuya posición es $X(t)$, entonces podemos escribir (24) como la **ecuación de Ornstein–Uhlenbeck**

$$mX''(t) + fX'(t) = W'(t), \quad \text{con} \quad X(0) = x_0,\ X'(0) = v_0.$$

Para cada $t \geq 0$, $\mathrm{X}(t)$ se puede calcular integrando $\mathrm{X}'(t) = V(t)$ y se obtiene

$$\begin{aligned} \mathrm{X}(t) &= x_0 + \int_0^t V(r)dr \\ &= x_0 + \frac{v_0}{\alpha}(e^{\alpha t} - 1) + \frac{1}{m\alpha}\int_0^t [e^{\alpha(t-s)} - 1]dW(s). \end{aligned}$$

Ejercicios § 22

22.1. Sea $\mathrm{X}(\cdot)$ el PE definido en el Ejemplo 22.5, es decir

$$\mathrm{X}(t) := \int_a^t f\ dW.$$

Demuestre que $\mathrm{X}(\cdot)$ es L_2–continuo y continuo c.s. (Con respecto a continuidad c.s., recuerde el Teorema 16.10 y el Ejercicio 16.5.

22.2. Sea α una constante y $\mathrm{X}(\cdot)$ un PE L_2–diferenciable. Demuestre que la derivada en L_2 de $e^{-\alpha t}\mathrm{X}(t)$ es

$$\frac{d}{dt}\left[e^{-\alpha t}\mathrm{X}(t)\right] = e^{-\alpha t}[\mathrm{X}'(t) - \alpha \mathrm{X}(t)].$$

22.3. Sea α una constante y $\mathrm{X}(\cdot)$ el PE definido como

$$\mathrm{X}(t) := \int_0^t e^{\alpha(t-s)}dW(s) \quad \forall\, t \geq 0.$$

Demuestre que $\mathrm{X}(\cdot)$ tiene media cero y función de covarianza

$$K_{\mathrm{X}}(s,t) = \frac{\sigma^2}{2\alpha}\left[e^{\alpha(s+t)} - e^{\alpha(t-s)}\right] \quad \text{si} \quad 0 \leq s \leq t.$$

Por simetría, concluya que

$$K_{\mathrm{X}}(s,t) = \frac{\sigma^2}{2\alpha}\left[e^{\alpha(s+t)} - e^{\alpha|t-s|}\right] \quad \forall\, s,t \geq 0.$$

22.4. Considere la EDE (21) con condición inicial $\mathrm{X}(0) \in L_2$. Demuestre que en este caso, la solución de la EDE tiene media

$$E\mathrm{X}(t) = e^{\alpha t}E\mathrm{X}(0)$$

y función de covarianza

$$K_{\mathrm{X}}(s,t) = e^{\alpha(t+s)}\mathrm{Var}[\mathrm{X}(0)] + \frac{\sigma^2}{2ab}\left[e^{\alpha|t-s|} - e^{\alpha(s+t)}\right].$$

22.5. Encuentre la media y la función de covarianza del proceso definido por

(a) $\mathrm{X}(t) := \int_0^t r dW(r), \quad t \geq 0;$

(b) $\mathrm{X}(t) := \int_0^1 \cos tr \, dW(r), \quad t \in \mathbb{R};$

(c) $\mathrm{X}(t) := \int_{t-1}^t (t-r) dW(r), \quad t \in \mathbb{R}.$

22.6. Demuestre que el PE $\mathrm{X}(\cdot)$ definido como

$$\mathrm{X}(t) = \int_{t_0}^t e^{\alpha(t-s)} dW(s) \quad \text{para} \quad t \geq t_0,$$

se puede escribir como

$$\mathrm{X}(t) = W(t) - e^{\alpha(t-t_0)} W(t_0) + \alpha \int_{t_0}^t e^{\alpha(t-s)} W(s) ds.$$

22.7. Sea $Y(\cdot)$ un PE L_2–estacionario, y considérese la EDE

$$a\mathrm{X}'(t) + b\mathrm{X}(t) = Y(t) + W'(t) \quad \text{para} \quad t \geq 0.$$

Sea $\alpha := -b/a$. Demuestre que

$$\begin{aligned} \mathrm{X}(t) &= \mathrm{X}(0)e^{\alpha t} + \frac{1}{a}\int_0^t e^{\alpha(t-s)} Y(s) ds \\ &\quad + \frac{1}{a}\int_0^t e^{\alpha(t-s)} dW(s), \quad t \geq 0. \end{aligned}$$

(*Sugerencia*: compare con (17) y (19).)

22.8. Supóngase que $\mathrm{X}(\cdot) = \{\mathrm{X}(t), -\infty < t < \infty\}$ y $Y(\cdot) = \{Y(t), -\infty < t < \infty\}$ satisfacen las EDEs

$$a\mathrm{X}'(t) + b\mathrm{X}(t) = W'(t) \quad \text{y} \quad \alpha Y'(t) + \beta Y(t) = \mathrm{X}(t),$$

respectivamente. Demuestre que $Y(\cdot)$ satisface la EDE

$$a\alpha Y''(t) + (a\beta + b\alpha)Y'(t) + \beta b Y(t) = W'(t).$$

22.9. Integrales impropias en L_2. Sea $f : \mathbb{R} \to \mathbb{R}$ una función en $C^1(\mathbb{R})$ tal que

$$\int_{-\infty}^b f^2(t) dt < \infty \quad \forall\, b \in \mathbb{R}. \tag{22.25}$$

(a) Demuestre que existe una v.a. $I(b) \in L_2$ tal que

$$I(b) = \lim_{a \to -\infty} \int_a^b f(t) dW \quad \text{(límite en } L_2\text{)},$$

en cuyo caso escribimos

$$I(b) = \int_{-\infty}^b f(t)\, dW(t).$$

(*Sugerencia*: use la ecuación (13) con $f = g$ y el Lema 27.9.)

(b) $I(b)$ es una v.a. gaussiana con media $EI(b) = 0$ y varianza

$$\mathrm{Var} I(b) = \sigma^2 \int_{-\infty}^b f^2(t) dt.$$

(c) Si $g \in C^1(\mathbb{R})$ es otra función que satisface (25), entonces

$$E\left(\int_{-\infty}^b f dW \cdot \int_{-\infty}^c g dW \right) = \sigma^2 \int_{-\infty}^{\min(b,c)} f(t) g(t) dt.$$

(Compare con (13).)

22.10. Sea $\alpha < 0$ una constante y $Y(\cdot)$ el PE definido por

$$Y(t) := \int_{-\infty}^t e^{\alpha(t-r)} dW(r) \quad \text{para} \quad -\infty < t < \infty.$$

(a) Verifique la condición (25) en el Ejercicio 9, i.e.

$$\int_{-\infty}^t e^{2\alpha(t-r)} dr = -1/2\alpha < \infty.$$

(b) Demuestre que $Y(\cdot)$ es un proceso gaussiano con media 0 y función de covarianza $K_Y(s,t) = \frac{\sigma^2}{-2\alpha} e^{\alpha|t-s|}$; luego, $Y(\cdot)$ es L_2–estacionario con función de covarianza $r_Y(t) := K_Y(0,t) = \frac{\sigma^2}{-2\alpha} e^{\alpha|t|}\ \forall\, t \in \mathbb{R}$.

22.11. Sea $Y(\bullet)$ el PE definido en el Ejercicio 10, con $\alpha := -b/a < 0$, y sea

$$\hat{\mathrm{X}}(t) := \frac{1}{a}Y(t) = \frac{1}{a}\int_{-\infty}^{t} e^{\alpha(t-r)}dW(r) \quad \forall\ -\infty < t < \infty.$$

Nótese que

$$\hat{\mathrm{X}}(t) = e^{\alpha t}\hat{\mathrm{X}}(0)\ +\ \frac{1}{a}\int_{0}^{t} e^{\alpha(t-r)}dW(r) \quad \forall\, t \geq 0; \tag{22.26}$$

por lo tanto, $\hat{\mathrm{X}}(t)$ coincide con la solución (19) de la EDE (17) con $t_0 = 0$ y condición inicial $\hat{\mathrm{X}}(0)$. Además, por el Ejercicio 10, $\hat{\mathrm{X}}(\bullet)$ es un PE L_2–estacionario, gaussiano, con media 0 y función de covarianza

$$r_{\hat{\mathrm{X}}}(t) = \frac{-\sigma^2}{2\alpha a^2}e^{\alpha|t|} = -\frac{\sigma^2}{2ab}e^{\alpha|t|} \quad \forall\ -\infty < t < \infty.$$

Demuestre que $\hat{\mathrm{X}}(\cdot)$ es el **único** PE L_2–estacionario que satisface (17). (*Sugerencia*: suponga que existe otro PE $\mathrm{X}(\bullet)$ que satisface (26) y observe que $\mathrm{X}(\bullet)$ y $\hat{\mathrm{X}}(\bullet)$ coinciden si $\mathrm{X}(0) = \hat{\mathrm{X}}(0)$ con probabilidad 1.)

22.12. Sea $Y(\bullet) = \{Y(t), -\infty < t < \infty\}$ un PE L_2–estacionario y L_2–continuo y considere la EDE

$$a\mathrm{X}'(t) + b\mathrm{X}(t) = Y(t). \tag{22.27}$$

(a) Demuestre que, con $\alpha := -b/a$,

$$\mathrm{X}(t) = e^{\alpha(t-t_0)}\mathrm{X}(t_0) + \frac{1}{a}\int_{t_0}^{t} e^{\alpha(t-s)}Y(s)ds \quad \text{para} \quad t \geq t_0$$

es solución de (27) con condición inicial $\mathrm{X}(t_0)$. (*Sugerencia:* compare con (17) y (19).)

(b) Demuestre que si $\alpha = -b/a < 0$, entonces $\hat{\mathrm{X}}(t) := \frac{1}{a}\int_{-\infty}^{t} e^{\alpha(t-s)}Y(s)ds$ es la única solución L_2–estacionaria de (27). (Compare con el Ejercicio 11.)

23 Movimiento Browniano

Contenido: Movimiento Browniano, variación cuadrática.

En esta sección probamos la existencia del movimiento Browniano (MB), también llamado proceso de Wiener, y estudiamos algunas de sus propiedades.

La siguiente demostración es un refinamiento de la idea original de N. Wiener propuesto en 1961 por Z. Ciesielski.

Teorema 23.1. *Existe un espacio de probabilidad* $(\Omega, \mathcal{F}, P)$ *donde está bien definido el MB* $\{B_t,\ t \in [0,\infty)\}$.

Demostración. Parte I. Definimos primero las funciones de Haar de la siguiente manera: $H_1^{(0)} := 1,\ t \in [0,1]$, y para $n \geq 1$ y $k \in I(n) := \{k \leq 2^n,\ k \ \text{ impar}\} = \{1, 3, \ldots, 2^n - 1\}$

$$H_k^{(n)}(t) := \begin{cases} 2^{\frac{n-1}{2}} & t \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right) \\ -2^{\frac{n-1}{2}} & t \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) \\ 0 & e.o.c. \end{cases}$$

Ahora construimos las funciones de Shauder

$$S_k^{(n)}(t) := \int_0^t H_k^{(n)}(s)ds,\ t \in [0,1],\ n \geq 0,\ k \in I(n).$$

La función $S_k^{(n)}$ está centrada en $\frac{k}{2^n}$ y la altura máxima es $2^{-\frac{n+1}{2}}$, además no hay traslape para los diferentes indices $k \in I(n)$. Definamos pues

$$B_t^{(n)} := \sum_{m=0}^{n} \sum_{k \in I(m)} \xi_k^{(m)} S_k^{(m)}(t),\ t \in [0,1],\ n \geq 0,$$

donde $\{\xi_k^{(m)},\ m \geq 1,\ k \in I(m)\}$ es una colección de v.a. normales $(0,1)$ i.i.d. Mostremos que para ω fija $B^{(0)}(\omega), B^{(1)}(\omega), \ldots$ es una sucesión convergente en $C[0,1]$ con la norma del supremo.

Parte II. Tenemos que

$$P(|\xi_k^{(n)}| > x) = \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-u^2/2} du \leq \sqrt{\frac{2}{\pi}} \int_x^\infty \frac{u}{x} e^{-u^2/2} du = \sqrt{\frac{2}{\pi}} \frac{e^{-x^2/2}}{x},$$

luego $P(A_{nk}) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-n^2/2}}{n}$, donde $A_{nk} := \{|\xi_k^{(n)}| > n\}$.

Sea $D_n := \{\max_{k\in I(n)} |\xi_k^{(n)}| > n\}$, luego

$$P(D_n) = P\left(\bigcup_{k\leq 2^n \text{ impar}} A_{nk}\right) \leq 2^n \sqrt{\frac{2}{\pi}} \frac{e^{-n^2/2}}{n},$$

por lo que $\sum_{n=1}^{\infty} P(D_n) < \infty$. Por el Lema de Borel-Cantelli $P(\bigcap_{m=1}^{\infty} \bigcup_{i=m}^{\infty} D_i) = 0$.

Esto quiere decir que para casi toda $\omega \in \Omega$, hay $N(\omega) < \infty$ tal que $|\xi_k^{(n)}(\omega)| < n$ si $n \geq N(\omega)$. Por lo que

$$\left| \sum_{m=N(\omega)}^{\infty} \sum_{k\in I(m)} |\xi_k^{(m)}(\omega) S_k^{(m)}(t)| \right| \leq \sum_{m=N(\omega)}^{\infty} m 2^{-\frac{m+1}{2}} < \infty$$

para toda $t \in [0,1]$. Notemos ahora que para t fija, $S_k^{(m)}(t) \neq 0$ solo para un índice $k \in I(m)$, esto implica que para casi toda $\omega \in \Omega$, $B^{(n)}$ converge uniformemente en $t \in [0,1]$, por lo que $B := \lim_{n\to\infty} B^{(n)}$ está en $C[0,1]$ casi siempre. Ahora veamos que B es en efecto un MB.

Parte III. Observamos que $\{H_k^{(n)}, n \geq 0,\ k \in I(n)\}$ es una base ortonormal en $L_2[0,1]$. Para ver que es completa, supongamos que hay $f \in L_2[0,1]$ ortonormal a todo $H_k^{(n)}$. Entonces la función continua $F(t) := \int_0^t f(s)ds$ se anularía en 0 y 1, pues $f \perp H_1^{(0)}$; pero también se anularía en 1/2, pues $f\perp\ H_1^{(1)}$; y también en 1/4 y 3/4, pues $f\perp\{H_1^{(2)}, H_3^{(2)}\}$ y por lo anterior; y así sucecivamente se tendría que F se alenula en un conjunto denso, lo que implicaría que F es nula en todo $t \in [0,1]$. Tenemos pues que $(\forall f, g \in L_2[0,1])$

$$\langle f, g\rangle = \sum_{n=1}^{\infty} \sum_{k\in I(n)} \langle f, H_k^{(n)}\rangle \langle g, H_k^{(n)}\rangle.$$

En particular si $f(t) := I_{[0,t]}$ y $g(t) := I_{[0,s]}$,

$$\min(t,s) = \sum_{n=1}^{\infty} \sum_{k\in I(n)} \langle f, H_k^{(n)}\rangle \langle g, H_k^{(n)}\rangle. \tag{23.1}$$

Ahora, sean $0 = t_0 < t_1 < \ldots < t_n \leq 1$ y $\{\theta_1, \ldots, \theta_n\} \subset \mathbb{R}$ y $\theta_{n+1} = 0$;

tenemos que

$$
\begin{aligned}
E\left[e^{-i\sum_{j=1}^{n}(\theta_{j+1}-\theta_j)B^{(N)}_{t_j}}\right] &= E\left[\exp\left\{-i\sum_{m=0}^{N}\sum_{k\in I(m)}\xi_k^{(m)}\sum_{j=1}^{n}(\theta_{j+1}-\theta_j)S_k^{(m)}(t_j)\right\}\right]\\
&= \prod_{m=0}^{N}\prod_{k\in I(m)} e^{-\frac{1}{2}\left(\sum_{j=1}^{n}(\theta_{j+1}-\theta_j)S_k^{(m)}(t_j)\right)^2}
\end{aligned}
$$

$$
= \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n}(\theta_{j+1}-\theta_j)(\theta_{i+1}-\theta_i)\sum_{m=0}^{N}\sum_{k\in I(m)}S_k^{(m)}(t_i)S_k^{(m)}(t_j)\right]
$$

Usando (23.1), al tomar $N\to\infty$ llegamos a

$$
E\left[e^{i\sum_{j=1}^{n}\theta_j(B_{t_j}-B_{t_{j-1}})}\right] = E\left[e^{-i\sum_{j=1}^{n}(\theta_{j+1}-\theta_j)B_{t_j}}\right]
$$

$$
\begin{aligned}
&= \exp\left[-\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}(\theta_{j+1}-\theta_j)(\theta_{i+1}-\theta_i)t_j-\frac{1}{2}\sum_{j=1}^{n}(\theta_{j+1}-\theta_j)^2t_j\right]\\
&= \exp\left[-\sum_{j=1}^{n-1}(\theta_{j+1}-\theta_j)(-\theta_{j+1})t_j-\frac{1}{2}\sum_{j=1}^{n}(\theta_{j+1}^2-2\theta_{j+1}\theta_j+\theta_j^2)t_j\right]\\
&= \exp\left[\frac{1}{2}\sum_{j=1}^{n-1}(\theta_{j+1}^2-\theta_j^2)t_j-\frac{1}{2}\theta_n^2t_n\right]=\prod_{j=1}^{n}e^{-\frac{1}{2}\theta_j^2(t_j-t_{j-1})}.
\end{aligned}
$$

Por lo tanto $\{B_{t_j}-B_{t_{j-1}}\}_{j=1}^{n}$ son normales con media cero, varianzas t_j-t_{j-1} y son independientes. Por eso B es MB.

Finalmente, sean $X^{(1)},X^{(2)},\ldots$ una sucesión de MBs en $[0,1]$ independientes, y construyamos el MB en $[0,\infty)$ como

$$
B_t := X_1^{(1)}+\ldots+X_1^{(n-1)}+X_t^{(n)}
$$

para $t\in[n,n+1)$. □

Ahora veremos propiedades del MB. Para ello enunciamos algunos resultados básicos de teoría de martingalas en tiempo continuo.

Definición 23.2. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad.
i) $\overline{\mathcal{F}} := \{\mathcal{F}_t,\ t \geq 0\}$ es una filtración si $(\forall t > s \geq 0)$ $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$.
ii) Un proceso X está adaptado a $\overline{\mathcal{F}}$ si X_t es $\mathcal{F}_t$-medible.
iii) $\overline{\mathcal{F}}$ es continua por la derecha si $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$.
iv) $T \in \mathbb{R}^+$ es tiempo de paro c.r.a $\overline{\mathcal{F}}$ si $(\forall t \geq 0)$ $\{T \leq t\} \in \mathcal{F}_t$.

Observación 23.3. La colección $\{\mathcal{F}_{t+}\}$ es una filtración continua por la derecha.

Definición 23.4. Sea X un proceso estocásticos real valuado y $A \subset \mathbb{R}$. El primer tiempo de llegada a A es

$$H_A := \inf\{t \geq 0 : X_t \in A\},$$

y el primer paso a A es

$$T_A := \inf\{t > 0 : X_t \in A\}.$$

Proposición 23.5. Sea X un proceso con trayectorias continuas y $\overline{\mathcal{F}}$ la filtración generada por X, i.e. $\mathcal{F}_t := \sigma(\{X_s, s \leq t\})$. Si A es abierto o cerrado entonces H_A es tiempo de paro c.r.a $\overline{\mathcal{F}}$. Si A es cerrado, entonces T_A es tiempo de paro c.r.a $\{\mathcal{F}_{t+}\}$.

Definición 23.6. Sea T un tiempo de paro c.r.a una filtración $\overline{\mathcal{F}}$, definimos la siguiente colección de conjuntos

$$\mathcal{F}_T := \left\{A \in \mathcal{F} : A \bigcap \{T \leq t\} \in \mathcal{F}_t\right\}.$$

Nota 23.7. Se puede mostrar que $\mathcal{F}_T$ es una $\sigma-$álgebra.

Teorema 23.8. *Sea X una martingala uniformemente integrable, es decir*

$$\sup_t E\left[|X_t| I_{\{|X_t|>r\}}\right] \to 0 \text{ cuando } r \to \infty.$$

Entonces, hay $X_\infty \in L_1$ tal que $X_t \xrightarrow{L_1} X_\infty$ cuando $t \to \infty$. Además, $X_t = E[X_\infty | \mathcal{F}_t]$. Aqui $\{\mathcal{F}_t\}$ es la filtración generada por X.

Teorema 23.9. ***(paro opcional)*** *Sea X una martingala y S y T tiempos de paro con $S \leq T \leq \infty$ c.s. Entonces $X_S = E[X_T | \mathcal{F}_S]$.*

Teorema 23.10. *Sea B un MB. Entonces los siguientes procesos son martingalas:*
i) B_t, $t \geq 0$.
ii) $B_t^2 - t$, $t \geq 0$.
iii) $(\forall u \in \mathbb{C})$ $X_t := \exp(uB_t - \frac{1}{2}u^2 t)$, $t \geq 0$.

Demostración. iii) Como $E(\exp(uB_t)) = \exp(u^2 t/2)$, entonces $(\forall t \geq 0) E(X_t) = 1$, o sea que está en L_1. También

$$E\left(\exp\left\{uB(t+s) - \frac{u^2}{2}(t+s)\right\} | \mathcal{F}_t\right) = e^{uB_t - \frac{u^2}{2}t}.$$

□

Teorema 23.11. *El MB tiene la propiedad fuerte de Markov, i.e. para cualquier tiempo de paro T c.r.a la filtración natural $\overline{\mathcal{F}}$ y tal que $P(T < \infty) = 1$, se tiene que*

$$X_t := B_{T+t} - B_T, \ t \geq 0,$$

es un MB independiente de $\mathcal{F}_T$.

Demostración. Como $M_t := \exp(i\theta B_t + \frac{\theta^2}{2}t)$ es una martingala, por el teorema de paro opcional

$$E\left[\exp\left(i\theta B_{\min(S,n)+t} + \frac{\theta^2}{2}(\min(S,n)+t)\right) | \mathcal{F}_{\min(S,n)}\right]$$

$$= \exp\left(i\theta B_{\min(S,n)} + \frac{\theta^2}{2}\min(S,n)\right),$$

donde $n \geq 0$, S es un tiempo de paro finito c.r.a $\overline{F}$. Si $n \to \infty$, tenemos

$$E\left[\exp\left(i\theta(B_{S+t} - B_S)\right) | \mathcal{F}_S\right] = e^{-\theta^2 t/2}.$$

Así, si $0 = t_0 \leq t_1 \leq \ldots \leq t_n < \infty$, $\{\theta_1, \ldots, \theta_n\} \subset \mathbb{R}$ y $A \in \mathcal{F}_T$,

$$\begin{aligned}
& E\left[I_A \exp\left(i \sum_{k=1}^{n} \theta_k (X_{t_k} - X_{t_{k-1}})\right)\right] \\
= \;& E\left[I_A \exp\left(i \sum_{k=1}^{n} \theta_k (B_{T+t_k} - B_{T+t_{k-1}})\right)\right] \\
= \;& E\left[E\left[I_A \exp\left(i \sum_{k=1}^{n} \theta_k (B_{T+t_k} - B_{T+t_{k-1}})\right) | \mathcal{F}_{T+t_{n-1}}\right]\right] \\
= \;& E\left[I_A \exp\left(i \sum_{k=1}^{n-1} \theta_k (B_{T+t_k} - B_{T+t_{k-1}})\right)\right] e^{-\frac{\theta_n^2}{2}(t_n - t_{n-1})}
\end{aligned}$$

Al seguir aplicando este procedimiento llegamos a

$$E\left[I_A \exp\left(i \sum_{k=1}^{n} \theta_k (X_{t_k} - X_{t_{k-1}})\right)\right] = E[I_A] \prod_{k=1}^{n} e^{-\frac{\theta_k^2}{2}(t_k - t_{k-1})},$$

lo cual muestra lo deseado. □

Ahora enunciamos una lista de resultados útiles sobre el MB. Definimos $T_x := \inf\{t > 0 : B_t = x\}$ con $a < 0 < b$.

- Para $x \in (a, b)$ y $\tau := \min(T_a, T_b)$ se tiene $P_x(\tau < \infty | B_0 = x) = 1$ y $E[\tau | B_0 = x] < \infty$.
- También $P(T_b < \infty | B_0 = a) = 1$.
- Sea $M_t := \max_{0 \leq s \leq t} B_s$, entonces para toda $x > 0$.

  $$P(M_t \geq x | B_0 = 0) = 2P(B_t \geq x | B_0 = 0)$$

- $P(T_x \leq t) = \int_0^t \frac{|x|}{\sqrt{2\pi}} s^{-3/2} e^{-\frac{x^2}{2s}} ds$.
- Principio de reflexión. Sea T un tiempo de paro, entonces

  $$X_t := \begin{cases} B_t & t \leq T \\ 2B_T - B_t & t \geq T, \end{cases}$$

  es un MB.

- La densidad conjunta de (B_t, M_t) es

$$f_{(B,M)}(x,y) := \sqrt{\frac{2}{\pi}} \frac{2y-x}{t^{3/2}} \exp\left(\frac{(2y-x)^2}{2t}\right) I_{\{y\geq 0, x\leq y\}}.$$

- La ley arcoseno:

$$P((\forall s \in [a,b])B_s \neq 0) = \frac{2}{\pi} \arcsin \sqrt{a/b}.$$

Nos gustaría ahora definir la integral estocástica de un proceso X c.r.al MB: $\int X(s)dW(s)$. Sin embargo, por los siguientes resultados, veremos que esto no lo podemos llevar a cabo usando la integral de Stieljes, la cual pide que W sea una función de variación acotada.

Proposición 23.12. (**Variación cuadrática de** $W(\cdot)$.) Para cada $m = 1, 2, \ldots$, sea $\pi_m : a = t_0^m < t_1^m < \cdots < t_{n_m}^m = b$ una partición de $[a,b]$ con módulo $|\pi_m| := \max_i |t_{i+1}^m - t_i^m|$. Sea

$$S_m := \sum_{i=0}^{n_m-1} (\Delta W_i^m)^2 \equiv \sum_{i=0}^{n_m-1} [W(t_{i+1}^m) - W(t_i^m)]^2.$$

(a) Si $|\pi_m| \to 0$, entonces

$$S_m \to b - a \quad \text{en} \quad L_2. \tag{23.2}$$

(b) Si $\sum_m |\pi_m| < \infty$, entonces la convergencia en (23.2) se cumple **c.s.**

Nota 23.13. Tóme $[a,b] = [0,t]$. Entonces (23.2) implica que, cuando $|\pi_m| \to 0$, el siguiente límite existe

$$\langle W \rangle_t := \lim S_m = t \quad \textit{en probabilidad} \quad \forall\, t \geq 0.$$

Al PE $\{\langle W \rangle_t = t,\ t \geq 0\}$ se le llama la **variación cuadrática** de $W(\cdot)$ y tiene la propiedad de ser un proceso único, continuo, con trayectorias no–decrecientes y, además, $W^2(t) - \langle W \rangle_t$ es una martingala. Estos conceptos y observaciones son válidos si en lugar de $W(\cdot)$ tomamos *cualquier martingala de cuadrado integrable*. Por ejemplo, si $N(\cdot)$ es un proceso de Poisson con parámetro λ, entonces su variación cuadrática es

$$\langle N \rangle_t = EN(t) = \lambda t$$

porque $N(t) - \lambda t$ es martingala.

Demostración. Recuérdese que si $\mathrm{X} \sim N(0,1)$, entonces $E(\mathrm{X}^k) = 0$ para k impar y

$$E(\mathrm{X}^k) = \frac{k!}{2^{k/2}(k/2)!} \quad \text{si} \quad k \quad \text{es par.}$$

En particular, como

$$\Delta W_i := W(t_{i+1}) - W(t_i) \sim N(0, t_{i+1} - t_i) =: N(0, \Delta t_i),$$

vemos que

$$E(\Delta W_i)^2 = \Delta t_i \quad \text{y} \quad E(\Delta W_i)^4 = 3(\Delta t_i)^2. \tag{23.3}$$

De aquí se sigue que $E(S_m) = \sum_{i=0}^{n_m-1} \Delta t_i^m = b - a$, y

$$\mathrm{Var}(S_m) = E[S_m - (b-a)]^2;$$

luego, (23.2) es equivalente a: $\mathrm{Var}(S_m) \to 0$ cuando $m \to \infty$.

Ahora, como $W(\cdot)$ tiene incrementos independientes,

$$\mathrm{Var}(S_m) = \sum_{i=0}^{n_m-1} \quad \mathrm{Var}[(\Delta W_i^m)^2].$$

Pero

$$\begin{aligned} \mathrm{Var}[(\Delta W_i^m)^2] &= E(\Delta W_i^m)^4 - [E(\Delta W_i^m)^2]^2 \\ &= 3(\Delta t_i^m)^2 - (\Delta t_i^m)^2 \quad \text{(por (23.3))} \\ &= 2(\Delta t_i^m)^2 \leq 2 \cdot |\pi_m| \cdot \Delta t_i^m. \end{aligned}$$

Luego

$$\mathrm{Var}(S_m) \leq 2|\pi_m| \sum_{i=0}^{n_m-1} \Delta t_i^m = 2|\pi_m|(b-a) \tag{23.4}$$

y cuando $m \to \infty$ se obtiene (a).

Por otra parte, si $\sum_m |\pi_m| < \infty$, entonces, por (23.4),

$$\sum_m \mathrm{Var}(S_m) \leq 2(b-a) \sum_m |\pi_m| < \infty.$$

Por lo tanto, (b) se sigue del Ejercicio 24.3. $\square$

Teorema 23.14. *Casi ninguna trayectoria del MB W es de variación acotada.*

Demostración. Usando la notación anterior, tenemos que

$$\sum_{i=0}^{n_m-1}(\Delta W_i^m)^2 = \sum_{i=0}^{n_m-1}|\Delta W_i^m||\Delta W_i^m| \leq \max_i |\Delta W_i^m| \sum_{i=0}^{n_m-1}|\Delta W_i^m|.$$

Como el lado izquierdo converge a $b-a>0$ en L_2, entonces hay una subsucesión de $\{n_m\}$ donde la convergencia a $b-a$ es casi segura. Para esa subsucesión se tiene que $\max_i |\Delta W_i^m| \to 0$ c.s, pues las trayectorias del MB son continuas. Así, la única posibilidad para no tener contradicción es que cuando $m\to\infty$

$$\sum_{i=0}^{n_m-1}|\Delta W_i^m| \to \infty \ c.s.$$

$\square$

De lo anterior vemos que no podemos usar la integral de Stieljes para definir $\int X(s)dW(s)$ trayectorialmente, sin embargo al mostrar que W tiene variación cuadrática finita se ofrece una posibilidad para definir la integral; veamos el siguiente resultado artesanal para definir $\int W(s)dW(s)$.

Proposición 23.15. Sean $a=0$ y $b=t$. Se tiene que

$$\sum_{i=0}^{n_m-1} W(s_i^m)(W(t_{i+1}^m)-W(t_i^m)) \xrightarrow{L_2} \frac{W^2(t)}{2}-\frac{t}{2},\ m\to\infty,$$

donde $s_i^m := t_i^m$.

Demostración. Usando la identidad $\alpha(\beta-\alpha)=\frac{1}{2}(\beta^2-\alpha^2)-\frac{1}{2}(\beta-\alpha)^2$ tenemos que

$$\begin{aligned} & \sum_{i=0}^{n_m-1} W(t_i^m)(W(t_{i+1}^m)-W(t_i^m)) \\ = \ & \frac{1}{2}\sum_{i=0}^{n_m-1} W^2(t_{i+1}^m)-W^2(t_i^m)-\frac{1}{2}\sum_{i=0}^{n_m-1}(W(t_{i+1}^m)-W(t_i^m))^2. \end{aligned}$$

La primera es una suma telescópica, y la segunta converge a la variación cuadrática, cuando $m\to\infty$. $\square$

Proposición 23.16. En la proposition anterior, si $s_i^m := \gamma t_i^m + (1-\gamma)t_{i+1}^m$, con $\gamma \in [0,1]$, el límite es

$$\frac{W^2(t)}{2} - \frac{t}{2} + \gamma t.$$

24 La integral de Ito

Contenido: La integral de Ito, la isometría de Ito, propiedades de la integral de Ito.

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad. Un intervalo $[a, b]$, para $a < b < \infty$, siempre estará dotado con la σ–álgebra de Borel $\mathcal{B}[a, b]$, y la medida de Lebesgue $\lambda(dt) = dt$. Consideraremos el espacio usual $L_2(\Omega) \equiv L_2(\Omega, \mathcal{F}, \mathrm{P})$ y también el espacio $L_2([a, b] \times \Omega) \equiv L_2([a, b] \times \Omega,\, \mathcal{B}[a, b] \times \mathcal{F})$ con el producto interno

$$\langle f, g \rangle := E\left(\int_a^b f(t)g(t)dt\right) = \int_a^b E[f(t)g(t)]dt. \tag{24.1}$$

Luego, la norma en $L_2([a, b] \times \Omega)$ es

$$\| f \|_2 := \sqrt{\langle f, f \rangle} = \sqrt{\int_a^b E[f^2(t)]dt}. \tag{24.2}$$

Observación 24.1. Recuérdese que un PE $f(\cdot) = \{f(t),\, t \geq 0\}$ sobre $(\Omega, \mathcal{F}, \mathrm{P})$ es de hecho una función de dos variables $(t, \omega) \mapsto f(t, \omega)$, de $[a, b] \times \Omega \to \mathbb{R}$, tal que

- $f(t, \omega)$ es una v.a. sobre Ω para cada t fijo, y
- $f(t, \omega)$ es función de $t \geq 0$ para cada ω fijo (en este caso se dice que $t \mapsto f(t, \omega)$ es una **trayectoria** o **realización** del proceso $f(\cdot)$).

Usaremos la notación usual $f(t) \equiv f(t, \omega)$ ó $f_t \equiv f_t(\omega)$.

Considérese un proceso de Wiener o movimiento Browniano $W(\cdot) = \{W(t),\, t \geq 0\}$ **estándar**, es decir, con parámetro $\sigma^2 = 1$. Sea $\mathcal{F}^W = \{\mathcal{F}_t^W,\, t \geq 0\}$ la **filtración natural** de $W(\cdot)$, i.e. $\mathcal{F}_t^W := \sigma\{W(s),\, 0 \leq s \leq t\}$.

Definición 24.2. Para $0 \leq a < b$, sea $N[a, b]$ la familia de PEs $f(\cdot) = \{f(t),\, t \geq 0\}$ tales que

(a) $(t, \omega) \mapsto f(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$ es medible;

(b) $f(\cdot)$ está adaptado a $\mathcal{F}^W$ (i.e. $f(t)$ es $\mathcal{F}_t^W$–medible para cada $t \geq 0$);

(c) $f \in L_2([a,b] \times \Omega)$, es decir, $\int_a^b E[f^2(t)]dt = E[\int_a^b f^2(t)dt] < \infty$; ver (2).

Decimos que $f \in N[a,b]$ es un PE **escalonado** (o **elemental** o **simple**) si existe una partición $a = t_0 < t_1 < \cdots < t_n = b$ de $[a,b]$ tal que

$$f(t) \equiv f(t_i) \quad \forall\, t \in [t_i, t_{i+1}), \quad i = 0, 1, \ldots, n-1;$$

en otras palabras,

$$f(t) = \sum_{i=0}^{n-1} f(t_i) I_{[t_i, t_{i+1})}(t),$$

en donde, por convención, $[t_{n-1}, t_n) \equiv [t_{n-1}, b]$. Denotaremos por $\mathcal{E}[a,b]$ la subfamilia de PEs escalonados $f \in N[a,b]$.

Definición 24.3. (**La integral de Ito para PEs en** $\mathcal{E}[a,b]$.) Si $f \in \mathcal{E}[a,b]$ es escalonado, definimos la **integral de Ito** de f como

$$\int_a^b f(t)dW(t) := \sum_{i=0}^{n-1} f(t_i)\Delta W_i, \tag{24.3}$$

donde $\Delta W_i := W(t_{i+1}) - W(t_i)$. Algunas veces escribimos la integral de Ito como $\int_a^b f\, dW$ ó como $I(f)$.

Proposición 24.4. (Propiedades de $I(f)$ para $f \in \mathcal{E}[a,b]$) Sean f, g procesos en $\mathcal{E}[a,b]$ y α, β números reales. Entonces:

(a) $\int_a^b (\alpha f + \beta g)dW = \alpha \int_a^b f\, dW + \beta \int_a^b g\, dW$;

(b) la integral de Ito $\int_a^b f\, dW$ es una v.a. con media cero, i.e.

$$E\left(\int_a^b f\, dW\right) = 0, \tag{24.4}$$

y varianza

$$E\left(\int_a^b f\, dW\right)^2 = \int_a^b E[f^2(t)]dt = E\left[\int_a^b f^2(t)dt\right]. \tag{24.5}$$

Al resultado en (5) se le llama **isometría de Ito**. (Vea la Observación 24.5.)

Demostración. La demostración de (a) se deja como ejercicio para el lector. (Obsérvese que $\alpha f + \beta g$ es un PE en $\mathcal{E}[a,b]$.)

(b) En la demostración de (4) y (5) usaremos que por la independencia de incrementos de $W(\bullet)$, la v.a. $\Delta W_i := W(t_{i+1}) - W(t_i)$ *es independiente de* $\mathcal{F}_{t_i}^W$, que a su vez implica

$$E(\Delta W_i | \mathcal{F}_{t_i}^W) = E(\Delta W_i) = 0 \tag{24.6}$$

y

$$E[(\Delta W_i)^2 | \mathcal{F}_{t_i}^W] = E[(\Delta W_i)^2] = t_{i+1} - t_i. \tag{24.7}$$

Ahora, por (6),

$$\begin{aligned} E[f(t_i)\Delta W_i] &= E(E[f(t_i)\Delta W_i | \mathcal{F}_{t_i}^W]) \\ &= E(f(t_i) E[\Delta W_i | \mathcal{F}_{t_i}^W]) \\ &= 0 \quad \text{para} \quad i = 0, 1, \ldots, n-1, \end{aligned}$$

que junto con la definición (3) da la igualdad (4) porque

$$E\left[\int_a^b f\, dW\right] = \sum_{i=0}^{n-1} E[f(t_i)\Delta W_i] = 0.$$

Finalmente, para demostrar (5) usamos de nuevo la definición (3) de la integral de Ito para obtener

$$E\left(\int_a^b f\, dW\right)^2 = A + B, \tag{24.8}$$

donde

$$A := E\left[\sum_{i=0}^{n-1} f(t_i)^2 (\Delta W_i)^2\right]$$

y

$$B := 2\, E\left[\sum_{i<j} f(t_i) f(t_j) \Delta W_i \Delta W_j\right].$$

Usando (6) se puede ver que $B = 0$ porque, para $t_i < t_j$,

$$\begin{aligned} E[f(t_i)f(t_j)\Delta W_i \Delta W_j] &= E(E[\cdots|\mathcal{F}_{t_j}^W]) \\ &= E(f(t_i)f(t_j)\Delta W_i E[\Delta W_j|\mathcal{F}_{t_j}^W]) \\ &= 0. \end{aligned}$$

Por otra parte,

$$A = \sum_{i=0}^{n-1} E[f(t_i)^2(\Delta W_i)^2]$$

y como

$$\begin{aligned} E[f(t_i)^2(\Delta W_i)^2] &= E(E[\cdots|\mathcal{F}_{t_i}^W]) \\ &= E(f(t_i)^2 E[(\Delta W_i)^2|\mathcal{F}_{t_i}^W]) \\ &= E[f(t_i)^2](t_{i+1} - t_i) \quad \text{por (7)}, \end{aligned}$$

vemos que

$$A = \int_a^b E[f^2(t)]dt.$$

Sustituyendo estos resultados en (8) se obtiene la igualdad (5). □

Observación 24.5. Sean $(V_1, \|\cdot\|_1)$ y $(V_2, \|\cdot\|_2)$ dos espacios vectoriales normados con productos internos $\langle\cdot,\cdot\rangle_1$ y $\langle\cdot,\cdot\rangle_2$, respectivamente. Se dice que una función lineal $I : V_1 \to V_2$ es una **isometría** si para todo $v \in V_1$ se cumple que $\| I(v) \|_2 = \| v \|_1$. En este caso es fácil ver que

$$\langle v, v' \rangle_1 = \langle I(v), I(v') \rangle_2 \quad \forall\, v, v' \in V_1. \tag{24.9}$$

(Esta igualdad se puede obtener usando el análogo de la fórmula

$$x \cdot y = \frac{1}{2}[x^2 + y^2 - (x-y)^2]$$

para x, y en $\mathbb{R}$.) Ahora interpretemos la integral de Ito como una función $I : \mathcal{E}[a,b] \to L_2(\Omega)$ que a cada proceso f en $\mathcal{E}[a,b]$ le asocia la v.a. $I(f) = \int_a^b f\, dW$ definida en (3). Por la Proposición 24.4(a), la función I es lineal, mientras que la igualdad (5) establece que la norma de la v.a. $I(f) \in L_2(\Omega)$

coincide con la norma de $f \in L_2([a,b]\times\Omega)$ definida en (2). En otras palabras, $I : \mathcal{E}[a,b] \to L_2(\Omega)$ es en efecto una **isometría**.

Esto significa, en particular, que en lugar de (9) podemos escribir

$$\langle f, g\rangle = \langle I(f), I(g)\rangle \quad \forall\, f, g \in \mathcal{E}[a,b],$$

en donde $\langle f, g\rangle$ está dado por (1), mientras que

$$\langle I(f), I(g)\rangle = E[I(f)\cdot I(g)] = E\left[\int_a^b f\,dW \cdot \int_a^b g\,dW\right],$$

es decir,

$$E\left[\int_a^b f\,dW \cdot \int_a^b g\,dW\right] = E\left[\int_a^b f(t)g(t)dt\right] = \int_a^b E[f(t)g(t)]dt \quad (24.10)$$

para todo f, g en $\mathcal{E}[a,b]$.

Lema 24.6. El espacio $\mathcal{E}[a,b]$ **es denso en** $N[a,b]$ en la norma de $L_2([a,b]\times\Omega)$; es decir, para cada $f(\cdot) \in N[a,b]$ existe una sucesión $\{f_n\}$ en $\mathcal{E}[a,b]$ tal que

$$E\left(\int_a^b |f_n(t) - f(t)|^2 dt\right) \to 0 \quad \text{cuando} \quad n \to \infty. \quad (24.11)$$

Demostración. Caso 1: f es acotada y tiene trayectorias $t \mapsto f(t,\omega)$ continuas para cada $\omega \in \Omega$.

Como f es continua en el compacto $[a,b]$, es también uniformemente continua, entonces para cada $n = 1, 2, \ldots,$ podemos definir $\pi_n = \{a = t_0 < t_1 < \cdots < t_{m_n} = b\}$ una partición de $[a,b]$ de tal manera que $|f(t_i) - f(t)| \leq 1/n$ si $t \in [t_i, t_{i+1})$, y sea $f_n \in \mathcal{E}[a,b]$ la función

$$f_n(t) := \sum_{i=0}^{m_n} f(t_i) I_{[t_i,t_{i+1})}(t).$$

Así, se tiene que

$$\int_a^b |f_n(t) - f(t)|^2 dt = \sum_{i=0}^{m_n} \int_{t_i}^{t_{i+1}} |f_n(t) - f(t)|^2 dt \to 0$$

cuando $n \to \infty$. Esto implica (11), por el Teorema de Convergencia Acotada.

Caso 2: $f \in N[a,b]$ es acotada, es decir, existe una constante M tal que $|f(t,\omega)| \leq M$ para todo $(t,\omega) \in [a,b] \times \Omega$.

En este caso, por el Ejercicio 2 existe una sucesión de funciones $f_n \in N[a,b]$ tales que $t \mapsto f_n(t,\omega)$ es continua para todo $\omega \in \Omega$ y, además, $|f_n(t,\omega)| \leq M$ y $f_n(t,\omega) \to f(t,\omega)$ para todo $(t,\omega) \in [a,b] \times \Omega$. Esto implica (11), por el Teorema de Convergencia Acotada.

Caso 3: $f \in N[a,b]$ es arbitraria.

Para cada $n = 1, 2, \ldots$, sea $f_n \in N[a,b]$ la función **truncada**

$$\begin{aligned} f_n(t) &:= f(t) \quad \text{si} \quad |f(t)| \leq n \quad (\text{i.e. } -n \leq f(t) \leq n), \\ &:= n \quad \text{si} \quad f(t) > n, \\ &:= -n \quad \text{si} \quad f(t) < -n. \end{aligned}$$

Nótese que cada f_n es acotada (porque $|f_n| \leq n$), está dominada por f ($|f_n| \leq |f|$ para todo n), y $f_n(t) \to f(t)$ para todo $t \in [a,b]$ cuando $n \to \infty$. Esto implica (11), por el Teorema de Convergencia Dominada. □

Definición de la integral de Ito para $f \in N[a,b]$. Si $f \in N[a,b]$, por el Lema 24.6 existe una sucesión $\{f_n\}$ en $\mathcal{E}[a,b]$ que satisface (11). Entonces, por la propiedad de isometría, la sucesión de integrales $I(f_n) = \int_a^b f_n \, dW$ es una **sucesión de Cauchy** en $L_2(\Omega) = L_2(\Omega, \mathcal{F}, \mathrm{P})$ porque

$$\begin{aligned} E|I(f_n) - I(f_m)|^2 &= E|\int_a^b (f_n - f_m) dW|^2 \\ &= E[\int_a^b (f_n - f_m)^2 dt] \quad (\text{por } (5)) \\ &\to 0 \quad \text{cuando} \quad n, m \to \infty \quad (\text{por } (11)). \end{aligned}$$

Por lo tanto, existe una v.a. $I(f) \in L_2$ tal que $I(f_n) \to I(f)$ en L_2. Al límite $I(f)$ se le llama la **integral de Ito** de f y escribimos

$$I(f) := \int_a^b f(t) dW(t) = (L_2) \lim_{n\to\infty} \int_a^b f_n dW. \tag{24.12}$$

Propiedades de $I(f)$ para $f \in N[a,b]$:

(a) $I(f)$ es independiente de la sucesión aproximante $\{f_n\}$.

(b) Por (12) y el Ejercicio 4, la integral $I(f)$ satisface (4), (5) y (10), es decir, para todo f, g en $N[a,b]$:

$$EI(f) = E\left(\int_a^b f\,dW\right) = 0, \tag{24.13}$$

$$\mathrm{Var}(I(f)) = E\left(\int_a^b f\,dW\right)^2 = \int_a^b E[f^2(t)]dt, \tag{24.14}$$

$$\begin{aligned}\mathrm{Cov}\left(\int_a^b f\,dW, \int_a^b g\,dW\right) &= E\left(\int_a^b f\,dW \cdot \int_a^b g\,dW\right) \\ &= \int_a^b E[f(t)g(t)]dt. \end{aligned} \tag{24.15}$$

Relacionado con la Proposición 23.16 tenemos lo siguiente.

Definición 24.7. (**La integral de Stratonovich**) Si en (3) reemplazamos $f(t_i)$ por $f(t_i^*)$ con $t_i^* = \frac{1}{2}(t_i + t_{i+1})$, que es el *punto medio* de intervalo (t_i, t_{i+1}), la integral resultante se llama la integral de Stratonovich de f que denotamos por

$$S(f) \equiv \int_a^b f(t) \circ dW(t).$$

Ejemplo 24.8. Sea $\pi : a = t_0 < t_1 < \cdots < t_n = b$ una partición de $[a,b]$ y sea $\alpha \in [0,1]$ un número fijo. Para cada $i = 0, \ldots, n-1$, sea

$$t_i^* = (1-\alpha)t_i + \alpha t_{i+1}.$$

En particular, si $\boldsymbol{\alpha}\mathbf{=0}$, entonces $t_i^* = t_i$ es el extremo izquierdo del intervalo $[t_i, t_{i+1})$; si $\boldsymbol{\alpha}\mathbf{=1/2}$, entonces $t_i^* = \frac{1}{2}(t_i + t_{i+1})$ es el punto medio de $[t_i, t_{i+1})$. Sea

$$S_n := \sum_{i=0}^{n-1} W(t_i^*) \cdot \Delta W_i \quad [\text{con} \quad \Delta W_i := W(t_{i+1}) - W(t_i)].$$

En todo caso, usando la variación cuadrática se puede demostrar que para $\alpha \in [0,1]$

$$(L_2) \lim_{|\pi| \to 0} S_n = \frac{1}{2}[W^2(b) - W^2(a)] + (\alpha - \frac{1}{2})(b-a). \tag{24.16}$$

Por lo tanto, como $W(\cdot)$ está en $N[a,b]$, con $\alpha = 0$ obtenemos la integral de Ito

$$\int_a^b W\, dW = \frac{1}{2}[W^2(b) - W^2(a)] - \frac{1}{2}(b-a);$$

con $\alpha = 1/2$ obtenemos la integral de Stratonovich

$$\int_a^b W \circ dW = \frac{1}{2}[W^2(b) - W^2(a)].$$

De hecho, **para cualquier** $\alpha \in [0,1]$ podemos definir una integral estocástica

$$(\alpha) \int_a^b W\, dW = \frac{1}{2}[W^2(b) - W^2(a)] + (\alpha - \frac{1}{2})(b-1).$$

(La demostración de (24.16) se puede ver en el libro de Arnold (1974), pp. 58–61, o en el de Mikosch (1998), pp. 96–98, o de hecho en cualquier texto sobre ecuaciones diferenciales estocásticas.) $\square$

Definición 24.9. Si $f \in N[a,b]$ y $A \in \mathcal{B}[a,b]$, entonces $f \cdot I_A$ está en $N[a,b]$ y definimos

$$\int_A f(t) dW(t) := \int_a^b f(t) I_A(t) dW(t).$$

En particular, si $f \in N[0,T]$ para algún $T > 0$, entonces podemos definir la **integral indefinida**

$$\mathrm{X}(t) := \int_0^t f(s) dW(s) \quad \forall\, t \in [0,T]. \tag{24.17}$$

Proposición 24.10. Sea $f \in N[0,T]$ y $\mathrm{X}(\cdot)$ el proceso en (24.17). Entonces

(a) $\mathrm{X}(\cdot)$ está adaptado a $\mathcal{F}^W$, i.e. $\mathrm{X}(t)$ es $\mathcal{F}_t^W$–medible $\forall\, 0 \leq t \leq T$; y

(b) $\mathrm{X}(\cdot)$ es un proceso en L_2 y una martingala c.r.a. $\mathcal{F}^W$. Por lo tanto:

(c) $\mathrm{P}(\sup\limits_{0 \leq t \leq T} |\mathrm{X}(t)| > \varepsilon) \leq \frac{1}{\varepsilon^2} \int_0^T E[f^2(s)] ds$, y

(d) $E(\sup\limits_{0 \leq t \leq T} |\mathrm{X}(t)|^2) \leq 4 \int_0^T E[f^2(s)] ds$.

Demostración. (a) Para un PE escalonado $f \in \mathcal{E}[0,T]$, el inciso (a) se sigue de la definición (3); para $f \in N[a,b]$, (a) se sigue de la definición de $I(f)$ en (12).

(b) Por (a) y (13)–(14), la integral $\mathrm{X}(\cdot)$ está adaptada a $\mathcal{F}^W$, tiene media $E\mathrm{X}(t) = 0$ y varianza $\mathrm{Var}(\mathrm{X}(t)) = E\mathrm{X}^2(t) = \int_0^t E[f^2(s)]ds$. Luego, para demostrar (b) sólo falta verificar que $E[\mathrm{X}(t)|\mathcal{F}_s^W] = \mathrm{X}(s)$ para todo $0 \leq s \leq t \leq T$, ó equivalentemente

$$E[\mathrm{X}(t) - \mathrm{X}(s)|\mathcal{F}_s^W] = 0 \quad \forall\, 0 \leq s \leq t \leq T.$$

De hecho, como

$$\mathrm{X}(t) := \int_0^t f\,dW = \int_0^s f\,dW + \int_s^t f\,dW = \mathrm{X}(s) + \int_s^t f\,dW,$$

debemos probar que

$$E\left(\int_s^t f\,dW|\mathcal{F}_s^W\right) = 0. \tag{24.18}$$

Supóngase primero que $f \in \mathcal{E}[0,T]$. Entonces, para alguna partición $s = t_0 < t_1 < \cdots < t_n = t$ de $[s,t]$, tenemos

$$\int_s^t f\,dW := \sum_{i=0}^{n-1} f(t_i)\Delta W_i \quad (\text{con}\ \Delta W_i := W(t_{i+1}) - W(t_i));$$

luego

$$E\left(\int_s^t f\,dW|\mathcal{F}_s^W\right) = \sum_{i=0}^{n-1} E[f(t_i)\Delta W_i|\mathcal{F}_s^W].$$

Pero, como $\mathcal{F}_s^W \subset \mathcal{F}_{t_i}^W$,

$$\begin{aligned} E[f(t_i)\Delta W_i|\mathcal{F}_s^W] &= E[E(f(t_i)\Delta W_i|\mathcal{F}_{t_i}^W)|\mathcal{F}_s^W] \\ &= E[f(t_i)E(\Delta W_i|\mathcal{F}_{t_i}^W)|\mathcal{F}_s^W] \\ &= 0 \quad \text{por (6)}. \end{aligned}$$

Esto demuestra (21) para $f \in \mathcal{E}[0,T]$. De este hecho y de (12) se obtiene (21) para $f \in N[0,T]$ arbitraria.

Finalmente, por (b) y (14), los incisos (c) y (d) se siguen de las desigualdades para martingalas (vea la Nota 1 al final de esta sección.) $\square$

Definición 24.11. Sean $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ y $Y(\cdot) = \{Y(t),\, t \geq 0\}$ dos PEs sobre $(\Omega, \mathcal{F}, \mathrm{P})$. Decimos que $\mathrm{X}(\cdot)$ es una **versión** de $Y(\cdot)$ si

$$\mathrm{P}[\mathrm{X}(t) = Y(t)] = 1 \quad \forall\, t \geq 0.$$

Teorema 24.12. *(***Continuidad de la integral***.) Si $f \in N[0,T]$ y $\mathrm{X}(t) := \int_0^t f\, dW$, entonces existe una versión $Y(\cdot)$ de $\mathrm{X}(\cdot)$ que tiene trayectorias continuas c.s.*

Demostración. Caso 1: $f \in \mathcal{E}[0,T]$. Sea $0 = t_0 < t_1 < \cdots < t_n = T$ una partición de $[0,T]$, y para cada $t \in [0,T]$ sea $\hat{t} := \max\{t_i | t_i \leq t\}$. Entonces

$$\mathrm{X}(t) = \int_0^t f\, dW = \sum_{i=0}^{\hat{t}-1} f(t_i)\Delta W_i + f(\hat{t})[W(t) - W(\hat{t})].$$

Luego, por la continuidad de $W(\cdot)$, el proceso integral $\mathrm{X}(\cdot)$ es continuo.

Caso 2: $f \in N[0,T]$**.** Por el Lema 24.6 existe una sucesión $\{f_n\}$ en $\mathcal{E}[0,T]$ tal que

$$\lim_{n\to\infty} \int_0^T E|f_n(s) - f(s)|^2 ds = 0. \tag{24.19}$$

Sea $\mathrm{X}_n(t) := \int_0^t f_n\, dW$. Por la Proposición 24.10(c) y (24.19), cuando $n \to \infty$ tenemos

$$\mathrm{P}(\sup_{0\leq t\leq T} |\mathrm{X}_n(t) - \mathrm{X}(t)| > \varepsilon) \leq \frac{1}{\varepsilon^2} \int_0^T E|f_n(s) - f(s)|^2 ds \to 0.$$

Por lo tanto, existe una subsucesión $\{n_k\}$ de $\{n\}$ para la cual

$$\mathrm{P}(\sup_{0\leq t\leq T} |\mathrm{X}_{n_k}(t) - \mathrm{X}(t)| > 1/2^k) \leq \frac{1}{2^k} \quad \forall k$$

y por lo tanto

$$\sum_{k=1}^{\infty} \mathrm{P}\left(\sup_{0\leq t\leq T} |\mathrm{X}_{n_k}(t) - \mathrm{X}(t)| > \frac{1}{2^k}\right) \leq \sum_{k=1}^{\infty} (1/2)^k < \infty.$$

Luego, por el Lema de Borel–Cantelli (Ejercicio 5), para casi todo $\omega \in \Omega$ existe un entero $k(\omega)$ para el cual

$$\sup_{0\leq t\leq T} |\mathrm{X}_{n_k}(t,\omega) - \mathrm{X}(t,\omega)| \leq 1/2^k \quad \forall\, k \geq k(\omega).$$

Es decir, para casi todo ω (en otras palabras, c.s.) $\mathrm{X}(\cdot)$ es el límite uniforme de una sucesión de vv.aa. continuas y, por lo tanto, $\mathrm{X}(\cdot)$ es continuo. $\square$

La integral de Ito para procesos no–anticipantes

Definición 24.13. Sea $W(\cdot) = \{W(t),\, t \geq 0\}$ un proceso de Wiener sobre $(\Omega, \mathcal{F}, \mathrm{P})$, y sea $\mathcal{G}(\cdot) = \{\mathcal{G}_t,\, t \geq 0\}$ una familia de sub–σ–álgebras de $\mathcal{F}$. Decimos que $\mathcal{G}(\cdot)$ es **no–anticipante** con respecto a $W(\cdot)$ si:

(a) $\mathcal{G}(\cdot)$ es creciente (i.e. $\mathcal{G}_s \subset \mathcal{G}_t \;\forall\, 0 \leq s \leq t$);

(b) $W(\cdot)$ está adaptado a $\mathcal{G}(\cdot)$ (i.e. $W(t)$ es $\mathcal{G}_t$–medible $\forall\, t \geq 0$, o equivalentemente, $\mathcal{F}_t^W \subset \mathcal{G}_t$);

(c) Para todo $t \geq 0$ y $h > 0, W(t+h) - W(t)$ es independiente de $\mathcal{G}_t$.

Ejemplo 24.14. (a) La filtración natural $\mathcal{G}(\cdot) \equiv \mathcal{F}^W$ (i.e. $\mathcal{G}_t = \mathcal{F}_t^W$ para todo $t \geq 0$) es no–anticipante con respecto a $W(\cdot)$. De hecho, $\mathcal{F}^W$ es la familia "más pequeña" de filtraciones no–anticipantes.

(b) Sea C una v.a. independiente de $\mathcal{F}_t^W$ para todo $t \geq 0$, y para cada $t \geq 0$ sea $\mathcal{G}_t$ la mínima σ–álgebra que contiene a $\sigma\{C\}$ y $\mathcal{F}_t^W$. Entonces $\mathcal{G}(\cdot) = \{\mathcal{G}_t,\, t \geq 0\}$ es una filtración no–anticipante con respecto a $W(\cdot)$.

(c) Sean $W_1(\cdot), \ldots, W_d(\cdot)$ procesos de Wiener **independientes**, es decir, las σ–álgebras $\sigma\{W_i(t),\, t \geq 0\}$, para $i = 1, \ldots, d$, son independientes. Entonces el PE $W(\cdot) = \{W(t),\, t \geq 0\}$ con $W(t) := (W_1(t), \ldots, W_d(t))' \in \mathbb{R}^d$ se llama **proceso de Wiener d–dimensional**. (Si A es una matriz, A' denota su *transpuesta*.) Sea

$$\mathcal{G}_t \equiv \mathcal{F}_t^W := \sigma\{W(s),\ 0 \leq s \leq t\}$$

la mínima σ–álgebra que contiene a $\mathcal{F}_t^{W_i}$ para $i = 1, \ldots, d$. Entonces $\{\mathcal{G}_t,\, t \geq 0\}$ es no–anticipante con respecto a cada $W_i(\cdot)\,(i = 1, \ldots, d)$ y también no–anticipante con respecto a $W(\cdot) = (W_1(\cdot), \ldots, W_d(\cdot))'$.

Observación 24.15. Notación de matrices. Si $F = (f_{ij})$ es una matriz cuadrada, definimos su **traza** como

$$\mathrm{Tr}(F) := \sum_i F_{ii}.$$

Si $F = (f_{ij})$ es una matriz $m \times d$ definimos su **norma** como

$$|F| := [\mathrm{Tr}(FF')]^{1/2} = \left(\sum_{i=1}^{m} \sum_{j=1}^{d} f_{ij}^2 \right)^{1/2}.$$

Definición 24.16. La integral vectorial de Ito. Sea $W(\cdot) = (W_1(\cdot), \ldots, W_d(\cdot))'$ un proceso de Wiener d–dimensional, y $\mathcal{G}(\cdot) = \{\mathcal{G}_t,\, t \geq 0\}$ una familia **no–anticipante** con respecto a $W(\cdot)$. Sea $N[a,b]_{na}^{m\times d}$ ("na" significa no–anticipante) la familia de PEs **matriciales** $F(\cdot) = [f_{ij}(\cdot)] : [0,\infty) \times \Omega \to \mathbb{R}^{m\times d}$ tales que (como en la Definición 24.2):

(a) $(t,\omega) \mapsto f_{ij}(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$ es medible;

(b) $F(\cdot)$ está adaptado a $\mathcal{G}(\cdot)$ (i.e., cada componente $f_{ij}(t)$ es $\mathcal{G}_t$–medible para $t \geq 0$);

(c) $F(\cdot) \in L_2([a,b] \times \Omega)$, i.e. $E[\int_a^b |F(t)|^2 dt] < \infty$, en donde $|F(t)|$ es la norma de $F(t) = (f_{ij}(t))$ (vea la Observación 24.15).

Entonces definimos la integral de Ito $\int_a^b F(t)dW(t) \equiv \int_a^b F\, dW$ como el vector aleatorio cuya i–ésima componente ($i = 1, \ldots, m$) es

$$\left(\int_a^b F\, dW \right)_i := \sum_{j=1}^{d} \int_a^b f_{ij} dW_j;$$

es decir

$$\int_a^b F\, dW = \left(\sum_{j=1}^{d} \int_a^b f_{1j} dW_j, \ldots, \sum_{j=1}^{d} \int_a^b f_{mj} dW_j \right)'. \tag{24.20}$$

Cada una de las integrales $\int_a^b f_{ij} dW_j$ se define como en (12) pero sustituyendo $\mathcal{F}^W$ con $\mathcal{G}(\cdot)$. Por otra parte, si $m = d = 1$, escribimos $N[a,b]_{na}^{m\times d}$ como $N[a,b]_{na}$.

La integral vectorial (24.20) satisface **todas** las propiedades vistas en esta sección, con los cambios adecuados de notación. Por ejemplo, comparando con (13), el valor esperado de la integral es el *vector*.

$$E\left(\int_a^b F\, dW \right) = 0 \in \mathbb{R}^m. \tag{24.21}$$

Asimismo, en lugar de (14) ahora tenemos la **matriz de covariancia**

$$E\left[\left(\int_a^b F\,dW\right)\cdot\left(\int_a^b F\,dW\right)'\right] = \int_a^b E(FF')ds. \tag{24.22}$$

En particular,

$$E|\int_a^b F\,dW|^2 = \int_a^b E|F|^2 ds. \tag{24.23}$$

Definición 24.17. La integral de Ito para procesos $f \notin L_2([a,b]\times\Omega)$. En la Definición 24.16 considérese el caso escalar $m = d = 1$ pero **en lugar de (c)** ahora tenemos

(c') $\int_a^b f^2(t)dt < \infty$ c.s.

A esta nueva familia de PEs la denotaremos por $M[a,b]_{na}$ (en lugar de $N[a,b]_{na}$) y definimos la integral de Ito como sigue:

Paso 1. Si $f \in M[a,b]_{na}$ es un PE **escalonado**, digamos

$$f(t) = \sum_{i=0}^{n-1} f(t_i) I_{[t_i,t_{i+1})}(t),$$

definimos la integral de Ito como en (3), i.e.

$$\int_a^b f(t)dW(t) := \sum_{i=0}^{n-1} f(t_i)\Delta W_i \quad (\text{con } \Delta W_i := W(t_{i+1}) - W(t_i)).$$

Paso 2. (Los procesos escalonados son densos en $M[a,b]_{na}$ pero **c.s.**: compare (24.24) con (11).) Para cada $f \in M[a,b]_{na}$ existe una sucesión $\{f_n\}$ de procesos escalonados en $M[a,b]_{na}$ tal que

$$\int_a^b |f_n(t) - f(t)|^2 dt \to 0 \quad \text{c.s.} \tag{24.24}$$

[**Demostración:** Ver Arnold (1974), Lemma 4.4.5.]

Paso 3. Si f y $\{f_n\}$ son como en el Paso 2, entonces la sucesión de integrales $\int_a^b f_n dW$ converge **en probabilidad** (**no** en L_2, como en (12)) y definimos la **integral de Ito** de f como

$$\int_a^b f\,dW := \lim_{n\to\infty} \int_a^b f_n dW \quad \text{límite } \textbf{en probabilidad}. \tag{24.25}$$

La definición de la integral (24.25) en el **caso vectorial**, con $f(\cdot) \in \mathbb{R}^{m\times d}$ y $W(\cdot) \in \mathbb{R}^d$ se hace exactamente igual que en los Pasos 1,2,3; véase (24.20). (Los detalles pueden consultarse en el libro de Arnold (1974), Sección 4.4, por ejemplo.)

La definición (24.25) para procesos f en $M[a,b]_{na}$ es más general que (12) o que la Definición 24.16 en el sentido que $M[a,b]_{na}$ es un espacio mucho mayor que (es decir, que contiene a)

$$N[a,b]_{na} \supset N[a,b].$$

Sin embargo, esta generalidad tiene un precio alto; por ejemplo, en general, la integral en (24.25) **no** tiene las propiedades (13) y (14).

Notas § 24

1. Desigualdades para martingalas. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \geq 0\}$ una martingala en L_p (i.e. $E|\mathrm{X}(t)|^p < \infty$ para cada $t \geq 0$). Entonces

(a) Para todo $p \geq 1$, $T \geq 0$ y $\varepsilon > 0$:

$$\mathrm{P}\left(\sup_{0\leq t\leq T} |\mathrm{X}(t)| \geq \varepsilon\right) \leq \varepsilon^{-p} E|\mathrm{X}(T)|^p;$$

(b) Para $p > 1$ se cumple la *desigualdad maximal de Doob*:

$$E\left(\left(\sup_{0\leq t\leq T} |\mathrm{X}(t)|\right)^p\right) \leq \left(\frac{p}{p-1}\right)^p E|\mathrm{X}(T)|^p.$$

La demostración de (a) y (b) se puede ver, por ejemplo, en el libro de Karatzas y Shreve (1991), p. 13, Theorem 3.8.

Ejercicios § 24

24.1. Demuestre la Proposición 24.4(a).

24.2. Sea $f : [a,b] \to \mathbb{R}$ una función Borel–medible, acotada por una constante M, y para cada $n = 1, 2, \ldots,$ sea $f_n : [a,b] \to \mathbb{R}$ la función continua definida como

$$f_n(t) := n \int_a^t e^{-n(t-s)} f(s)ds \quad \text{para} \quad a \leq t \leq b.$$

Demuestre que la sucesión $\{f_n\}$ está acotada uniformemente por M y que $f_n(t) \to f(t)$ para λ–casi todo $t \in [a,b]$, donde λ es la medida de Lebesgue.

24.3. Sean $\mathrm{X}, \mathrm{X}_n\,(n = 1, 2, \ldots)$ vv.aa. y considere las proposiciones:

(a) $(\mathrm{X}_n - \mathrm{X}) \in L_2$ y $\sum_{n=1}^{\infty} E(\mathrm{X}_n - \mathrm{X})^2 < \infty$.

(b) $\sum_{n=1}^{\infty} \mathrm{P}(|\mathrm{X}_n - \mathrm{X}| \geq \varepsilon) < \infty$ para cada $\varepsilon > 0$. (En este caso se dice que $\{\mathrm{X}_n\}$ **converge completamente a** X.)

(c) $\mathrm{X}_n \to \mathrm{X}$ c.s.

Demuestre que (a) $\Rightarrow$ (b) $\Rightarrow$ (c).

24.4. Sean $\mathrm{X}, \mathrm{X}_n\,(n = 1, 2, \ldots)$ vv.aa. en L_2 tales que $\mathrm{X}_n \to \mathrm{X}$ en L_2. Demuestre que (a) $E\mathrm{X}_n \to E\mathrm{X}$, y (b) $E\mathrm{X}_n^2 \to E\mathrm{X}^2$.

24.5. Demuestre el **Lema de Borel–Cantelli**: Si $\{A_n\}$ es una sucesión de eventos tales que $\sum\limits_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$, entonces $\mathrm{P}(\limsup A_n) = 0$, donde $\limsup A_n := \bigcap\limits_{k=1}^{\infty} \bigcup\limits_{n=k}^{\infty} A_n$.

24.6. Demuestre que $\mathrm{X}(t) := W^3(t) - 3tW(t)$ es una martingala.

24.7. Use directamente la definición de la integral de Ito para demostrar que

$$\int_0^t s\,dW(s) = tW(t) - \int_0^t W(s)ds.$$

24.8. (La integral de Ito **no depende** de la partición intermedia $t_i^* \in [t_i, t_{i+1})$ si el PE $f \in N[0,T]$ es suficientemente "suave".) Sea $f \in N[0,T]$ un PE tal que, para algún $K < \infty$ y $\varepsilon > 0$, $\qquad E|f(s) - f(t)|^2 \leq K|s - t|^{1+\varepsilon} \quad \forall\, s, t \in [0,T]$. (*)

Demuestre que

$$\int_0^t f\,dW = (L_1) \lim_{|\pi| \to 0} \sum_{i=0}^{n-1} f(t_i^*) \Delta W_i$$

para cualquier selección de los puntos $t_i^* \in [t_i, t_{i+1}]$. En particular, las integrales de Ito y de Stratonovich coinciden: $\int_0^T f\,dW = \int_0^T f \circ dW$. (Por ejemplo, una función determinista $f \in C^1[a,b]$ es de Lipschitz y, por lo tanto, satisface $(*)$. Luego, en este caso, las integrales de Ito y Stratonovich coinciden con la integral $\int_a^b f\,dW$ definida en las ecuaciones (22.9) ó (22.10).)

24.9. Demuestre que para cada $n = 1, 2, \ldots$, existe una constante c_n tal que $E|W(t) - W(s)|^{2n} = c_n |t - s|^n$.

24.10. Demuestre que $W(t)/t \to 0$ en probabilidad cuando $t \to \infty$ (i.e. $\mathrm{P}(|W(t)/t| \geq \varepsilon) \to 0$ para cualquier $\varepsilon > 0$. (**Nota:** De hecho, $W(t)/t \to 0$ **c.s.** cuando $t \to \infty$. En efecto, si $n \leq t < n + 1$, escriba

$$W(t) = W(t) - W(n) + W(n) = W(t) - W(n) + \sum_{i=1}^{n} [W(i) - W(i-1)]$$

y note que $\frac{1}{n} \sum_{i=1}^{n} [W(i) - W(i-1)] \to 0$ c.s., por la Ley Fuerte de los Grandes Números.)

24.11. Sea $\alpha > 0$ una constante y $\mathrm{X}(t) := e^{\alpha W(t) - \alpha^2 t/2}$. Demuestre que

(a) $E\mathrm{X}(t) = 1 \; \forall \, t \geq 0$;

(b) $(\mathrm{X}(\cdot), \mathcal{F}^W)$ es una martingala; y

(c) $\mathrm{P}[\sup_{0 \leq s \leq t} (W(s) - \alpha s/2) > \beta] \leq e^{-\alpha\beta}$ $(\beta > 0)$.

(*Sugerencia:* En (b), considere $\mathrm{X}(t)/\mathrm{X}(s)$; en (c), use los incisos (a) y (b) y la desigualdad en la Nota 1(a) en el final de la sección.)

24.12. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), \, t \geq 0\}$ una martingala en L_2. Demuestre que

$$E[(\mathrm{X}(t) - \mathrm{X}(s))^2 | \mathcal{F}_s^{\mathrm{X}}] = E[\mathrm{X}^2(t) - \mathrm{X}^2(s) | \mathcal{F}_s^{\mathrm{X}}] \quad \forall \, 0 \leq s \leq t.$$

24.13. Sea $f : [0, \infty) \to \mathbb{R}$ una función **determinística** tal que $\int_0^t f^2(s) ds < \infty$ para $t \geq 0$. (Nótese que f está en $N[0, t]$ para cada $t \geq 0$; vea la Definición 24.2.) Sea $\mathrm{X}(t) := \int_0^t f(r) dW(r)$. Demuestre que los PEs

$$\mathrm{X}^2(t) - \int_0^t f^2(r) dr \quad \text{y} \quad \exp\left(\mathrm{X}(t) - \frac{1}{2} \int_0^t f^2(r) dr\right)$$

son martingalas con respecto a $\mathcal{F}_t^W$.

24.14. Demuestre que si f es un PE en $N[0,T]$, entonces la integral de Ito $\mathrm{X}(t) := \int_0^t f\,dW$ tiene **incrementos no–correlacionados**, i.e. $E[(\mathrm{X}(b) - \mathrm{X}(a))(\mathrm{X}(d) - \mathrm{X}(c))] = 0 \; \forall \; a \leq b \leq c \leq d$. Además, para $f, g \in N[0,T]$

$$E\left[\left(\int_0^t f\,dW\right) \cdot \left(\int_0^s g\,dW\right)\right] = \int_0^{\min(s,t)} E[f(r)g(r)]dr$$

(vea la igualdad (15)).

24.15. Demuestre que $\mathrm{X}(t) := W^2(t) - t$ es una martingala con respecto a $\mathcal{F}^W$.

24.16. Sea $\mathrm{X}(\cdot)$ la integral indefinida en (24.17). Demuestre que si el proceso f es acotado, entonces

$$M(t) := \mathrm{X}(t) - \int_0^t f^2(r)dr$$

es una martingala. Por lo tanto, por la nota en la Nota 23.13, la variación cuadrática de $\mathrm{X}(\cdot)$ es

$$\langle \mathrm{X} \rangle_t = \int_0^t f^2(r)dr.$$

(Una versión más general de este resultado aparece en el Ejercicio 25.7.)

25 La regla diferencial de Ito

Contenido: Diferencial estocástica, regla diferencial de Ito, fórmula de integración por partes.

Sea $\mathcal{G} = \{\mathcal{G}_t,\, t \geq 0\}$ una filtración no–anticipante con respecto a $W(\bullet)$ (vea la Definición 24.13). Sean $u(t,\omega),\ v(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$ dos PEs medibles, adaptados a $\mathcal{G}(\bullet)$ y tales que **c.p.1**:

$$\int_0^t |u(s)|\; ds < \infty \quad \text{y} \quad \int_0^t v^2(s)\; ds < \infty \quad \forall\, t \geq 0. \tag{25.1}$$

Nótese que, en otras palabras, $v(\bullet)$ pertenece a la familia $M[0,t]_{na}$ para todo $t \geq 0$. (Vea la Definición 24.17.)

Definición 25.1. (**Diferencial estocástica**) Si $\mathrm{X}(\bullet) := \{\mathrm{X}(t),\, 0 \leq t \leq T\}$ es un PE tal que

$$\mathrm{X}(b) - \mathrm{X}(a) = \int_a^b u(t)dt + \int_a^b v(t)dW(t) \quad \forall\, 0 \leq a \leq b \leq T, \tag{25.2}$$

decimos entonces que $\mathrm{X}(\bullet)$ es un **proceso de Ito** o que $\mathrm{X}(\cdot)$ tiene una **diferencial estocástica** en $[0,T]$ y escribimos

$$d\mathrm{X}(t) = u(t)dt + v(t)dW(t). \tag{25.3}$$

Ejemplo 25.2. (a) Del Ejemplo 24.8 sabemos que

$$\int_a^b W\, dW = \frac{1}{2}[W^2(b) - W^2(a)] - \frac{1}{2}(b-a).$$

Equivalentemente,

$$W^2(b) - W^2(a) = (b-a) + \int_a^b 2W(t)dW(t) = \int_a^b 1 \bullet dt + \int_a^b 2W(t)dW(t).$$

Por lo tanto, $\mathrm{X}(t) = W^2(t)$ tiene la diferencial estocástica

$$dW^2(t) = dt + 2W(t)dW(t). \tag{25.4}$$

(b) Sea $f \in C^1$ una función determinística. Luego, de (23.22.9) tenemos

$$\int_a^b f(t)dW(t) = f(b)W(b) - f(a)W(a) - \int_a^b f'(t)W(t)dt,$$

i.e
$$f(b)W(b) - f(a)W(a) = \int_a^b f'(t)W(t)dt + \int_a^b f(t)dW(t).$$
Por lo tanto, $\mathrm{X}(t) = f(t)W(t)$ tiene la diferencial estocástica
$$d[f(t)W(t)] = f'(t)W(t)dt + f(t)dW(t). \tag{25.5}$$
Como ejemplo de esta fórmula general, si $f(t) = t^n$ entonces
$$d[t^n W(t)] = nt^{n-1}W(t)dt + t^n dW(t). \quad \square$$

Las siguientes observaciones serán útiles más adelante.

Observación 25.3. (a) Considérese la regla formal:

$$\begin{array}{c|c|c} \cdot & dt & dW \\ \hline dt & 0 & 0 \\ \hline dW & 0 & dt \end{array} \tag{25.6}$$

de modo que todos los productos indicados en (6) son cero, excepto $(dW)^2 = dt$. Entonces, de (3) vemos que, formalmente,
$$(d\mathrm{X})^2 = u^2(dt)^2 + v^2(dW)^2 + 2uv(dt)(dW) = v^2\, dt;$$
es decir,
$$(d\mathrm{X}(t))^2 = v^2(t)dt. \tag{25.7}$$

(b) **Notación**. Denotaremos por $C^{1,2}([0,\infty) \times \mathbb{R})$, o simplemente $C^{1,2}$, el espacio de funciones $g(t,x)$ que son de clase C^1 en $t \geq 0$ y de clase C^2 en $x \in \mathbb{R}$. Las derivadas parciales se denotan con subíndices, por ejemplo
$$g_t := \partial g/\partial t, \quad g_x := \partial g/\partial x, \quad \text{etc.}$$
Sea $\mathrm{X}(\cdot)$ como en (3), y sea $g(t,x) \in C^{1,2}([0,\infty) \times \mathbb{R})$. En este caso, escribimos
$$(Lg)(t, \mathrm{X}(t)) := g_t(t, \mathrm{X}(t)) + g_x(t, \mathrm{X}(t))u(t) + \frac{1}{2} g_{xx}(t, \mathrm{X}(t))v^2(t). \quad \square \tag{25.8}$$

La demostración del siguiente teorema aparece al final de la sección.

Teorema 25.4. *(***Regla diferencial de Ito: caso escalar***.) Sea* $\mathrm{X}(\cdot)$ *como en (3), y sea* $g(t,x) \in C^{1,2}([0,\infty) \times \mathbb{R})$. *Entonces el PE* $Y(t) := g(t, \mathrm{X}(t))$ *tiene la diferencial estocástica (en forma "compacta")*

$$dY(t) = g_t(t, \mathrm{X}(t))dt + g_x(t, \mathrm{X}(t))d\mathrm{X}(t) + \frac{1}{2} g_{xx}(t, \mathrm{X}(t))(d\mathrm{X}(t))^2. \tag{25.9}$$

Es decir, por (3) y (7),

$$\begin{aligned} dY(t) &= [g_t(t, \mathrm{X}(t)) + g_x(t, \mathrm{X}(t))u(t) + \frac{1}{2} g_{xx}(t, \mathrm{X}(t))v^2(t)]\ dt \\ &\quad + g_x(t, \mathrm{X}(t))v(t)dW(t); \end{aligned} \tag{25.10}$$

o bien, usando (8)

$$dY(t) = Lg(t, \mathrm{X}(t))dt + g_x(t, \mathrm{X}(t))v(t)dW(t). \tag{25.11}$$

Ejemplo 25.5. (a). Use la regla de Ito para verificar la igualdad (5), con $f \in C^1$ determinística.

Solución. Sea $Y(t) := g(t, \mathrm{X}(t))$ con $g(t,x) = f(t)x$ y $X(\cdot) \equiv W(\cdot)$ (es decir, $u \equiv 0$ y $v \equiv 1$). Entonces

$$g_t(t,x) = f'(t)x, \quad g_x(t,x) = f(t), \quad \text{y} \quad g_{xx}(t,x) = 0.$$

Sustituyendo estos valores en (10) se obtiene (5).

(b) Verifique que para cualquier entero $n \geq 2$:

$$d[W^n(t)] = nW^{n-1}(t)dW(t) + \frac{1}{2} n(n-1)W^{n-2}(t)dt. \tag{25.12}$$

Solución. Sea $Y(t) = g(t, \mathrm{X}(t))$ con $\mathrm{X}(\cdot) \equiv W(\cdot)$ y $g(t,x) = x^n$. Entonces $g_t = 0$, $g_x = nx^{n-1}$ y $g_{xx} = n(n-1)x^{n-2}$. Luego, (12) se sigue de (10).

(c). Demuestre que la diferencial del PE $Y(t) = e^{(\alpha - \beta^2/2)t + \beta W(t)}$ es

$$dY(t) = \alpha Y(t)dt + \beta Y(t)dW(t).$$

En particular, si $\alpha = 0$, la diferencial de $Y(t) = e^{(-\beta^2/2)t + \beta W(t)}$ es

$$dY(t) = \beta Y(t)dW(t).$$

Solución. Tómese $Y(t) = g(t, \mathrm{X}(t))$ con $g(t,x) = e^{(\alpha-\beta^2/2)t+\beta x}$ y $\mathrm{X}(\cdot) \equiv W(\cdot)$. Entonces $g_t = (\alpha - \beta^2/2)g$, $g_x = \beta g$ y $g_{xx} = \beta^2 g$. Sustituyendo estos valores en (10) vemos que

$$dY(t) = [(\alpha - \beta^2/2)Y(t) + \frac{1}{2}\beta^2 Y(t)]dt + \beta Y(t)dW(t)$$

que se reduce al resultado deseado. $\square$

A continuación veremos la regla diferencial de Ito en el caso vectorial. Si Q es una matriz (en particular, un vector) denotaremos por Q' su transpuesta.

Caso vectorial. Sean $u(\cdot) = (u_1(\cdot), \ldots, u_m(\cdot))' \in \mathbb{R}^m$ y $v(\cdot) = (v_{ij}(\cdot)) \in \mathbb{R}^{m\times d}$ PEs cuyas componentes $u_i(\cdot)$ y $v_{ij}(\cdot)$ satisfacen (1). Ahora, sea $W(\cdot) = (W_1(\cdot), \ldots, W_d(\cdot))' \in \mathbb{R}^d$ un proceso de Wiener. En este caso, la diferencial estocástica (3), es decir, $d\mathrm{X}(t) = u(t)dt + v(t)dW(t)$, significa que las componentes de $\mathrm{X}(\cdot)$ satisfacen

$$d\mathrm{X}_i(t) = u_i(t)dt + \sum_{j=1}^{d} v_{ij}(t)dW_j(t) \quad \forall\, i = 1, \ldots, m. \tag{25.13}$$

Observación 25.6. Sea $g(t,x) : [0,\infty) \times \mathbb{R}^m \to \mathbb{R}$ una función de clase $C^{1,2}$. El **gradiente** de g es el vector "fila" $g_x = (g_{x_1}, \ldots, g_{x_m})$ y la **matriz Hessiana** es $g_{xx} = (g_{x_i x_j}) \in \mathbb{R}^{m\times m}$.

Sea $\sigma := vv' \in \mathbb{R}^{m\times m}$. Entonces las componentes de σ son

$$\sigma_{ij} = (vv')_{ij} = \sum_{k=1}^{d} v_{ik} v_{jk}. \tag{25.14}$$

Además,

$$\mathrm{Tr}(\sigma g_{xx}) = \mathrm{Tr}(g_{xx}\sigma) := \sum_{i=1}^{m} (\sigma g_{xx})_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_{ij} g_{x_i x_j},$$

ó bien, por (14),

$$\mathrm{Tr}(g_{xx}\sigma) = \sum_{k=1}^{d}\sum_{i,j=1}^{m} g_{x_i x_j} v_{ik} v_{jk}. \tag{25.15}$$

Notación. En lugar de (8) ahora tenemos, con $\sigma := vv'$,

$$\begin{aligned}(Lg)(t, \mathrm{X}(t)) &:= g_t(t, \mathrm{X}(t)) + g_x(t, \mathrm{X}(t))u(t) + \frac{1}{2}\mathrm{Tr}[\sigma(t)g_{xx}(t, \mathrm{X}(t))] \\ &= g_t + \sum_{i=1}^{m} g_{x_i}u_i + \frac{1}{2}\sum_{k=1}^{d}\sum_{i,j=1}^{m} g_{x_ix_j}v_{ik}v_{jk} \qquad (25.16)\end{aligned}$$

Teorema 25.7. *(***Regla diferencial de Ito: caso vectorial***). Sea* $\mathrm{X}(\cdot)$ *como en (13) y* $g(t, x) := [0, \infty) \times \mathbb{R}^m \to \mathbb{R}$ *como en la Observación 25.6. Entonces el PE* $Y(t) := g(t, \mathrm{X}(t))$ *tiene la diferencial estocástica (en forma "compacta")*

$$\begin{aligned}dY(t) &= g_t(t, Y(t))dt + g_x(t, \mathrm{X}(t))d\mathrm{X}(t) \\ &\quad + \frac{1}{2}\mathrm{Tr}[g_{xx}(t, \mathrm{X}(t))\sigma(t)]dt, \qquad (25.17)\end{aligned}$$

donde $\sigma := vv'$ *y* $\mathrm{Tr}(g_{xx}\sigma)$ *se puede escribir como en (15). Más explícitamente, sustituyendo (13) en (17) se obtiene*

$$\begin{aligned}dY(t) &= [g_t(t, Y(t)) + g_x(t, \mathrm{X}(t))u(t) + \frac{1}{2}\mathrm{Tr}(g_{xx}(t, \mathrm{X}(t))\sigma(t))]dt \\ &\quad + g_x(t, \mathrm{X}(t))v(t)dW(t), \qquad (25.18)\end{aligned}$$

o bien, usando (16),

$$dY(t) = Lg(t, \mathrm{X}(t))dt + g_x(t, \mathrm{X}(t))v(t)dW(t). \qquad (25.19)$$

Un caso de la regla de Ito aún más general que el del Teorema anterior, 25.7, es cuando también la función g es vectorial. En tal caso se tiene lo siguiente.

Observación 25.8. (caso general). Si $g : [0, \infty) \times \mathbb{R}^m \to \mathbb{R}^p$, entonces $Y(t) := g(t, \mathrm{X}(t))$ es un PE en $\mathbb{R}^p$ y en tal caso la ecuación (25.19) (ó la (17) ó la (18)) es una ecuación **vectorial** con componentes

$$dY_\ell(t) = (Lg_\ell)(t, \mathrm{X}(t))dt + \sum_{k=1}^{d}\sum_{i=1}^{m} \frac{\partial g_\ell(t, \mathrm{X}(t))}{\partial x_i} v_{ik}(t)dW_k(t) \qquad (25.20)$$

para $\ell = 1, \ldots, p$.

Ejemplo 25.9. Fórmula de integración por partes. Supóngase que $\mathrm{X}(\bullet) \in \mathbb{R}^2$ tiene la diferencial estocástica

$$d\mathrm{X}(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} dt + \begin{pmatrix} v_1(t) \\ v_2(t) \end{pmatrix} dW(t),$$

es decir, $d\mathrm{X}$ tiene componentes, $d\mathrm{X}_i(t) = u_i(t)dt + v_i(t)dW(t)$ para $i = 1, 2$. Demuestre que

$$d[\mathrm{X}_1(t)\mathrm{X}_2(t)] = \mathrm{X}_1(t)d\mathrm{X}_2(t) + \mathrm{X}_2(t)d\mathrm{X}_1(t) + v_1(t)v_2(t)dt.$$

Solución. Sea $Y(t) := \mathrm{X}_1(t)\mathrm{X}_2(t) = g(t, \mathrm{X}(t))$ con $g(t,x) = x_1x_2$ para todo $t \geq 0$ y $x = (x_1, x_2) \in \mathbb{R}^2$. Entonces

$$g_t = 0, \quad g_x = (x_2, x_1), \quad \text{y} \quad g_{xx} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Además

$$\sigma := vv' = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} (v_1, v_2) = \begin{pmatrix} v_1^2 & v_1v_2 \\ v_1v_2 & v_2^2 \end{pmatrix},$$

de modo que $\mathrm{Tr}[g_{xx}\sigma] = 2v_1v_2$. Sustituyendo estos resultados en (17) se obtiene la fórmula deseada. $\square$

Como caso especial del Ejemplo 25.9 cabe mencionar el siguiente:

$$d(W_1W_2) = W_1dW_2 + W_2dW_1 + dt,$$

i.e.

$$W_1(t)W_2(t) = \int_0^t W_1dW_2 + \int_0^t W_2dW_1 + t \quad \text{para} \quad t \geq 0.$$

Ejemplo 25.10. Suponga que $\mathrm{X}(\bullet) \in \mathbb{R}^m$ satisface (13) y sea $Y(t) = |\mathrm{X}(t)|^2$, i.e. $Y(t) = g(t, \mathrm{X}(t))$ con $g(t,x) = x_1^2 + \cdots + x_m^2$. Entonces $g_t = 0$, $g_{x_i} = 2x_i$ y $g_{x_ix_j} = 2\delta_{ij}$. Por lo tanto, (17) se reduce a

$$d|\mathrm{X}(t)|^2 = 2\sum_{i=1}^m \mathrm{X}_i(t)d\mathrm{X}_i(t) + |v(t)|^2dt,$$

donde $|v(t)|^2 = \mathrm{Tr}(vv') = \sum_{i,j} v_{ij}(t)^2$. En particular, tomando $u_i(\cdot) \equiv 0$ y $v_{ij}(\cdot) \equiv \delta_{ij}$ obtenemos

$$d[W_1(t)^2 + \cdots + W_m(t)^2] = 2\sum_{i=1}^{m} W_i(t)dW_i(t) + m\,dt.$$

Análogamente, si $Y(t) := |\mathrm{X}(t)| > 0$ c.p.1, tomando $g(t,x) := |x| = (x_1^2 + \cdots + x_m^2)^{1/2}$ vemos que $g_t = 0$, $g_{x_i} = x_i/|x|$ y $g_{x_i x_j} = \delta_{ij}/|x| - x_i x_j/|x|^3$. Por lo tanto, por (17),

$$d|\mathrm{X}(t)| = \frac{1}{|\mathrm{X}(t)|}\sum_{i=1}^{m} \mathrm{X}_i d\mathrm{X}_i + \frac{1}{2}\left[\frac{|v(t)|^2}{|\mathrm{X}(t)|} - \frac{1}{|\mathrm{X}(t)|^3}\sum_{i,j=1}^{m} \mathrm{X}_i(t)\mathrm{X}_j(t)\sigma_{ij}(t)\right]dt$$

con $\sigma(t) = (\sigma_{ij}(t)) = v(t)v'(t)$. En particular, si $\mathrm{X}(\cdot) \equiv W(\cdot)$ es el proceso de Wiener m–dimensional, entonces $u_i(\cdot) \equiv 0$, $v_{ij}(\cdot) \equiv \delta_{ij}$ y $\sigma = vv' = I$. Luego

$$d|W(t)| = \frac{1}{|W(t)|}\sum_{i=1}^{m} W_i\,dW_i + \frac{m-1}{2|W(t)|}dt,$$

y se sigue que el PE $Y(t) := |W(t)|$, llamado el **proceso de Bessel m–dimensional**, satisface que

$$Y(t)\,dY(t) = \sum_{i=1}^{m} W_i\,dW_i + \frac{m-1}{2}dt \quad \forall\, m \geq 2. \quad \Box$$

Ejemplo 25.11. (Caso vectorial — ver (20).) Calcule la diferencial estocástica de

$$Y(t) = (W_1(t) + W_2(t) + W_3(t),\ W_2(t)^2 - W_1(t)W_3(t))' \in \mathbb{R}^2.$$

Solución. Primero escribimos $Y(t) = (g_1(t,\mathrm{X}(t)), g_2(t,\mathrm{X}(t)))'$ con $\mathrm{X}(\cdot) \equiv W(\cdot) \in \mathbb{R}^3$ (i.e. $d\mathrm{X} = u\;dt + v\,dW$ con $u(\cdot) \equiv 0 \in \mathbb{R}^3$, $v(\cdot) \equiv I \in \mathbb{R}^{3\times 3}$) y $g_i(t,x) : [0,\infty) \times \mathbb{R}^3 \to \mathbb{R}$ definida por $g_1(t,x) := x_1 + x_2 + x_3$, $g_2(t,x) := x_2^2 - x_1x_3$.

Por lo tanto, para $i,j = 1,2,3 : \partial g_1/\partial x_i = 1$ y $\partial^2 g_1/\partial x_i \partial x_j = 0$. Luego, por (20), $Y_1(t) = g_1(t,\mathrm{X}(t)) = W_1(t) + W_2(t) + W_3(t)$ tiene la diferencial estocástica

$$dY_1 = d\left(\sum_{i=1}^{3} W_i\right) = \sum_{i=1}^{3} dW_i.$$

Análogamente, el gradiente de g_2 con respecto a x es el vector (fila)

$$(\partial g_2/\partial x_1,\ \partial g_2/\partial x_2,\ \partial g_2/\partial x_3) = (-x_3, 2x_2, -x_1)$$

y la matriz Hessiana es

$$g_{xx} = (\partial^2 g_2/\partial x_i \partial x_j) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

De aquí se sigue que, por (20),

$$dY_2(t) = dt + (-W_3, 2W_2, -W_1)dW = dt - W_3 dW_1 + 2W_2 dW_2 - W_1 dW_3.$$

Resumiendo, la diferencial de $Y(t)$ es

$$dY(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} dt + \begin{pmatrix} 1 & 1 & 1 \\ -W_3 & 2W_2 & -W_1 \end{pmatrix} \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \end{pmatrix}. \quad \square$$

Demostración. (**del Teorema 25.4**, Regla de Ito en el caso escalar.) Se desea demostrar que $Y(\cdot)$ tiene la diferencial estocástica (11), i.e.

$$Y(t) - Y(t_0) = \int_{t_0}^{t} Lg(s, \mathrm{X}(s))ds + \int_{t_0}^{t} g_x(s, \mathrm{X}(s))v(s)dW(s) \quad \forall\, 0 \le t_0 \le t,$$

con Lg como en (8).

Caso 1. $u(t,\omega) \equiv u(\omega)$ y $v(t,\omega) \equiv v(\omega)$ son vv.aa. independientes de t (es decir, son constantes) y $\mathcal{G}_{t_0}$–medibles. Sea $\Pi : t_0 < t_1 < \cdots < t_n = t$ una partición de $[t_0, t]$. Usaremos la siguiente notación:

$$\Delta t_i := t_{i+1} - t_i,\ \Delta \mathrm{X}_i := \mathrm{X}(t_{i+1}) - \mathrm{X}(t_i) = u\Delta t_i + v\Delta W_i$$

con $\Delta W_i := W(t_{i+1}) - W(t_i)$. Nótese que

$$(\Delta \mathrm{X}_i)^2 = u^2(\Delta t_i)^2 + v^2(\Delta W_i)^2 + 2uv(\Delta t_i)(\Delta W_i). \tag{25.21}$$

Además, escribimos

$$Y(t) - Y(t_0) = \sum_{i=0}^{n-1}[Y(t_{i+1}) - Y(t_i)] = \sum_{i=0}^{n-1} \Delta g_i$$

donde

$$\Delta g_i := Y(t_{i+1}) - Y(t_i) = g(t_{i+1}, \mathrm{X}(t_{i+1})) - g(t_i, \mathrm{X}(t_i)).$$

Como $g \in C^{1,2}$, por la fórmula de Taylor existen números $0 < \Theta_i, \hat{\Theta}_i < 1$ tales que

$$\begin{aligned} \Delta g_i &= g_t(t_i + \Theta_i \Delta t_i, \mathrm{X}(t_i))\Delta t_i + g_x(t_i, \mathrm{X}(t_i))\Delta \mathrm{X}_i \\ &\quad + \frac{1}{2} g_{xx}(t_i, \mathrm{X}(t_i) + \hat{\Theta}_i \Delta \mathrm{X}_i)(\Delta \mathrm{X}_i)^2. \end{aligned} \tag{25.22}$$

Por la continuidad de g_t y de $\mathrm{X}(\cdot)$, cuando $|\Pi| \to 0$ tenemos

$$\sum_{i=0}^{n-1} g_t(t_i, \mathrm{X}(t_i))\Delta t_i \to \int_{t_0}^t g_t(s, \mathrm{X}(s))\, ds \quad \text{c.p.1} \tag{25.23}$$

y

$$\alpha_n := \sup_{1 \le i \le n} |g_t(t_i + \Theta_i \Delta t_i, \mathrm{X}(t_i)) - g_t(t_i, \mathrm{X}(t_i))| \to 0 \quad \text{c.p.1.}$$

Esto implica que

$$\begin{aligned} & \left| \sum_{i=0}^{n-1} g_t(t_i + \Theta_i \Delta t_i, \mathrm{X}(t_i))\Delta t_i - \sum_{i=0}^{n-1} g_t(t_i, \mathrm{X}(t_i))\Delta t_i \right| \\ \le \; & \alpha_n \sum_{i=0}^{n-1} \Delta t_i \\ = \; & \alpha_n \cdot (t - t_0) \to 0 \quad \text{c.p.1 cuando} \quad |\Pi| \to 0. \end{aligned}$$

Luego, por (23), el **primer término** en el lado derecho de (22) satisface:

$$\sum_{i=0}^{n-1} g_t(t_i + \Theta_i \Delta t_i, \mathrm{X}(t_i))\Delta t_i \to \int_{t_0}^t g_t(s, \mathrm{X}(s)) ds \quad \text{c.p.1.} \tag{25.24}$$

Análogamente, por la continuidad de g_x y $\mathrm{X}(\cdot)$, cuando $|\Pi| \to 0$ tenemos

$$\sum_{i=0}^{n-1} g_x(t_i, \mathrm{X}(t_i))\Delta t_i \to \int_{t_0}^t g_x(s, \mathrm{X}(s)) ds \quad \text{c.p.1}$$

y, por la convergencia en (24.24.25),

$$\sum_{i=0}^{n-1} g_x(t_i, \mathrm{X}(t_i))\Delta W_i \to \int_{t_0}^t g_x(s, \mathrm{X}(s)) dW(s) \quad \text{en probabilidad.}$$

Por lo tanto, el **segundo término** en el lado derecho de (22) satisface:

$$\sum_{i=0}^{n-1} g_x(t_i, \mathrm{X}(t_i))\Delta \mathrm{X}_i = u \cdot \sum_{i=0}^{n-1} g_x(t, \mathrm{X}(t_i))\Delta t_i + v \cdot \sum_{i=0}^{n-1} g_x(t_i, \mathrm{X}(t_i))\Delta W_i$$

$$\to \int_{t_0}^{t} u \cdot g_x(s, \mathrm{X}(s))ds + \int_{t_0}^{t} v g_x(s, \mathrm{X}(s))dW(s) \tag{25.25}$$

en probabilidad cuando $|\Pi| \to 0$.

Ahora usaremos (21) para estimar el **tercer término** en el lado derecho de (22). Primero nótese que por la continuidad de g_{xx} y de $\mathrm{X}(\cdot)$, cuando $|\Pi| \to 0$ tenemos:

$$\beta_n := \max_{1\leq i\leq n} |g_{xx}(t_i, \mathrm{X}(t_i) + \hat{\Theta}_i \Delta \mathrm{X}_i) - g_{xx}(t_i, \mathrm{X}(t_i))| \to 0 \quad \text{c.p.1}, \tag{25.26}$$

$$\sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))\Delta t_i \to \int_{t_0}^{t} g_{xx}(s, \mathrm{X}(s))ds \quad \text{c.p.1}, \tag{25.27}$$

y, por la variación cuadrática de $W(\cdot)$,

$$\sum_{i=0}^{n-1} (\Delta W_i)^2 \to t - t_0 \quad \text{en} \quad L_2 \quad \text{(por lo tanto, en probabilidad).} \tag{25.28}$$

En particular, por (28) y (21),

$$\begin{aligned} \sum_{i=0}^{n-1} (\Delta \mathrm{X}_i)^2 \quad &= \quad u^2 \sum_{i=0}^{n-1} (\Delta t_i)^2 + v^2 \sum_{i=0}^{n-1} (\Delta W_i)^2 + 2uv \sum_{i=0}^{n-1} (\Delta t_i)(\Delta W_i) \\ &\to v^2(t - t_0) \quad \text{en probabilidad} \end{aligned}$$

y combinando este hecho con (26) vemos que en el **último término** de (22) basta suponer que $\hat{\Theta}_i = 0$. Es decir, deseamos ver que cuando $|\Pi| \to 0$

$$\sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))(\Delta \mathrm{X}_i)^2 = u^2 \cdot A_n + v^2 \cdot B_n + 2uv \cdot C_n$$

$$\to v^2 \int_{t_0}^{t} g_{xx}(s, \mathrm{X}(s))ds \quad \text{en probabilidad,} \tag{25.29}$$

donde

$$\begin{aligned} A_n &:= \sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))(\Delta t_i)^2 \\ B_n &:= \sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))(\Delta W_i)^2 \\ C_n &:= \sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))\Delta t_i \Delta W_i. \end{aligned}$$

Para probar (29), observe que, por (27),

$$|A_n| \leq |\Pi| \left| \sum_{i=0}^{n-1} |g_{xx}(t_i, \mathrm{X}(t_i))| \Delta t_i \right| \to 0 \quad \text{c.p.1},$$

Similarmente,

$$|C_n| \leq |w_n| \left| \sum_{i=0}^{n-1} |g_{xx}(t_i, \mathrm{X}(t_i))| \Delta t_i \right| \to 0 \quad \text{c.p.1},$$

donde $w_n = \max_i |\Delta W_i|$, lo cual converge a 0 $c.p.1$. Se usó tambén que $\sum\limits_i |g_{xx}(t_i, \mathrm{X}(t_i))|(\Delta t_i) \to \int_{t_0}^t |g_{xx}(s, \mathrm{X}(s))| ds$ $c.p.1$.

Finalmente observe que

$$\begin{aligned} B_n &= \sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))\Delta t_i + \sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))[(\Delta W_i)^2 - \Delta t_i] \\ &\to \int_{t_0}^t g_{xx}(s, \mathrm{X}(s)) ds \quad \text{en probabilidad} \end{aligned}$$

porque como g_{xx} y $\mathrm{X}(\cdot)$ son continuos, se tiene que

$$\sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))\Delta t_i \to \int_{t_0}^t g_{xx}(s, \mathrm{X}(s)) ds \quad \text{c.s.},$$

y se puede verificar que

$$\sum_{i=0}^{n-1} g_{xx}(t_i, \mathrm{X}(t_i))((\Delta W_i)^2 - \Delta t_i) \to 0 \quad \text{en} \quad L_2$$

y por lo tanto, en probabilidad. Esto completa la demostración de (29). En conclusión, de (24), (25) y (29) se obtiene el Teorema 25.4 en el Caso 1, o sea, cuando u y v son "constantes".

Caso 2. $u(\cdot)$ y $v(\cdot)$ procesos **escalonados**. Este caso se obtiene del Caso 1 porque si $u(\cdot)$ y $v(\cdot)$ son escalonados, entonces el intervalo $[t_0, t]$ se puede particionar en un número finito de intervalos en cada uno de los cuales $u(\cdot)$ y $v(\cdot)$ son constantes.

Caso general. Por hipótesis casi seguramente $u(\cdot) \in L_1[0,1]$ y $v(\cdot) \in L_2[0,t]$, y para cada casi toda ω podemos suponer que u y v están acotadas, entonces se sabe que existen sucesiones $\{u_n(\cdot)\}$ y $\{v_n(\cdot)\}$ de PEs escalonados tales que, *con probabilidad 1*:

$$\int_{t_0}^{s} |u_n(r) - u(r)| dr \to 0, \quad \int_{t_0}^{s} |v_n(r) - v(r)|^2 dr \to 0.$$

y

$$\begin{aligned} \mathrm{X}_n(s) \quad &:= \quad \mathrm{X}(t_0) + \int_{t_0}^{s} u_n(r)dr + \int_{t_0}^{s} v_n(r)dW(r) \\ &\quad\to \mathrm{X}(s) \quad \text{uniformemente en}[t_0, t]. \end{aligned}$$

Como $g \in C^{1,2}$, esto implica que la sucesión $Y_n(t) := g(t, \mathrm{X}_n(t)) \to Y(t) = g(t, \mathrm{X}(t))$ uniformemente c.p.1. Por último tomando el límite cuando $n \to \infty$ de la sucesión

$$Y_n(t) - Y_n(t_0) = \int_{t_0}^{t} Lg(s, \mathrm{X}_n(s))ds + \int_{t_0}^{t} g_x(s, \mathrm{X}_n(s))v_n(s)dW(s)$$

se obtiene la "forma integral" de (11). $\square$

Ejercicios § 25

25.1. Use la regla de Ito para calcular la diferencial estocástica de los siguientes PEs:

(a) $Y(t) := 2 + t^2 + e^{W(t)}$;

(b) $Y(t) := (t, W^2(t))$;

(c) $Y(t) := (W_1(t) + W_2(t),\ t + W_1(t),\ W_2^2(t) - W_1(t))$ donde $(W_1(t), W_2(t))$ es un proceso de Wiener bi–dimensional.

25.2. Demuestre que $\int_0^t W^2(s)dW(s) = \frac{1}{3}W^3(t) - \int_0^t W(s)ds \ \ \forall\, t \geq 0$.

25.3. Para $n = 0, 1, \ldots$ y $t \geq 0$, sea $m_n(t) := E[W^n(t)]$. Use la regla de Ito para demostrar que $m_n(t) = \frac{1}{2}n(n-1)\int_0^t m_{n-2}(s)ds$ para $n \geq 2$. Demuestre que $m_n(\cdot) = 0$ si $n \geq 1$ es impar, y calcule $m_4(t)$ y $m_6(t)$.

25.4. (a) Sea $Y(t) := e^{\beta W(t) + ct}$ en donde β y c son constantes. Demuestre que

$$dY(t) = (c + \frac{1}{2}\beta^2)Y(t)dt + \beta Y(t)dW(t).$$

(Compare con el Ejemplo 25.5(c).)

(b) Sea $Y(t) := \exp\left[\sum_{j=1}^{d} \beta_j W_j(t) + ct\right]$ en donde $\beta_1, \ldots \beta_d, c$ son constantes, y $W = (W_1, \ldots, W_d)'$ es un proceso de Wiener d–dimensional. Demuestre que

$$dY(t) = \left(c + \frac{1}{2}\sum_{j=1}^{d} \beta_j^2\right) Y(t)\, dt + Y(t) \cdot \sum_{j=1}^{d} \beta_j dW_j(t).$$

25.5. Supóngase que $\mathrm{X}(\cdot)$ satisface $d\mathrm{X} = a\mathrm{X}\, dt + b\mathrm{X}\, dW$. Demuestre que

$$d\mathrm{X}^2 = (2a + b^2)\mathrm{X}^2\, dt + 2b\mathrm{X}^2\, dW.$$

Deduzca que si el cuarto momento $E[\mathrm{X}^4(t)]$ $(t \geq 0)$ es acotado, entonces

$$E\mathrm{X}^2(t) = E\mathrm{X}^2(0) + (2a + b^2)\int_0^t E\mathrm{X}^2(s)ds$$

y, por lo tanto, el segundo momento $m_2(t) := E\mathrm{X}^2(t)$ satisface que

$$m_2(t) = m_2(0)e^{(2a+b^2)t} \quad \forall\, t \geq 0.$$

Nota. La propiedad $E(\int f\, dW) = 0$ en *requiere la condición (c) de la Definición 24.16*, i.e. $E(\int f^2\, dt) < \infty$.

25.6. Sea $d\mathrm{X}(t) = -\frac{1}{2}\beta^2\,dt + \beta\,dW$ con β constante y $\mathrm{X}(0) = 0$. Sea $Y(t) := e^{\mathrm{X}(t)} = e^{\beta W(t) - \beta^2 t/2}$. Por el Ejemplo 25.5(c), $dY(t) = \beta Y(t)\,dW(t)$, i.e.

$$Y(t) = 1 + \int_0^t \beta Y(s)\,dW(s),$$

lo cual implica que $Y(\cdot)$ es martingala. (Véa la Proposición 24.10 (b) y la nota en el ejercicio anterior). Supóngase ahora que $\beta(\cdot)$ **es una función de** t, que pertenece a $M[0,t]_{na}$ para todo $t \geq 0$ y es tal que

$$E\int_0^t |\beta(s)|^4\,ds < \infty \quad \text{y} \quad E\left(e^{4\mathrm{X}(t)}\right) = E\left[Y^4(t)\right] \leq K$$

en donde K es una constante. Bajo estas condiciones, demuestre que $Y(t) = e^{\mathrm{X}(t)}$ es una martingala con respecto a $\{\mathcal{G}_t,\, t \geq 0\}$.

25.7. Supóngase que $d\mathrm{X}(t) = v(t)\,dW(t)$ con $v(\cdot) \in N[0,t]_{na}$ para todo $t \geq 0$; en particular, $E\left(\int_0^t v^2(s)\,ds\right) < \infty$ para $t \geq 0$. Demuestre que el PE $Y(t) := \mathrm{X}^2(t) - \int_0^t v^2(s)ds$ es una martingala con respecto a $\{\mathcal{G}_t\}$.

25.8. (Caso especial de la **fórmula de Dynkin** y del **generador infinitesimal**.) Suponga que se satisfacen las hipótesis del Teorema 25.7 y, además, $g(t,x)$ y $v(t)$ son tales que

$$E\int_0^t |g_x(s,\mathrm{X}(s))v(s)|^2\,ds < \infty \quad \forall\, t \geq 0.$$

(a) Deduzca de (19) la **fórmula de Dynkin**

$$Eg(t,\mathrm{X}(t)) = Eg(s,\mathrm{X}(s)) + E\int_s^t Lg(r,\mathrm{X}(r))dr \quad \forall\, 0 \leq s < t. \quad (*)$$

(b) Si $g(t,x) \equiv g(x)$ no depende de t, entonces el operador L en (16) se reduce a

$$(Lg)(t,\mathrm{X}(t)) = g_x(\mathrm{X}(t))u(t) + \frac{1}{2}\mathrm{Tr}[\sigma(t)g_{xx}(\mathrm{X}(t))]$$

con $\sigma(\cdot) := v(\cdot)v(\cdot)'$. Además, si $\mathrm{X}(0) = x$, vemos de (*) que

$$Eg(\mathrm{X}(t)) = g(x) + E\int_0^t Lg(r,\mathrm{X}(r))dr \quad \forall\, t \geq 0,$$

en donde $Eg(\mathrm{X}(t)) = E[g(\mathrm{X}(t))|\mathrm{X}(0) = x]$.

(c) Para cada estado inicial $\mathrm{X}(0) = x$, definimos el **generador infinitesimal** de $\mathrm{X}(\cdot)$ como

$$Ag(x) := \lim_{t \downarrow 0} \frac{1}{t} [Eg(\mathrm{X}(t)) - g(x)].$$

Demuestre que $Ag = Lg$.

26 Ecuaciones diferenciales estocásticas

Contenido: Ecuación diferencial estocástica, condiciones de Ito, ecuaciones lineales, propiedades de soluciones de EDEs.

A. Solubilidad de EDEs

En esta sección se dan condiciones para la existencia y unicidad de soluciones de una **ecuación diferencial estocástica** (EDE)

$$dX(t) = f(t, X(t))dt + G(t, X(t))dW(t) \quad \forall\, t_0 \leq t \leq T, \tag{26.1}$$

con condición inicial $X(t_0) = C$, o en forma integral

$$X(t) = C + \int_{t_0}^{t} f(s, X(s))ds + \int_{t_0}^{t} G(s, X(s))dW(s) \quad \forall\, t_0 \leq t \leq T, \tag{26.2}$$

en donde $W(\cdot) \in \mathbb{R}^d$, $X(\cdot) \in \mathbb{R}^m$, $f(t,x) \in \mathbb{R}^m$, $G(t,x) \in \mathbb{R}^{m \times d}$ y $C \in \mathbb{R}^m$. Supondremos que la condición inicial C es una v.a. independiente de $W(t) - W(t_0)$ para $t \geq t_0$. Usaremos la siguiente terminología:

- $f(t,x) \equiv$ vector (o coeficiente) de **tendencia** o **deriva**,
- $G(t,x) \equiv$ matriz (o coeficiente) de **dispersión**,
- $G(t,x)G(t,x)' \equiv$ matriz de **difusión**.

A (1) se le llama una EDE **de Ito** para distinguirla de otras EDEs, como las EDEs en L_2 (vea la Sección 22).

Sea $\mathcal{G}(\cdot) = \{\mathcal{G}_t,\, t \geq 0\}$ la filtración no–anticipante con respecto a $W(\cdot)$ definida como $\mathcal{G}_t := \sigma\{C, W(s) \text{ para } s \leq t\}$. (Vea el Ejemplo 24.14(b).)

Definición 26.1. Decimos que un PE $X(\cdot)$ es **solución** de la EDE (1) si:

(a) $X(\cdot)$ está adaptado a $\mathcal{G}(\cdot)$;

(b) $f(t,x)$ y $G(t,x)$ son funciones medibles y tales que

$$\hat{f}(t,\omega) := f(t, \mathrm{X}(t,\omega)) \quad \text{y} \quad \hat{G}(t,\omega) := G(t, \mathrm{X}(t,\omega))$$

satisfacen que c.p.1.

$$\int_{t_0}^{T} |\hat{f}(t,\omega)|dt < \infty \quad \text{y} \quad \int_{t_0}^{T} |\hat{G}(t,\omega)|^2 dt < \infty,$$

con $|\hat{G}|^2 = \mathrm{Tr}(\hat{G}\hat{G}')$ (en otras palabras, $\hat{G}$ está en $M[a,b]_{na}^{m\times d}$ — vea la Definición 24.17);

(c) $\mathrm{X}(t)$ satisface (2) para todo $t \in [t_0, T]$ c.s.

El siguiente teorema se demuestra al final de esta sección.

Teorema 26.2. *Supóngase que $f(t,x)$ y $G(t,x)$ satisfacen las siguientes hipótesis, llamadas las* **condiciones de Ito**, *para alguna constante $K \geq 0$:*

(a) *(***condición de Lipschitz***) para todo $t \in [t_0, T]$, $x, y \in \mathbb{R}^m$;*

$$|f(t,x) - f(t,y)| + |G(t,x) - G(t,y)| \leq K|x-y|; \tag{26.3}$$

(b) *(***condición de "crecimiento lineal"***) para todo $t \in [t_0, T]$, $x \in \mathbb{R}^m$:*

$$|f(t,x)| + |G(t,x)| \leq K(1 + |x|). \tag{26.4}$$

Entonces la EDE (1) tiene una única solución continua c.s., es decir, si $\mathrm{X}(\cdot)$ y $Y(\cdot)$ son dos PEs continuos que satisfacen (1), entonces

$$\mathrm{P}(\sup_{t_0 \leq t \leq T} |\mathrm{X}(t) - Y(t)| > 0) = 0. \tag{26.5}$$

Antes de demostrar el teorema veremos algunas observaciones y casos especiales.

Observación 26.3. (a) Si no se cumple la condición de Lipschitz (3), la EDE puede no tener una única solución. Por ejemplo, tomando $G(t,x) = 0$, sea

$$\dot{x}(t) := 3x(t)^{2/3} \quad \text{para} \quad t \geq 0, \quad \text{con} \quad x(0) = 0. \tag{26.6}$$

Entonces para **cualquier** $a > 0$, la función

$$x(t) = \begin{cases} 0 & \text{si} \quad t \leq a, \\ (t-a)^3 & \text{si} \quad t > a \end{cases}$$

es solución de (6).

(b) Si no se cumple la condición de crecimiento lineal (4), la solución de (1) puede "explotar" en un tiempo finito. Por ejemplo, si

$$\dot{x}(t) = x^2(t), \quad \text{con} \quad x(0) = 1,$$

entonces la ecuación tiene la (única) solución $x(t) = \frac{1}{1-t}$ para $0 \leq t < 1$, y $x(t) \to +\infty$ cuando $t \uparrow 1$.

(c) **Condición de Lipschitz local:** El Teorema 26.2 sigue valiendo si sustituimos (3) por una condición "local": para cada $N > 0$ existe una constante K_N tal que

$$|f(t,x) - f(t,y)| + |G(t,x) - G(t,y)| \leq K_N |x-y| \tag{26.7}$$

para todo $t \in [t_0, T]$ y $|x|$, $|y| \leq N$. (Por ejemplo, $f(x) = x^2$ satisface (7) pero no (3).)

(d) **EDE autónoma.** Si $f(t,x) \equiv f(x)$ y $G(t,x) \equiv G(x)$ no dependen de t, se dice que

$$d\mathrm{X}(t) = f(\mathrm{X}(t))dt + G(\mathrm{X}(t))dW(t) \tag{26.8}$$

es una EDE **autónoma**. En este caso, las condiciones de Ito (a) y (b) del Teorema 26.2 se pueden sustituir por la siguiente condición de Lipschitz: existe K tal que

$$|f(x) - f(y)| + |G(x) - G(y)| \leq K|x-y| \quad \forall\, x, y \in \mathbb{R}^m. \tag{26.9}$$

En efecto, (9) implica la condición de crecimiento lineal (4) en el caso autónomo. Esto se obtiene observando que si $\bar{x} \in \mathbb{R}^m$ es un punto fijo arbitrario, tenemos

$$|f(x)| \leq |f(x) - f(\bar{x})| + |f(\bar{x})| \leq K|x - \bar{x}| + |f(\bar{x})| \leq K'(1 + |x|)$$

donde $K' := \max\{K,\, K|\bar{x}| + |f(\bar{x})|\}$, y similarmente para $G(x)$. $\square$

Definición 26.4. Se dice que la EDE (1) es **lineal** si es de la forma

$$dX(t) = (A(t)X(t) + a(t))dt + \sum_{i=1}^{d}[B_i(t)X(t) + b_i(t)]dW_i(t), \tag{26.10}$$

con $A, B_1, \ldots, B_d$ matrices $m \times m$ y $a, b_1, \ldots, b_d \in \mathbb{R}^m$. Además, la EDE lineal (10) es:

- **homogénea** si $a(t)$, $b_1(t), \ldots, b_d(t)$ son $\equiv 0$, i.e.

$$dX(t) = A(t)X(t)dt + \sum_{i=1}^{d} B_i(t)X(t)dW_i(t);$$

- **autónoma** si es homogénea y $A(t) \equiv A$ y $B_i(t) \equiv B_i$ $(i = 1, \ldots, d)$ son matrices constantes, i.e.

$$dX(t) = AX(t)dt + \sum_{i=1}^{d} B_i X(t)dW_i(t);$$

- **lineal en el sentido restringido** si $B_i(t) \equiv 0$ para todo $i = 1, \ldots, d$, i.e.

$$dX(t) = (A(t)X(t) + a(t))dt + B(t)dW(t), \ X(t_0) = C, \tag{26.11}$$

  para alguna matriz $B(t) = (b_{ij}(t))\, m \times d$. **Nota:** Si $B(\cdot) \in C^1$ se pueden usar las técnicas de la Sección 22.

Solución de la ecuación (11). Considérese la ecuación determinista

$$\dot{x}(t) = A(t)x(t) + a(t) \quad \text{para} \quad t_0 \leq t \leq T, \quad \text{con} \quad x(t_0) = c, \tag{26.12}$$

y la correspondiente ecuación homogénea $\dot{x}(t) = A(t)x(t)$. La solución $\Phi(t) \equiv \Phi(t, t_0)$ de la ecuación matricial

$$\dot{\Phi}(t) = A(t)\Phi(t) \quad \text{con} \quad \Phi(t_0) = I \tag{26.13}$$

se llama la **matriz fundamental** de (12) y, además, la solución de (12) es

$$x(t) = \Phi(t)\left[c + \int_{t_0}^{t} \Phi^{-1}(s)a(s)ds\right].$$

En particular si $A(t) \equiv A$ es independiente de t, entonces

$$\Phi(t) = e^{A(t-t_0)} := \sum_{n=0}^{\infty} A^n (t-t_0)^n / n!$$

y la solución de (12) resulta

$$x(t) = e^{A(t-t_0)} c + \int_{t_0}^{t} e^{A(t-s)} a(s) ds.$$

Proposición 26.5. Sea $Y(\cdot)$ el PE definido como

$$Y(t) := C + \int_{t_0}^{t} \Phi^{-1}(s) a(s) ds + \int_{t_0}^{t} \Phi^{-1}(s) B(s) dW(s), \ t \geq t_0.$$

Entonces la solución de la ecuación lineal "restringida" (11) es

$$\mathrm{X}(t) = \Phi(t) Y(t) = \Phi(t) \left[C + \int_{t_0}^{t} \Phi^{-1}(s) \left(a(s) ds + B(s) dW(s) \right) \right].$$

En particular, si $A(t) \equiv A$ es una matriz constante, entonces

$$\mathrm{X}(t) = e^{A(t-t_0)} C + \int_{t_0}^{t} e^{A(t-s)} \left[a(s) ds + B(s) dW(s) \right]. \tag{26.14}$$

Demostración. El proceso $Y(\cdot)$ tiene la diferencial estocástica

$$dY(t) = \Phi(t)^{-1} [a(t) dt + B(t) dW(t)]. \tag{26.15}$$

Por lo tanto, aplicando la regla de Ito al PE $\mathrm{X}(t) = \Phi(t) Y(t)$ (es decir, $g(t, y) := \Phi(t) y$) vemos que, por (13) y (15),

$$\begin{aligned} d\mathrm{X}(t) &= \dot{\Phi}(t) Y(t) dt + \Phi(t) dY(t) \\ &= A(t) \Phi(t) Y(t) dt + a(t) dt + B(t) dW(t) \\ &= (A(t) \mathrm{X}(t) + a(t)) dt + B(t) dW(t). \quad \square \end{aligned}$$

$\square$

La solución de la EDE lineal general (10) se puede ver en el libro de Arnold (1974), Sección 8.5.

Ejemplo 26.6. La ecuación de Langevin. Del Ejemplo 22.9 la ecuación de Langevin es

$$m\,V'(t) + f\,V(t) = W'(t), \quad t \geq 0, \quad \text{con} \quad V(0) = v_0 \in \mathbb{R},$$

que escrita como EDE de Ito resulta

$$dV(t) = \alpha V(t) + \frac{1}{m} dW(t), \quad \text{con} \quad \alpha := -f/m.$$

Esta es una EDE "escalar" ($m = d = 1$), lineal en el sentido restringido (11) con $A(t) \equiv \alpha$, $a(t) \equiv 0$ y $B(t) \equiv 1/m$. Como $t_0 = 0$, de (14) obtenemos

$$V(t) = v_0 e^{\alpha t} + \frac{1}{m} \int_0^t e^{\alpha(t-s)} dW(s) \quad \forall\, t \geq 0, \tag{26.16}$$

y $V(t) = \mathrm{X}'(t)$ es la velocidad de una partícula con posición $\mathrm{X}(t)$. Para encontrar la posición $\mathrm{X}(\cdot)$, que se conoce como el **proceso de Ornstein–Uhlenbeck**, podemos integrar $V(t)$ como en el Ejemplo 22.9 para obtener que

$$\mathrm{X}(t) = x_0 + \frac{v_0}{\alpha}(e^{\alpha t} - 1) + \frac{1}{m\alpha} \int_0^t (e^{\alpha(t-s)} - 1) dW(s) \quad \forall\, t \geq 0, \tag{26.17}$$

con $\mathrm{X}(0) = x_0$. Alternativamente, podemos considerar la **ecuación de Ornstein–Uhlenbeck** $m\mathrm{X}'' + f\mathrm{X}' = W'$, i.e.

$$\mathrm{X}'' = \alpha \mathrm{X}' + \frac{1}{m} W', \quad \text{con} \quad \mathrm{X}(0) = x_0,\ \mathrm{X}'(0) = V(0) = v_0, \tag{26.18}$$

con $\alpha := -f/m$. En efecto, sea $\mathrm{X}_1(t) := \mathrm{X}(t)$ y $\mathrm{X}_2(t) := \mathrm{X}'(t)(= V(t))$. Entonces podemos reescribir (18) como una EDE de Ito en $\mathbb{R}^2$:

$$\begin{cases} d\mathrm{X}_1(t) = \mathrm{X}_2(t)dt \\ d\mathrm{X}_2(t) = \alpha \mathrm{X}_2(t)dt + \frac{1}{m} dW, \end{cases}$$

i.e.

$$d\begin{pmatrix} \mathrm{X}_1(t) \\ \mathrm{X}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} \mathrm{X}_1(t) \\ \mathrm{X}_2(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \frac{1}{m} \end{pmatrix} dW(t), \tag{26.19}$$

con

$$\begin{pmatrix} \mathrm{X}_1(0) \\ \mathrm{X}_2(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ v_0 \end{pmatrix}.$$

Esta es una EDE lineal en el sentido restringido (11) con coeficientes

$$A = \begin{pmatrix} 0 & 1 \\ 0 & \alpha \end{pmatrix}, \quad a(\cdot) \equiv 0 \in \mathbb{R}^2, \quad B(t) = \begin{pmatrix} 0 \\ 1/m \end{pmatrix}.$$

Como $A^0 = I$ y $A^n = \begin{pmatrix} 0 & \alpha^{n-1} \\ 0 & \alpha^n \end{pmatrix}$ para $n \geq 1$, se puede ver (Ejercicio 1) que la solución de (19) está dada por (14) con componente $\mathrm{X}_1 = \mathrm{X}$ y $\mathrm{X}_2 = \mathrm{X}'(= V)$ como en (17) y (16), respectivamente. $\square$

Demostración. Del Teorema 26.2 en el caso escalar $(m = d = 1)$ **y suponiendo:**

$$E(C^2) < \infty. \tag{26.20}$$

Bajo esta condición se puede demostrar (Ejercicio 6) que

$$\text{(a) } E[\mathrm{X}^2(t)] < \infty \;\; \forall\, t_0 \leq t \leq T, \quad \text{y} \quad \text{(b) } \int_{t_0}^{T} E[G(s, \mathrm{X}(s))^2] ds < \infty. \tag{26.21}$$

Unicidad. Sean $\mathrm{X}(\cdot)$ y $Y(\cdot)$ dos soluciones de (2), de modo que para $t_0 \leq t \leq T$:

$$\begin{aligned} \mathrm{X}(t) - Y(t) &= \int_{t_0}^{t} [f(s, \mathrm{X}(s)) - f(s, Y(s))] ds \\ &+ \int_{t_0}^{t} [G(s, \mathrm{X}(s)) - G(s, Y(t))] dW(s). \end{aligned} \tag{26.22}$$

Primero demostraremos que

$$E|\mathrm{X}(t) - Y(t)|^2 = 0 \quad \forall\, t_0 \leq t \leq T. \tag{26.23}$$

Con este fin, en (22) tome cuadrados en ambos lados de la desigualdad y use que $(a + b)^2 \leq 2(a^2 + b^2)$; después use la desigualdad de Schwartz

$$\left(\int_{t_0}^{t} fg \right)^2 \leq \left(\int_{t_0}^{t} f^2 \right) \left(\int_{t_0}^{t} g^2 \right) \quad \text{con} \quad g(\cdot) \equiv 1,$$

tome esperanzas y use la isometría de Ito para obtener

$$\begin{aligned} E(\mathrm{X}(t) - Y(t))^2 &\leq 2(t - t_0) \int_{t_0}^{t} E[f(s, \mathrm{X}(s)) - f(s, Y(s))]^2 ds \\ &+ 2 \int_{t_0}^{t} E[G(s, \mathrm{X}(s)) - G(s, Y(s))]^2 ds. \end{aligned} \tag{26.24}$$

Por la condición de Lipschitz (3),

$$E[f(s, \mathrm{X}(s)) - f(s, Y(s))]^2 \leq K^2 E[\mathrm{X}(s) - Y(s)]^2 \tag{26.25}$$

y similarmente para G en lugar de f. Por lo tanto, definiendo $g(t) := E(\mathrm{X}(t) - Y(t))^2$, de (24) vemos que

$$g(t) \leq L \int_{t_0}^{t} g(s)ds \quad \text{con} \quad L := (2(t - t_0) + 2)K^2. \tag{26.26}$$

Luego, por la desigualdad de Bellman–Gronwall (Ejercicio 5), de (26) se obtiene (23).

Finalmente, para obtener (5), nótese que (23) implica

$$\mathrm{P}(|\mathrm{X}(t) - Y(t)| = 0) = 1 \quad \forall\, t \in [t_0, T].$$

Luego, si Q es el conjunto de los números racionales, se sigue que

$$\mathrm{P}(|\mathrm{X}(t) - Y(t)| = 0 \quad \forall\, t \in Q \cap [t_0, T]) = 1.$$

Por lo tanto, por la continuidad de $t \mapsto |\mathrm{X}(t) - Y(t)|$, concluímos que

$$\mathrm{P}(|\mathrm{X}(t) - Y(t)| = 0 \quad \forall\, t \in [t_0, T]) = 1.$$

Existencia. Usaremos el método de Picard o de **aproximaciones sucesivas**: definimos una sucesión de PEs $\mathrm{X}^{(n)}(\cdot) := \{\mathrm{X}^{(n)}(t),\, t_0 \leq t \leq T\}$ para $n = 0, 1, \ldots$, con $\mathrm{X}^{(0)}(\cdot) \equiv C$ y para $n \geq 1$:

$$\mathrm{X}^{(n)}(t) := C + \int_{t_0}^{t} f(s, \mathrm{X}^{(n-1)}(s))ds + \int_{t_0}^{t} G(s, \mathrm{X}^{(n-1)}(s))dW(s) \ \ \forall\, t \in [t_0, T]. \tag{26.27}$$

Por inducción, para cada $n = 0, 1, \ldots$, el proceso $\mathrm{X}^{(n)}(\cdot)$ está adaptado a $\mathcal{G}$ y es continuo.

Deseamos demostrar que existe un PE $\mathrm{X}(\cdot)$ tal que

$$\lim_{n \to \infty} \mathrm{X}^{(n)}(t) = \mathrm{X}(t) \quad \text{c.s. uniformemente en} \quad t \in [t_0, T] \tag{26.28}$$

y que

$$\mathrm{X}(\cdot) \quad \text{satisface la EDE} \quad (1) - (2). \tag{26.29}$$

Para probar (28) verificaremos que

$$\sup_{t_0 \le t \le T} E|\mathrm{X}^{(n+1)}(t) - \mathrm{X}^{(n)}(t)|^2 \le L_1 \cdot (L_0(T-t_0))^n/n! \quad \forall\, n \ge 0, \qquad (26.30)$$

en donde $L_0 = L_0(K, T-t_0)$ y $L_1 = L_1(L_0, T-t_0, EC^2)$ son constantes, y que

$$\sum_{n=1}^{\infty} \mathrm{P}(D_n \ge 1/n^2) < \infty, \quad \text{donde} \quad D_n := \sup_{t_0 \le t \le T} |\mathrm{X}^{(n+1)}(t) - \mathrm{X}^{(n)}(t)|. \qquad (26.31)$$

Demostración de (30): caso $n = 0$. Como $\mathrm{X}^{(0)}(\cdot) \equiv C$,

$$\mathrm{X}^{(1)}(t) - \mathrm{X}^{(0)}(t) = \int_{t_0}^{t} f(s, C) ds + \int_{t_0}^{t} G(s, C) dW(s).$$

Usando el mismo argumento que seguimos para obtener (24) se ve que

$$E(\mathrm{X}^{(1)}(t) - \mathrm{X}^{(0)}(t))^2 \le 2(t-t_0) \int_{t_0}^{t} E(f^2) ds + 2 \int_{t_0}^{t} E(G^2) ds.$$

Por otra parte, de (4), $E[f^2(s, C)] \le 2K^2(1 + EC^2)$ y análogamente para $E[G^2(s, C)]$. Por lo tanto,

$$\begin{aligned} E(\mathrm{X}^{(1)}(t) - \mathrm{X}^{(0)}(t))^2 &\le 2(t-t_0)[2K^2(1+EC^2)](t-t_0+1) \\ &\le 2(T-t_0)[2K^2(1+EC^2)](T-t_0+1) =: L_1. \end{aligned}$$

Esto implica (30) para $n = 0$. Para $\boldsymbol{n \ge 1}$, un argumento similar da

$$E(\mathrm{X}^{(n+1)}(t) - \mathrm{X}^{(n)}(t))^2 \le L_0 \int_{t_0}^{t} E(\mathrm{X}^{(n)}(s) - \mathrm{X}^{(n-1)}(s))^2 ds,$$

con $L_0 := 2(T - t_0 + 1)K$. Finalmente, iterando la desigualdad anterior se obtiene

$$E(\mathrm{X}^{(n+1)}(t) - \mathrm{X}^{(n)}(t))^2 \le L_0^n \frac{1}{(n-1)!} \int_{t_0}^{t} (t-s)^{n-1} E(\mathrm{X}^{(1)}(s) - \mathrm{X}^{(0)}(s))^2 ds$$

Luego, como $E((\mathrm{X}^{(1)}(\cdot) - \mathrm{X}^{(0)}(\cdot))^2) \le L_1$, se sigue (30).

Demostración de (31). Por la desigualdad de Chebyshev,

$$\sum_{n=1}^{\infty} \mathrm{P}(D_n \ge 1/n^2) \le \sum_{n=1}^{\infty} n^4 E(D_n^2).$$

Por lo tanto, para demostrar (31) veremos que, para alguna constante M,

$$E(D_n^2) \leq M \cdot (L_0(T-t_0))^{n-1}/(n-1)! \quad \forall\, n \geq 1 \tag{26.32}$$

porque entonces (usando el criterio del cociente para convergencia de series)

$$\sum_{n=1}^{\infty} \mathrm{P}(D_n \geq 1/n^2) \leq M \sum_{n=1}^{\infty} n^4 (L_0(T-t_0))^{n-1}/(n-1)! < \infty.$$

Para demostrar (32) primero nótese que [por la definición (27) de $\mathrm{X}^{(n)}(t)$]

$$D_n := \sup_{t_0 \leq t \leq T} |\mathrm{X}^{(n+1)}(t) - \mathrm{X}^{(n)}(t)| \leq A_n + B_n$$

donde

$$A_n := \int_{t_0}^{T} |f(s, \mathrm{X}^{(n)}(s)) - f(s, \mathrm{X}^{(n-1)}(s))| ds$$

y

$$B_n := \sup_{t_0 \leq t \leq T} |\int_{t_0}^{t} [G(s, \mathrm{X}^{(n)}(s)) - G(s, \mathrm{X}^{(n-1)}(s)] dW(s)|.$$

Luego [como $(a+b)^2 \leq 2(a^2+b^2)$]

$$E(D_n^2) \leq 2\, E(A_n^2) + 2\, E(B_n^2),$$

de modo que para demostrar (32) basta ver que existen constantes M_1, M_2 tales que

$$E(A_n^2) \leq M_1 \bullet (L_0(T-t_0))^{n-1}/(n-1)! \tag{26.33}$$

$$E(B_n^2) \leq M_2 \bullet (L_0(T-t_0))^{n-1}/(n-1)! \tag{26.34}$$

Demostración de (33). Por la condición de Lipschitz (3):

$$A_n \leq K \int_{t_0}^{T} |\mathrm{X}^{(n)}(s) - \mathrm{X}^{(n-1)}(s)| dt,$$

así que (por la desigualdad de Cauchy–Schwarz)

$$A_n^2 \leq K^2 (T-t_0) \int_{t_0}^{T} |\mathrm{X}^{(n)}(s) - \mathrm{X}^{(n-1)}(s)|^2 ds.$$

Esto implica que

$$E(A_n^2) \leq K^2(T-t_0)\int_{t_0}^{T} E|\mathrm{X}^{(n)}(s)-\mathrm{X}^{(n-1)}(s)|^2 ds$$

$$\leq K^2(T-t_0)^2 L_1 \cdot (L_0(T-t_0))^{n-1}/(n-1)! \quad \text{(por (30))}$$

y así se obtiene (33) con $M_1 := K^2(T-t_0)^2 L_1$.

Demostración de (34). Por la desigualdad de Doob para martingalas (vea la Nota 1(b) en el final de la Sección 24, con $p=2$)

$$\begin{aligned}
E(B_n^2) &\leq 4\,E\left[\int_{t_0}^{T}(G(s,\mathrm{X}^{(n)}(s))-G(s,\mathrm{X}^{(n-1)}(s))dW(s)\right]^2 \\
&= 4\int_{t_0}^{T} E[G(s,\mathrm{X}^{(n)}(s))-G(s,\mathrm{X}^{(n-1)}(s))]^2 ds \quad \text{(Isometría de Ito)} \\
&\leq 4\,K^2\int_{t_0}^{T} E[\mathrm{X}^{(n)}(s)-\mathrm{X}^{(n-1)}(s)]^2 ds \quad \text{(condición de Lipschitz)} \\
&\leq M_2 \cdot (L_0(T-t_0))^{n-1}/(n-1)! \quad \text{(por (30))},
\end{aligned}$$

que da (34) con $M_2 := 4\,K^2(T-t_0)L_1$.

Como ya se tienen (33) y (34) entonces, como ya se mencionó, se sigue (31), i.e.

$$\sum_{n=1}^{\infty} \mathrm{P}(D_n \geq 1/n^2) < \infty.$$

Esto a su vez implica que (por el Lema de Borel–Cantelli; vea el Ejercicio 24.5)

$$\mathrm{P}(\limsup_{n\to\infty}\{D_n \geq 1/n^2\}) = 0$$

o, equivalentemente,

$$\mathrm{P}(\liminf_{n\to\infty}\{D_n < 1/n^2\}) = 1.$$

Es decir, para casi todo $\omega \in \Omega$, existe $N(\omega)$ tal que

$$D_n := \sup_{t_0\leq t\leq T}|\mathrm{X}^{(n+1)}(t)-\mathrm{X}^{(n)}(t)| < 1/n^2 \quad \forall\, n \geq N(\omega).$$

De aquí se sigue que la sucesión $\{\mathrm{X}^{(n)}(t)\}$ es de Cauchy c.p. 1 porque

$$\mathrm{X}^{(n)}(t) = \mathrm{X}^{(0)}(t) + \sum_{i=0}^{n-1} [\mathrm{X}^{(i+1)}(t) - \mathrm{X}^{(i)}(t)].$$

Por lo tanto, existe un PE $\mathrm{X}(\cdot)$ para el cual

$$\mathrm{X}^{(n)}(t) \to \mathrm{X}(t) \quad \text{c.p.1 } \textbf{uniformemente sobre } [t_0, T]. \tag{26.35}$$

Esta convergencia uniforme implica que $\mathrm{X}(\cdot)$ está adaptado a $\mathcal{G}(\cdot)$ y que $\mathrm{X}(t)$ es continuo en todo $t \in [t_0, T]$. Queda entonces mostrado (28).

- **Verificación de que** $\mathrm{X}(\cdot)$ **satisface la EDE (2)**, i.e.

$$\mathrm{X}(t) = C + \int_{t_0}^{t} f(s, \mathrm{X}(s)) ds + \int_{t_0}^{t} G(s, \mathrm{X}(s)) dW(s) \quad \forall t_0 \leq t \leq T. \tag{26.36}$$

Esto es obvio para $t = t_0$ porque como $\mathrm{X}^{(n)}(t_0) = C \;\; \forall\, n \geq 0$, se sigue de (35) que

$$\mathrm{X}(t_0) = \lim_{n \to \infty} \mathrm{X}^{(n)}(t_0) = C \quad \text{c.p.1.}$$

Por otra parte, también se tiene que

$$\int_{t_0}^{T} f(s, \mathrm{X}^{(n)}(s)) ds \to \int_{t_0}^{t} f(s, \mathrm{X}(s)) ds \quad \text{c.p.1.} \quad \forall\, t \in [t_0, T] \tag{26.37}$$

porque (35) y la condición de Lipschitz (3) dan que

$$|\int_{t_0}^{t} f(s, \mathrm{X}^{(n)}(s)) ds - \int_{t_0}^{t} f(s, \mathrm{X}(s)) ds| \leq K \int_{t_0}^{t} |\mathrm{X}^{(n)}(s) - \mathrm{X}(s)| ds$$

$$\to 0 \quad \text{c.p.1.} \quad (\text{por } (35)).$$

Finalmente, por (35) y (30), cuando $n \to \infty$ se tiene que

$$\mathrm{X}^{(n)}(t) \to \mathrm{X}(t) \quad \text{uniformemente sobre} \quad [t_0, T], \quad \text{en} \quad L_2, \tag{26.38}$$

y por lo tanto

$$\int_{t_0}^{t} G(s, \mathrm{X}^{(n)}(s)) dW(s) \to \int_{t_0}^{t} G(s, \mathrm{X}(s)) dW(s) \quad \text{en} \quad L_2 \quad \forall\, t_0 \leq t \leq T.$$

En efecto

$$
\begin{aligned}
& E\left[\int_{t_0}^{t} G(s,\mathrm{X}^{(n)}(s))dW(s) - \int_{t_0}^{t} G(s,\mathrm{X}(s))dW(s)\right]^2 \\
& = E\left[\int_{t_0}^{t} (G(s,\mathrm{X}^{(n)}(s)) - G(s,\mathrm{X}(s)))dW(s)\right]^2 \\
& = \int_{t_0}^{t} E[G(s,\mathrm{X}^{(n)}(s)) - G(s,\mathrm{X}(s))]^2 ds \quad \text{(isometría de Ito)} \\
& \leq K^2 \int_{t_0}^{t} E|\mathrm{X}^{(n)}(s) - \mathrm{X}(s)|^2 ds \quad \text{(condición de Lipschitz)} \\
& \to 0 \quad \text{(por (38))}.
\end{aligned}
$$

Un argumento similar demuestra que la convergencia en (37) también se cumple en L_2. En conclusión, $\mathrm{X}(\cdot)$ satisface (36). $\square$

B. Propiedades de las soluciones de EDEs

simon

En esta sección consideramos la EDE

$$
d\mathrm{X}(t) = f(t,\mathrm{X}(t))\, dt + G(t,\mathrm{X}(t))\, dW(t) \quad \forall\, t_0 \leq t \leq T, \tag{26.39}
$$

con condición inicial $\mathrm{X}(t_0) = C$, bajo la hipótesis de que $f(t,x)$ y $G(t,x)$ satisfacen las *condiciones de Ito* en el Teorema 26.2. Dichas condiciones aseguran la existencia y unicidad de soluciones de (26.39). Ahora veremos algunas propiedades de tales soluciones.

Teorema 26.7. (Estimación de los momentos de las soluciones.) *Suponga que se satisfacen las condiciones de Ito y, además,*

$$
E|C|^{2n} < \infty,
$$

en donde n es un entero positivo. Entonces

(a) $E|\mathrm{X}(t)|^{2n} \leq (1 + E|C|^{2n})e^{D_1(t-t_0)}$ *para todo* $t \in [t_0, T]$, *en donde* $D_1 = 2n(2n+1)K^2$, *y*

(b) $E|\mathrm{X}(t) - C|^{2n} \leq D_2(1 + E|C|^{2n})(t-t_0)^n e^{D_1(t-t_0)}$ *para todo* $t \in [t_0, T]$, *en donde* D_2 *es una constante que sólo depende de* n *,*K *y* $T - t_0$.

La demostración se puede encontrar en Arnold (1974), Teorema 7.1.2.

Para $s \in [t_0, T]$ y $s \le t \le T$, denotaremos por $\mathrm{X}(t; s, x)$ la solución de (26.39) con condición inicial $\mathrm{X}(s) = x$; es decir, $\mathrm{X}(t; s, x)$ satisface

$$\mathrm{X}(t) = x + \int_s^t f(r, \mathrm{X}(r))\, dr + \int_s^t G(r, \mathrm{X}(r))\, dW(r), \quad s \le t \le T. \quad (26.40)$$

Teorema 26.8. (La propiedad de Markov.) *Supóngase que la EDE (26.39), con condición inicial* $\mathrm{X}(t_0) = C$*, satisface las condiciones de Ito. Entonces la solución* $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t_0 \le t \le T\}$ *de (26.39) es un proceso de Markov cuya distribución inicial (es decir, en el tiempo* t_0*) es la distribución de* C *y cuyas probabilidades de transición están dadas por*

$$\mathrm{P}(s, x, t, B) := \mathrm{P}[\mathrm{X}(t) \in B | \mathrm{X}(s) = x] = \mathrm{P}[\mathrm{X}(t; s, x) \in B]$$

para todo $t_0 \le s \le t \le T$*,* $x \in \mathbb{R}^m$ *y* $B \in \mathcal{B}(\mathbb{R}^m)$*, donde* $\mathrm{X}(t; s, x)$ *es la solución de (26.40).*

La demostración se puede encontrar en Arnold (1974), Teorema 9.2.3. (De hecho, se puede demostrar algo más: $\mathrm{X}(\cdot)$ satisface la propiedad de Markov **fuerte**. Vea el libro de Friedman (1975), pág 112, Teorema 3.4.)

Ejemplo 26.9. Considérese la ecuación lineal en el sentido restringido (25.11), i.e.

$$d\mathrm{X}(t) = [A(t)\mathrm{X}(t) + a(t)]\, dt + B(t)\, dW(t) \quad \text{para} \quad t_0 \le t \le T,$$

con condición inicial $\mathrm{X}(t_0) = C$. Por el Ejercicio 26.3, si C es un vector gaussiano (independiente de $W(t) - W(t_0)$ para todo $t \ge t_0$) o un vector constante, entonces la solución $\mathrm{X}(\cdot)$ es un proceso gaussiano. Por lo tanto, en vista del Teorema 26.8 se dice que $\mathrm{X}(\cdot)$ es un proceso de **Gauss–Markov**. □

Notación: $\mathrm{P}_{s,x}(B) := \mathrm{P}(B|\mathrm{X}(s) = x)$, $E_{s,x}(Y) := E(Y|\mathrm{X}(s) = x)$.

Procesos de difusión

Definición 26.10. Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t), t \ge t_0\}$ un proceso de Markov en $\mathbb{R}^m$, con trayectorias continuas c.s., y probabilidades de transición $\mathrm{P}(s, x, t, B)$. *Supóngase que* $\mathrm{X}(\cdot)$ *está en* L_2. Se dice que $\mathrm{X}(\cdot)$ es un **proceso de difusión** si existen funciones $f(s, x) \in \mathbb{R}^m$ y $\sigma(s, x) \in \mathbb{R}^{m \times m}$ tales que para todo $s \ge t_0$ y $x \in \mathbb{R}^m$, cuando $t \downarrow s$ se tiene que

(a) $\mathrm{P}_{s,x}(|\mathrm{X}(t)-\mathrm{X}(s)|>\epsilon)=o(t-s)$,

(b) $E_{s,x}[\mathrm{X}(t)-\mathrm{X}(s)]=f(s,x)(t-s)+o(t-s)$,

(c) $E_{s,x}[(\mathrm{X}(t)-\mathrm{X}(s))(\mathrm{X}(t)-\mathrm{X}(s))']=\sigma(s,x)(t-s)+o(t-s)$.

A f y σ se les llama los **coeficientes** del proceso de difusión $\mathrm{X}(\cdot)$:

- $f(s,x)$ es el coeficiente (o vector) de **tendencia** o de **deriva**,
- $\sigma(s,x)$ es el coeficiente (o matriz) de **difusión**.

Si $\mathrm{X}(\cdot)\notin L_2$, entonces (b) y (c) se deben reemplazar por las siguientes "condiciones locales": para cada $\epsilon>0$

(b′) $\int\limits_{|y-x|\leq\epsilon}(y-x)\mathrm{P}(s,x,t,dy)=f(s,x)(t-s)+o(t-s)$,

(c′) $\int\limits_{|y-x|\leq\epsilon}(y-x)(y-x)'\,\mathrm{P}(s,x,t,dy)=\sigma(s,x)(t-s)+o(t-s)$.

Nótese que $\sigma(s,x)$ es una matriz simétrica y definida no–negativa.

Ejemplo 26.11. (a) Un proceso de Wiener estándar $W(\cdot)\in\mathbb{R}^m$ es un proceso de difusión con vector de deriva $f(s,x)\equiv 0$ y matriz de difusión $\sigma(s,x)\equiv I$.

(b) Considere la ecuación diferencial (determinista) en $\mathbb{R}^m$

$$\dot{x}(t)=f(t,x(t))\quad\forall\,t\geq t_0,\quad\text{con}\quad x(s)=x.$$

Si f satisface las condiciones de Ito y es **continua en** t, entonces $x(\cdot)$ es un proceso de difusión con vector de deriva $f(s,x)$ y matriz de transición $\sigma(s,x)\equiv 0$.

(c) Sean $u_0,c\in\mathbb{R}^m$ y $v_0\in\mathbb{R}^{m\times d}$ constantes. Considérese la diferencial estocástica

$$d\mathrm{X}(t)=u_0\,dt+v_0\,dW\quad\forall\,t\geq t_0,\quad\text{con}\quad\mathrm{X}(s)=c$$

y $W(\cdot)\in\mathbb{R}^d$. Luego

$$\mathrm{X}(t)-\mathrm{X}(s)=u_0\cdot(t-s)+v_0\cdot(W(t)-W(s))\quad\forall\,t\geq s\geq t_0,$$

de modo que $\mathrm{X}(\bullet)$ es un proceso de Markov cuya probabilidad de transición $\mathrm{P}(s, x, t, \bullet)$ es la distribución normal

$$N(x + u_0 \cdot (t - s),\ v_0 v_0' \cdot (t - s)).$$

Además

$$E_{s,x}[\mathrm{X}(t) - \mathrm{X}(s)] = u_0 \cdot (t - s)$$

y

$$E_{s,x}[(\mathrm{X}(t) - x)(\mathrm{X}(t) - x)'] = \sigma \cdot (t - s) + u_0 u_0' \cdot (t - s)^2, \quad \text{con} \quad \sigma := v_0 v_0'.$$

Por lo tanto, $\mathrm{X}(\bullet)$ es un proceso de difusión con vector de deriva $f(s, x) \equiv u_0$ y matriz de difusión $\sigma(s, x) \equiv \sigma := v_0 v_0'$.

Teorema 26.12. *Considérese la EDE (26.39) con coeficientes $f(t, x)$ y $G(t, x)$ que satisfacen las condiciones de Ito. Si además $f(t, x)$ y $G(t, x)$ son* **funciones continuas en** t*, entonces la solución* $\mathrm{X}(\bullet)$ *de (26.39) es un proceso de difusión con vector de deriva $f(t, x)$ y matriz de difusión $\sigma(t, x) := G(t, x)G(t, x)'$. (En este caso se dice que* $\mathrm{X}(\bullet)$ *es una* **difusión de Ito***). En particular, si la EDE es autónoma (i.e. $f(t, x) \equiv f(x)$ y $(t, x) \equiv G(x)$), entonces* $\mathrm{X}(\cdot)$ *es un* **proceso de difusión homogéneo***.*

Uns desmostración se encuentra en Arnold (1974), Teorema 9.3.1; o Friedman (1975), p. 115, Teorema 4.2.

Tiempos de paro

Definición 26.13. Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y $\mathcal{F}_\bullet = \{\mathcal{F}_t,\ t \geq 0\}$ una filtración de $\mathcal{F}$. Se dice que una v.a. $\tau : \Omega \to [0, \infty]$ es un **tiempo de paro** con respecto a $\mathcal{F}_\bullet$ si

$$\{\tau \leq t\} \in \mathcal{F}_t \quad \forall\, t \geq 0.$$

Nótese que un tiempo de paro τ puede tomar el valor $+\infty$, y que cualquier constante $\tau \equiv t$ es un tiempo de paro.

Ejemplo 26.14. Sea $\mathrm{X}(t) \in \mathbb{R}^d$, $t \geq 0$, la solución de la EDE (26.39) con condición inicial $\mathrm{X}(0) = \mathrm{X}_0$. Si B es un subconjunto no–vacío de $\mathbb{R}^d$ definimos

$$\tau_B := \inf\{t \geq 0 \mid \mathrm{X}(t) \in B\}.$$

Si B es abierto o cerrado, entonces τ_B es un tiempo de paro con respecto a la filtración natural de $\mathrm{X}(\cdot)$. (De hecho, este resultado es cierto para cualquier conjunto de Borel B pero la demostración es más complicada.)

Demostración. Sea $\{t_i,\ i = 1, 2, \ldots\}$ un conjunto numerable y denso en $[0, \infty)$. Supóngase primero que B es abierto. Dado $t \geq 0$, el evento

$$\{\tau_B \leq t\} = \bigcup_{t_i \leq t} \{\mathrm{X}(t_i) \in B\}$$

pertenece a $\mathcal{F}_t^{\mathrm{X}}$ porque $\{\mathrm{X}(t_i) \in B\} \in \mathcal{F}_{t_i}^{\mathrm{X}} \subset \mathcal{F}_t^{\mathrm{X}}$ para todo $t_i \geq t$.

Ahora supóngase que B es cerrado. Para cada $n = 1, 2, \ldots$ considere el conjunto abierto

$$B_n := \{x \in \mathbb{R}^d \mid \mathrm{dist}\,(x, B) < 1/n\}.$$

Entonces el evento

$$\{\tau_B \leq t\} = \bigcap_{n=1}^{\infty} \bigcup_{t_i \leq t} \{\mathrm{X}(t_i) \in B_n\}$$

pertenece a $\mathcal{F}_t^{\mathrm{X}}$. $\square$ $\square$

En contraste con el Ejemplo 26.14, la v.a. $\sigma := \sup\{t \geq 0 \mid \mathrm{X}(t) \in B\}$, que representa el último instante de tiempo que $\mathrm{X}(\cdot)$ está en B, en general *no* es un tiempo de paro porque el evento $\{\sigma \leq t\}$ puede depender de toda la historia *futura* del proceso $\mathrm{X}(\cdot)$, no sólo de $\mathcal{F}_t^{\mathrm{X}}$.

Definición 26.15. Si $f \in N[0, T]$ y τ es un tiempo de paro con respecto a $\mathcal{F}^W$ y tal que $0 \leq \tau \leq T$, definimos

$$\int_0^\tau f(t) dW(t) := \int_0^T f(t) I_{\{t \leq \tau\}} dW(t). \tag{26.41}$$

Observe que el PE $f(t) I_{\{t \leq \tau\}}, 0 \leq t \leq T$, en (26.41) está en $N[0, T]$ y por lo tanto se cumple lo siguiente.

Lema 26.16. La integral en (26.41) satisface:

$$E\left(\int_0^\tau f\, dW\right) = 0 \tag{26.42}$$

$$E\left(\int_0^\tau f\, dW\right)^2 = E\left(\int_0^\tau f^2(t) dt\right). \tag{26.43}$$

Sea $g(t,x)$ una función de clase $C^{1,2}$ en $[0,\infty)\times\mathbb{R}^m$ y sea $\mathrm{X}(t)\in\mathbb{R}^m$ la solución de (1) para $t\geq 0$. Por la fórmula de Ito (Teorema 25.7)

$$dg(t,\mathrm{X}(t)) = Lg(t,\mathrm{X}(t))dt + g_x(t,x(t))G(t,\mathrm{X}(t))dW(t), \tag{26.44}$$

donde Lg es el operador definido en (25.25.16), i.e.

$$Lg = g_t + g_x f(t,x) + \frac{1}{2}Tr[\sigma g_{xx}], \quad \text{con} \quad \sigma := GG'. \tag{26.45}$$

Reescribiendo (26.44) en forma integral tenemos

$$g(t,\mathrm{X}(t)) = g(0,\mathrm{X}(0)) + \int_0^t (Lg)ds + \int_0^t g_x G\,dW \quad \forall\, 0\leq t\leq T.$$

Por lo tanto, si $g_x(t,\mathrm{X}(t))G(t,\mathrm{X}(t))$ está en $N[0,T]$ y τ es un tiempo de paro con $0\leq\tau\leq T$, entonces (usando (26.41)) obtenemos la **fórmula de Ito con tiempos de paro**

$$g(\tau,\mathrm{X}(t)) = g(0,\mathrm{X}(0)) + \int_0^\tau (Lg)ds + \int_0^\tau g_x G\;dW. \tag{26.46}$$

Además, usando (26.42), se obtiene de (26.46) la siguiente forma general de la **fórmula de Dynkin** (ver el Ejercicio 25.8):

$$E[g(\tau,\mathrm{X}(\tau))] = E[g(0,\mathrm{X}(0))] + E[\int_0^\tau (Lg)ds].$$

Ejercicios § 26

26.1. Calcule la solución $(\mathrm{X}_1(\bullet),\mathrm{X}_2(\bullet))'$ de (19) mediante la fórmula (14).

26.2. Considere la ecuación lineal en el sentido retringido (11) y suponga que la condición inicial C satisface que $E(C^2)<\infty$. Demuestre:

(a) La media $m_{\mathrm{X}}(t) := E\mathrm{X}(t)$ satisface (12) con condición inicial $m_{\mathrm{X}}(t_0) = EC$, i.e.

$$\dot{m}_{\mathrm{X}}(t) = A(t)m_{\mathrm{X}}(t) + a(t) \quad \text{para} \quad t\geq t_0, \quad \text{con} \quad m_{\mathrm{X}}(t_0) = EC.$$

(b) La covarianza $K_{\mathrm{X}}(s,t)$ de $\mathrm{X}(\cdot)$ está dada por

$$K_{\mathrm{X}}(s,t) = \Phi(s)\ \{E[(C-EC)(C-EC)'] \\ + \int_{t_0}^{\min(s,t)} \Phi(r)^{-1}B(r)B(r)'(\Phi(r)^{-1})'dr\}\Phi(t)'.$$

(c) La matriz de covarianza de $\mathrm{X}(t)$, es decir $K(t) := K_{\mathrm{X}}(t,t)$, es la única solución simétrica y positiva definida de la ecuación matricial

$$\dot{K}(t) = A(t)K(t) + K(t)A(t)' + B(t)B(t)', \quad \text{con} \\ K(t_0) = E[(C-EC)(C-EC)'].$$

26.3. Considérese la ecuación lineal en el sentido restringido (11). Demuestre que si la condición inicial $\mathrm{X}(t_0) = C$ es una v.a. gaussiana o una constante, entonces $\mathrm{X}(\cdot)$ es un PE gaussiano (cuya media y covarianza están dados en el Ejercicio 2.)

26.4. Considérese la ecuación lineal autónoma escalar ($m = 1$)

$$d\mathrm{X}(t) = A\mathrm{X}(t)dt + \mathrm{X}(t)\sum_{i=1}^{d} B_i dW_i(t) \quad \forall\, t \geq 0, \quad \text{con} \quad \mathrm{X}(0) = C.$$

Demuestre que esta EDE tiene la solución

$$\mathrm{X}(t) = C\, e^{(A-\frac{1}{2}\sum_{i=1}^{d} B_i^2)t + \sum_{i=1}^{d} B_i dW_i(t)}.$$

(*Sugerencia*: use el resultado en el Ejercicio 25.4(b).)

26.5. Demuestre el **Lema de Bellman–Gronwall**: Sean $g(\cdot)$ y $h(\cdot)$ funciones integrables sobre $[t_0,T]$, con $g(\cdot) \geq 0$, y tales que, para alguna constante $L > 0$,

$$g(t) \leq L\int_{t_0}^{t} g(s)ds + h(t) \quad \forall t\ \in [t_0,T]. \tag{$*$}$$

Entonces

$$g(t) \leq h(t) + L\int_{t_0}^{t} e^{L(t-s)}h(s)ds \quad \forall\, t \in [t_0,T].$$

En particular, si $(*)$ se cumple con $h(\cdot) \equiv 0$, entonces $g(\cdot) \equiv 0$.

26.6. Suponga que se satisfacen las "condiciones de Ito" del Teorema 26.2 y que $\mathrm{X}(\cdot)$ es solución de la EDE (1)–(2) en el caso escalar $m = d = 1$. Demuestre que si la condición inicial $\mathrm{X}(t_0) = C$ satisface que $E(C^2) < \infty$, entonces existen constantes $M_1, M_2 \geq 0$ tales que

(a) $E[\mathrm{X}^2(t)] \leq M_1 \quad \forall\, t \in [t_0, T]$, y

(b) $\int_{t_0}^{T} E[G(s, \mathrm{X}(s))^2]ds < \infty$.

(*Sugerencia*: nótese que (b) se sigue de (a) y de la condición de crecimiento lineal (4). Para obtener (a), primero aplique a (2) la desigualdad $(\sum_{i=1}^{n} x_i)^2 \leq n \sum_{i=1}^{n} x_i^2$. Después use la isometría de Ito, la condición (4) y el Lema de Bellman–Gronwall (Ejercicio 5).)

26.7. Verifique que $\mathrm{X}(t) := e^{W(t)}$ satisface la EDE $d\mathrm{X}(t) = \frac{1}{2}\mathrm{X}(t)dt + \mathrm{X}(t)dW(t)$ por dos métodos distintos:

(a) usando la regla de Ito, y

(b) mediante el resultado en el Ejercicio 4.

26.8. Verifique que el PE dado en cada uno de los siguientes incisos satisface la EDE dada:

(a) $\mathrm{X}(t) = W(t)/(1+t)$; $d\mathrm{X}(t) = -\frac{\mathrm{X}(t)}{1+t}dt + \frac{1}{1+t}dW(t)$, $\mathrm{X}(0) = 0$.

(b) $(\mathrm{X}_1(t), \mathrm{X}_2(t))' = (\cos W(t), \operatorname{sen} W(t))'$;

$$\begin{pmatrix} d\mathrm{X}_1(t) \\ d\mathrm{X}_2(t) \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} \mathrm{X}_1(t) \\ \mathrm{X}_2(t) \end{pmatrix} dt + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \mathrm{X}_1(t) \\ \mathrm{X}_2(t) \end{pmatrix} dW(t).$$

(c) $(\mathrm{X}_1(t), \mathrm{X}_2(t))' = (\cosh W(t), \operatorname{senh} W(t))'$;

$$\begin{pmatrix} d\mathrm{X}_1(t) \\ d\mathrm{X}_2(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \mathrm{X}_1(t) \\ \mathrm{X}_2(t) \end{pmatrix} dt + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \mathrm{X}_1(t) \\ \mathrm{X}_2(t) \end{pmatrix} dW(t).$$

(d) $(\mathrm{X}_1(t), \mathrm{X}_2(t))' = (t, e^t W(t))'$;

$$\begin{pmatrix} d\mathrm{X}_1(t) \\ d\mathrm{X}_2(t) \end{pmatrix} = \begin{pmatrix} 1 \\ \mathrm{X}_2(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ e^{\mathrm{X}_1(t)} \end{pmatrix} dW(t).$$

26.9. Resuelva las siguientes EDEs:

(a) $\begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & X_1(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}$.

(b) $dX(t) = X(t)dt + dW(t)$

(c) $dX(t) = -X(t)dt + e^{-t}dW(t)$.

26.10. (**El puente Browniano de** a **a** b **sobre** $[0, T]$.) Considére el PE

$$X(t) := a(1 - \frac{t}{T}) + b\frac{t}{T} + (T - t)\int_0^t \frac{dW(s)}{T - s} \quad \text{para} \quad t \in [0, T].$$

Nótese que $X(0) = a$ y $X(T) = b$ y, por tal motivo, $X(\cdot)$ se conoce como el "puente Browniano" de a a b sobre $[0, T]$. Demuestre que:

(a) $X(\cdot)$ es un PE gaussiano, con trayectorias continuas c.s., y con media y función de covarianza

$$m_X(t) = EX(t) = a(1 - \frac{t}{T}) + b\frac{t}{T},$$
$$K_X(s, t) = \min(s, t) - \frac{st}{T},$$

respectivamente, para todo $0 \leq s, t \leq T$. (*Sugerencia*: use la Proposición 22.3.)

(b) $X(\cdot)$ es solución de la EDE $dX(t) = \frac{b - X(t)}{T - t}dt + dW(t)$ para $0 \leq t \leq T$, con $X(0) = a$. (Nótese que esta EDE es de la forma (11) con $m = d = 1$.)

(c) $X(t) \to b$ en L_2 cuando $t \to T-$.

26.11. Sea $dX(t) = \alpha X(t)dt + \beta X(t)dW(t)$ para $t \geq 0$, con $X(0) = x$ y $\beta > 0$. Demuestre que $\log\, [X(t)/x] \sim N((\alpha - \frac{1}{2}\beta^2)t,\ \beta^2 t)$. (*Sugerencia*: use la regla de Ito para calcular $d(\log\ X(t))$. Vea también el Ejemplo 24.5(c). Al PE $X(\cdot)$ se le llama **movimiento browniano geométrico**.)

26.12. Sea $X(\cdot)$ como en el Ejercicio 11 y supóngase que $E[X^4(t)] < \infty$ para todo $t \in [0, T]$. Demuestre:

(a) $dX^2 = (2\alpha + \beta^2)X^2 dt + 2\beta X^2 dW$, y

(b) $E\mathrm{X}^2(t) = x^2 + (2\alpha+\beta^2)\int_0^t E\mathrm{X}^2(s)ds$, de modo que $E\mathrm{X}^2(t) = x^2 e^{(2\alpha+\beta^2)t}$.

(c) Suponiendo que $E[\mathrm{X}^{2n}(t)] < \infty$ para $t \in [0,T]$ y $n = 1, 2, \ldots$, demuestre que

$$E\mathrm{X}^n(t) = x^n \ \exp[(n\alpha + \frac{n(n-1)}{2}\beta^2)t] \quad \forall\, n = 1, 2, \ldots.$$

26.13. (Oscilador armónico perturbado aleatoriamente). Sean

$$d\mathrm{X}_1(t) = \mathrm{X}_2(t)dt, \ d\mathrm{X}_2(t) = -\mathrm{X}_1(t)dt + \alpha\mathrm{X}_1(t)dW(t), \quad \text{con} \quad \alpha > 0.$$

Calcule ecuaciones para los momentos $E\mathrm{X}_1^2$, $E\mathrm{X}_2^2$, $E(\mathrm{X}_1\mathrm{X}_2)$ y para la "energía total" $E|\mathrm{X}(t)|^2$.

26.14. Sea $d\mathrm{X}(t) = A(t)\mathrm{X}(t)dt + B(t)dW(t)$ para $s \leq t \leq T$, con $\mathrm{X}(\cdot) \in \mathbb{R}^m$ y $W(\cdot) \in \mathbb{R}^d$. Sean $m(t)$ y $K(t)$ la media y la matriz de covarianza de $\mathrm{X}(t)$, i.e.

$$m(t) := E\mathrm{X}(t) \quad \text{y} \quad K(t) := E[(\mathrm{X}(t) - m(t))(\mathrm{X}(t) - m(t))'].$$

Demuestre que $m(\cdot)$ y $K(\cdot)$ satisfacen las ecuaciones diferenciales ordinarias

$$\dot{m}(t) = A(t)m(t), \quad \dot{K}(t) = A(t)K(t) + K(t)A(t)' + \sigma(t),$$

con $\sigma := BB'$, suponiendo que la condición inicial $\mathrm{X}(s) = C$ es una v.a. gaussiana o una constante. (Compare con el Ejercicio 2.)

26.15. Sean τ, τ_1 y τ_2 tiempos de paro con respecto a una filtración $\{\mathcal{F}_t,\ t \geq 0\}$. Demuestre que

(a) $\{\tau < t\} \in \mathcal{F}_t$, y por tanto $\{\tau = t\} \in \mathcal{F}_t$, para todo $t \geq 0$.

(b) $\min(\tau_1, \tau_2)$ y $\max(\tau_1, \tau_2)$ son tiempos de paro.

26.16. Demuestre (26.42) y (26.43).

27 Apéndice

Contenido: Espacio normado, producto interno, convergencia en L_2.

A. Terminología de análisis

Definición 27.1. Sea V un espacio lineal o espacio vectorial (sobre $\mathbb{R}$). Una **seminorma** sobre V es una función $\| \cdot \|$ de V en $\mathbb{R}_+ := [0, \infty)$ que satisface:

(a) $x = 0 \;\Rightarrow \| x \| = 0$.

(b) Homogeneidad absoluta: $\| ax \| = |a| \, \| x \| \;\; \forall\, a \in \mathbb{R},\, x \in V$.

(c) Subaditividad o desigualdad del triángulo:

$$\| x + y \| \leq \| x \| + \| y \| \;\; \forall\, x, y \in V.$$

Si además se cumple el recíproco de (a), i.e.

(a') $\| x \| = 0 \Rightarrow x = 0$,

se dice que $\| \cdot \|$ es una **norma**. En este caso se dice que la pareja $(V, \| \cdot \|)$ es un **espacio lineal normado**.

Si $(V, \| \cdot \|)$ es un espacio lineal normado, entonces la función $d : V \times V \to \mathbb{R}_+$ definida por $d(x, y) := \| x - y \|$ es una **métrica** (o distancia) en V, y se dice (V, d) es un **espacio métrico**. Este espacio es **completo** si toda sucesión de Cauchy (con respecto a d) converge; es decir, si $\{x_n\} \subset V$ es una sucesión tal que $d(x_n, x_m) \to 0$ cuando $n, m \to \infty$, entonces existe $x \in \mathrm{X}$ tal que $x = \lim\limits_{n\to\infty} x_n$.

Un espacio lineal normado y completo se llama **espacio de Banach**.

Hipótesis 27.2. Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$, identificaremos cualquiera dos vv.aa. X y Y tales que $\mathrm{P}(\mathrm{X} = Y) = 1$. (En otras palabras, si $\mathrm{P}(\mathrm{X} \neq Y) = 0$, entonces $\mathrm{X} = Y$.)

Ejemplo 27.3. Sea $p \geq 1$, y sea $L_p \equiv L_p(\Omega, \mathcal{F}, \mathrm{P})$ el conjunto de todas las vv.aa. X tales que $E|\mathrm{X}|^p < \infty$. Entonces

$$\| \mathrm{X} \|_p := (E|\mathrm{X}|^p)^{1/p} \tag{27.1}$$

define una **seminorma** sobre L_p. En efecto, $\|\mathrm{X}\|_p = 0$ **no** implica que $\mathrm{X} = 0$; sólo implica que $\mathrm{P}(\mathrm{X} = 0) = 1$. Sin embargo, bajo la Hipótesis 27.2, la expresión (1), define una norma sobre L_p. De hecho, $(L_p, \| \cdot \|_p)$ es un **espacio de Banach**.

Definición 27.4. Un **producto interno** sobre un espacio vectorial V es una función $(x, y) \mapsto \langle x, y \rangle$ de $V \times V \to \mathbb{R}$ que satisface:

(a) $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle \quad \forall\, a, b \in \mathbb{R};\ x, y \in V$

(b) $\langle x, y \rangle = \langle y, x \rangle \quad \forall\, x, y \in V$

(c) $\langle x, x \rangle \geq 0 \quad \forall\, x \in \mathrm{X}$

(d) $\langle x, x \rangle = 0$ ssi $x = 0$.

Proposición 27.5. (**Propiedades de** $\langle \cdot, \cdot \rangle$) Sea $\langle \cdot, \cdot \rangle$ un producto interno sobre un espacio vectorial V, y sea

$$\| x \| := \sqrt{\langle x, x \rangle}. \tag{27.2}$$

(a) **Desigualdad de Cauchy–Schwarz:** $|\langle x, y \rangle| \leq \| x \| \| y \| \quad \forall\, x, y \in V$. Además, se cumple la igualdad ssi x, y son linealmente dependientes (es decir, existe $a \in \mathbb{R}$ tal que $x + ay = 0$).

(b) La expresión (2) define una **norma** sobre V.

(c) El producto interno es continuo en ambas variables, es decir, si $x_n \to x$ y $y_m \to y$, entonces $\langle x_n, y_m \rangle \to \langle x, y \rangle$.

Definición 27.6. Un espacio vectorial V con un producto interno $\langle \cdot, \cdot \rangle$ se dice que es un espacio **pre–Hilbert**. Si además es un espacio completo con respecto a la norma definida por (2), entonces V es un **espacio de Hilbert**. (En particular, un espacio de Hilbert es de Banach.)

Ejemplo 27.7. En el Ejemplo 27.3 tómese $p = 2$. Entonces $L_2 \equiv L_2(\Omega, \mathcal{F}, \mathrm{P})$ con el producto interno

$$\langle \mathrm{X}, Y\rangle := E(\mathrm{X}\bar{Y}) \tag{27.3}$$

es un **espacio de Hilbert**; por (2), la **norma** en L_2 es

$$\| \mathrm{X} \|_2 := \sqrt{\langle \mathrm{X}, \mathrm{X}\rangle} = \sqrt{E(\mathrm{X}^2)}. \tag{27.4}$$

(Recuérdese la Hipótesis 27.2.) En este caso podemos escribir la **desigualdad de Cauchy–Schwarz** como

$$|E(\mathrm{X}Y)| \leq \| \mathrm{X} \|_2 \cdot \| Y \|_2 = \sqrt{E(\mathrm{X}^2)} \cdot \sqrt{E(Y^2)}. \tag{27.5}$$

Convergencia en L_2: Por (4)

$$\mathrm{X}_n \to \mathrm{X} \;\text{ en }\; L_2 \;\text{ ssi }\; \| \mathrm{X}_n - \mathrm{X} \|_2 \to 0 \;\text{ ssi }\; E|\mathrm{X}_n - \mathrm{X}|^2 \to 0.$$

Además, por la continuidad de un producto interno (ver Proposición 27.5 (c)) tenemos:

Lema 27.8. Si $\mathrm{X}_n \to \mathrm{X}$ y $Y_m \to Y$ en L_2 entonces

$$\langle \mathrm{X}_n, Y_m\rangle = E(\mathrm{X}_n Y_m) \to E(\mathrm{X}Y) = \langle \mathrm{X}, Y\rangle.$$

Lema 27.9. (**Criterio para la existencia de un límite en L_2.**) Sea $\mathrm{X}(\cdot) = \{\mathrm{X}(t),\, t \in T\}$ un PE en L_2, y sea $t_0 \in T$. Las siguientes proposiciones son equivalentes:

(a) Existe una v.a. $\mathrm{X} \in L_2$ tal que $\mathrm{X}(t) \to \mathrm{X}$ en L_2 cuando $t \to t_0$.

(b) Existe un número $\ell \in \mathbb{R}$ tal que

$$E[\mathrm{X}(t_n)\mathrm{X}(t'_m)] \to \ell \quad \text{cuando} \quad n, m \to \infty \tag{27.6}$$

para cualesquiera dos sucesiones $t_n \to t_0$, $t'_m \to t_0$.

Demostración. (a)⇒(b). Si $t_n \to t_0$ y $t'_m \to t_0$, entonces (por (a))

$$\mathrm{X}(t_n) \to \mathrm{X} \quad \text{y} \quad \mathrm{X}(t'_m) \to \mathrm{X} \quad \text{en} \quad L_2.$$

Luego, por el Lema 27.8,

$$E[\mathrm{X}(t_n)\mathrm{X}(t'_m)] \to E(\mathrm{X}^2) =: \ell.$$

(b)⇒(a). Supóngase que $t_n \to t_0$ y en (b) tómese $t'_m \equiv t_m$. Entonces, por (6),

$$
\begin{aligned}
E[(\mathrm{X}(t_n) - \mathrm{X}(t_m))^2] &= E[\mathrm{X}^2(t_n)] + E[\mathrm{X}^2(t_m)] - 2E[\mathrm{X}(t_n)\mathrm{X}(t_m)] \\
&\to \ell + \ell - 2\ell = 0 \quad \text{cuando} \quad n, m \to \infty;
\end{aligned}
$$

es decir, $\| \mathrm{X}(t_n) - \mathrm{X}(t_m) \|_2 \to 0$ lo cual significa que $\{\mathrm{X}(t_n)\}$ es una sucesión de Cauchy en L_2. Luego, como L_2 es **completo**, existe $\mathrm{X} \in L_2$ tal que $\mathrm{X}(t_n) \to \mathrm{X}$ en L_2.

Ahora sea $\{s_n\}$ cualquier otra sucesión con $s_n \to t_0$. Entonces

$$
\| \mathrm{X}(s_n) - \mathrm{X} \|_2 \leq \| \mathrm{X}(s_n) - \mathrm{X}(t_n) \|_2 + \| \mathrm{X}(t_n) - \mathrm{X} \|_2,
$$

y como $\| \mathrm{X}(t_n) - \mathrm{X} \|_2 \to 0$ y

$$
\begin{aligned}
E[(\mathrm{X}(s_n) - \mathrm{X}(t_n))^2] &= E[\mathrm{X}^2(s_n)] + E[\mathrm{X}^2(t_n)] - 2E[\mathrm{X}(s_n)\mathrm{X}(t_n)] \\
&\to \ell + \ell - 2\ell = 0,
\end{aligned}
$$

concluimos que $\mathrm{X}(s_n) \to \mathrm{X}$. $\square$

28 Bibliografía

D. Applebaum (2004). *Lévy Processes and Stochastic Calculus*, Cambridge University Press.

L. Arnold (1974). *Stochastic Differential Equations*, Wiley.

R.B. Ash (1970). *Basic Probability Theory*. Wiley.

R.B. Ash (1972). *Real Analysis and Probability*, Academic Press. Segunda edición (2000): *Probability and Measure Theory*.

R.B. Ash, M.F. Gardner (1975). *Topics in Stochastic Processes*, Academic Press.

J. Bertoin (1996). *Lévy Processes*, Cambridge University Press.

L. Breiman (1968). *Probability*, Addison–Wesley. (Second edition, SIAM, 1992.)

R.M. Dudley (2003). *Real Analysis and Probability, Second Edition*, Cambridge University Press.

R. Durrett (1999). *Essentials of Stochastic Processes*, Springer–Verlag.

E.B. Dynkin, A.A. Yushkevich (1956). "Strong Markov processes", *Theory Probab. Appl.*, Vol. 1, pp. 134–139.

A. Friedman (1975). *Stochastic Differential Equations and Applications*, Vol. 1, Academic Press.

I.I. Gikhman, A.V. Skorokhod (1969). *Introduction to the Theory of Random Processes*, Saunders.

D.P. Heyman, M.J. Sobel (1982). *Stochastic Models in Operations Research*, Vol. 1, McGraw–Hill.

P.G. Hoel, S.C. Port, C.J. Stone (1972). *Introduction to Stochastic Processes*, Houghton Mifflin.

O. Kallenberg (2002). *Foundations of Modern Probability, Second Edition*, Springer–Verlag.

J. Jacod, P. Protter (2003). *Probability Essentials, Second Edition.* Springer.

I. Karatzas, S.E. Shreve (1991). *Brownian Motion and Stochastic Calculus, Second Edition*, Springer–Verlag.

D. Khoshnevisan (2007). *Probability.* American Math. Soc.

R.G. Laha, V.K. Rohatgi (1979). *Probability Theory*, Wiley.

T. Mikosch (1998). *Elementary Stochastic Calculus with Finance in View*, World Scientific.

J.R. Norris (1997). *Markov Chains*, Cambridge University Press.

B. Oksendal (1998). *Stochastic Differential Equations, Fifth Edition*, Springer–Verlag.

E. Parzen (1962). *Stochastic Process*, Holden–Day.

I.K. Rana (2002). *An Introduction to Measure Theory and Integration. Second Edition.* American Math. Soc.

J.S. Rosenthal (2000). *A First Look at Rigorous Probability Theory.* World Scientific.

K.–I. Sato (1999). *Lévy Processes and Infinite Divisibility*, Cambridge University Press.

H.G. Tucker (1967). *A Graduate Course in Probability*, Academic Press.