Procesos Estocásticos I
Semestre 2013-II

Gerónimo Uribe Bravo
Instituto de Matemáticas
Universidad Nacional Autónoma de México

CAPÍTULO 1

Martingalas

En este capítulo nos enfocaremos en el estudio de las martingalas. Esta es una clase de procesos fundamental para la teoría moderna de la probabilidad. Tanto así que la herramienta teórica sobre la cual se construye la teoría moderna de las finanzas matemáticas (el llamado cálculo estocástico) es una teoría basada en las martingalas.

1. Recordatorio sobre esperanza condicional

Si $(\Omega, \mathscr{F}, \mathbb{P})$ es un espacio de probabilidad y $B \in \mathscr{F}$ es tal que $\mathbb{P}(B) > 0$, podemos definir la probabilidad condicional de A dado B mediante la fórmula

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

que se puede entender a través de la interpretación frecuentista de la probabilidad. Así, para una variable aleatoria discreta[1] estamos acostumbrados a expresiones como $\mathbb{P}(A \,|X = j)$ y a la conotación que que se les ha dado. Desafortunadamente, una extensión del concepto de probabilidad condicional a eventos cualquiera no es tan inmediata[2], por lo que primero desarrollaremos algunas propiedades de la esperanza condicional que nos permitan entender la solución que se le ha dado a este problema de extensión, definiendo algunos conceptos y verificando algunas propiedades de las variables aleatorias que nos faciliten el camino.

1.1. Preliminares. A lo largo de la sección, $(\Omega, \mathscr{F}, \mathbb{P})$ designará a un espacio de probabilidad arbitrario y al las funciones medibles de Ω en $\mathbb{R}$ las llamaremos variables aleatorias reales, aunque generalmente se omitirá la palabra reales. Si $X : \Omega \to \mathbb{R}$ es una variable aleatoria en $(\Omega, \mathscr{F}, \mathbb{P})$ se utilizará la notación

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

También, m_n (m) representará a la medida de Lebesgue sobre los Borelianos de $\mathbb{R}^n$ $(\mathbb{R})$.

Nota. Si $f : \mathbb{R} \to \mathbb{R}$ es Borel medible, entonces $f \circ X$ es borel medible, por lo que está definida su esperanza cuando la integral de la composición esté definida.

[1] Esto es, una función $X : \Omega \to \mathbb{R}$ Borel medible tal que $X(\Omega)$ sea a lo más numerable.

[2] ¿Qué pasaría en el caso de eventos condicionantes de probabilidad cero?

Definición. Si $X : \Omega \to \mathbb{R}$ es una variable aleatoria, la medida de probabilidad inducida por X, es la función $\mathbb{P}_X : \mathscr{B}_{\mathbb{R}} \to [0,1]$ dada por:

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B)\,.$$

Nota. Si $\mathbb{P}_X$ es absolutamente continua con respecto a la medida de Lebesgue sobre los Borelianos de $\mathbb{R}$, diremos que X es absolutamente continua y en este caso, por el teorema de Radon-Nikodym existe una densidad $g_X : \mathbb{R} \to \mathbb{R}$ tal que

$$\mathbb{P}(X \in B) = \mathbb{P}_X(B) = \int_B g_X \, dm.$$

Teorema 1.1 (Teorema de Cambio de variable). *Si $X : \Omega \to \mathbb{R}$ es una variable aleatoria en $(\Omega, \mathscr{F}, \mathbb{P})$ y $f : \mathbb{R} \to \mathbb{R}$ es Borel medible tal que la integral de $f \circ X$ está definida, entonces:*

$$\mathbb{E}(f \circ X) = \int f \, d\mathbb{P}_X.$$

Ejercicio 1.1. Sea X una variable aleatoria normal centrada de varianza 1. Utilice el teorema de cambio de variable para calcular

$$\mathbb{E}\big(X^{2n}\big)$$

para toda $n \in \mathbb{N}$.

Proposición 1.1. *Sea $Z : \Omega \to \mathbb{R}$ una variable aleatoria fija en $(\Omega, \mathscr{F}, \mathbb{P})$ y $\mathscr{G} = \sigma(Z)$. Si $X : \Omega \to \mathbb{R}$ es $\mathscr{G}$ medible, entonces existe $f : \mathbb{R} \to \mathbb{R}$ Borel-medible, tal que $X = f \circ Z$.*

1.2. Esperanza Condicional. Si $Z : \Omega \to \mathbb{R}$ es una variable aleatoria simple en $(\Omega, \mathscr{F}, \mathbb{P})$ y $Y : \Omega \to \mathbb{R}$ es una variable aleatoria, una definición natural de la probabilidad condicional $\mathbb{P}(Y \in B \,|Z)$ es la siguiente:

$$\mathbb{P}(Y \in B \,|Z) = \sum_{i \in \mathscr{R}_Z} \mathbb{P}(Y \in B \,|Z = i) \, \mathbf{1}_{\{Z=i\}},$$

donde $\mathscr{R}_Z = Z(\Omega) \subset \mathbb{R}$ es un conjunto finito. Notemos que en este caso, la probabilidad condicional es una función de la variable aleatoria Z, por lo que resulta ser $\sigma(Z)$-medible y que cumple la relación

$$\mathbb{P}(Y \in B, A) = \int_A \mathbb{P}(Y \in B \,|Z) \; d\mathbb{P}, \quad A \in \sigma(Z)\,,$$

que es equivalente a

$$\int_A \mathbf{1}_{Y \in B} \, d\mathbb{P} = \int_A \mathbb{P}(Y \in B \,|Z) \; d\mathbb{P},$$

esto es, obtenemos información (la integral sobre un conjunto) de la variable $\mathbf{1}_{Y \in B}$, que no es necesariamente $\sigma(Z)$-medible a través de la variable $\mathbb{P}(Y \in B \,|Z)$ que si lo es, aunque sea para una clase restringida de eventos ($\sigma(Z)$, que resulta ser una σ-álgebra). Además, en la propiedad anterior de probabilidad condicional, la variable aleatoria Z solo juega un papel secundario, y la σ-álgebra $\sigma(Z)$ se torna imprescindible. Como un comentario adicional, recordemos que dos variables aleatorias Y y Z son iguales $\mathbb{P}$-p.s. si y solo si $\int_A Y\, d\mathbb{P} = \int_A Z\, d\mathbb{P}$ para todo $A \in \mathscr{F}$ (una propiedad parecida a la encontrada en la probabilidad condicional), por lo que la función que a cada elemento A de $\mathscr{F}$ le asigna el número $\int_A Y\, d\mathbb{P}$ (que resulta ser una medida con signo si la integral de Y está definida) determina completamente a la variable aleatoria Y. El comentario anterior puede motivar la definición de esperanza condicional, de la cual la probabilidad condicional es un caso particular[3], en la que se condiciona con respecto a una σ-álgebra:

Definición. Si X es una variable aleatoria en $(\Omega, \mathscr{F}, \mathbb{P})$, y $\mathscr{G} \subset \mathscr{F}$ es una σ-álgebra, la **esperanza condicional** de X dado $\mathscr{G}$, denotada por $\mathbb{E}(X \mid \mathscr{G})$, es una variable aleatoria $\mathscr{G}$-medible que cumple

$$\int_A X\, d\mathbb{P} = \int_A \mathbb{E}(X \mid \mathscr{G})\, d\mathbb{P}$$

para todo $A \in \mathscr{G}$.

Proposición 1.2. *Si $X : \Omega \to \mathbb{R}$ es una variable aleatoria en $(\Omega, \mathscr{F}, \mathbb{P})$ cuya integral está definida y $\mathscr{G} \subset \mathscr{F}$ es una σ-álgebra, entonces existe una variable aleatoria $Y : \Omega \to \mathbb{R}$ tal que $\int_A X\, d\mathbb{P} = \int_A Y\, d\mathbb{P}$, para $A \in \mathscr{G}$. Además, si Z cumple la misma propiedad, entonces $Y = Z$ casi seguramente respecto a $\mathbb{P}|_{\mathscr{G}}$.*

Ejercicio 1.2. Si (X, Y) son dos variables aleatorias con densidad conjunta $f(x, y)$, pruebe que:

$$\mathbb{E}(g(Y) \mid X) = \frac{\int f(X, y)\, g(y)\, dy}{\int f(X, y)\, dy}.$$

Ejercicio 1.3. Sean $X_1, X_2, \ldots$ vaiids. Sea K una variable aleatoria independiente de $X_1, X_2, \ldots$ y con valores en $\mathbb{N}$. Cacule

$$\mathbb{E}(X_1 + \cdots + X_K \mid K).$$

Sugerencia: ¿Qué pasa cuando K toma sólo un valor?

1.3. Propiedades de la esperanza condicional. Las siguientes son algunas propiedades de la esperanza condicional, en las que consideramos $\mathscr{G} \subset \mathscr{F}$ una σ-álgebra. Si X y Y son variables aleatorias, la ecuación $Y = \mathbb{E}(X \mid \mathscr{G})$ significa que el lado izquierdo de la ecuación es $\mathscr{G}$-medible y que $\int_A Y\, d\mathbb{P} = \int_A X\, d\mathbb{P}$ para $A \in \mathscr{G}$. Consideraremos solo variables aleatorias cuya integral esté definida, por lo que la existencia de la esperanza condicional queda garantizada.

Propiedad 1 (Linealidad de la esperanza condicional). *Si X y Y son variables aleatorias integrables y $a, b \in \mathbb{R}$, entonces $\mathbb{E}(aX + bY \mid \mathscr{G})$ existe y es igual a $a\mathbb{E}(X \mid \mathscr{G}) + b\mathbb{E}(Y \mid \mathscr{G})$.*

[3] Se utilizará la relación $\mathbb{P}(A) = \mathbb{E}(\mathbf{1}_A)$ para este efecto.

Propiedad 2 (Monotonía de la esperanza condicional). *Si X es no negativa $\mathbb{P}$-p.s. entonces $\mathbb{E}(X \mid \mathscr{G})$ existe y es no negativa $\mathbb{P}|_{\mathscr{G}}$-p.s..*

Propiedad 3. *Si la integral de X está definida entonces $|\mathbb{E}(X \mid \mathscr{G})| \leq \mathbb{E}(|X| \mid \mathscr{G})$.*

Propiedad 4. *Si X es $\mathscr{G}$-medible, entonces $\mathbb{E}(X \mid \mathscr{G}) = X$.*

Propiedad 5. *Si X es independiente de $\mathscr{G}$ entonces $\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(X)$.*

Propiedad 6. $\mathbb{E}(\mathbb{E}(X \mid \mathscr{G})) = \mathbb{E}(X)$.

Propiedad 7 (Propiedad de torre). *Si $\mathscr{D} \subset \mathscr{G}$ y $\mathscr{D}$ es σ-álgebra entonces*

$$\mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{D}) = \mathbb{E}(X \mid \mathscr{D}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{D}) \mid \mathscr{G}).$$

Propiedad 8 (Teorema de Convergencia Monótona para la Esperanza Condicional). *Si $(X_n)_{n\in\mathbb{N}}$ son variables aleatorias tal que $0 \leq X_n \leq X_{n+1}$ y $X = \lim_{n\to\infty} X_n$, entonces $\mathbb{E}(X \mid \mathscr{G})$ existe y*

$$\lim_{n\to\infty} \mathbb{E}(X_n \mid \mathscr{G}) = \mathbb{E}(X \mid \mathscr{G}).$$

Propiedad 9 (Lema de Fatou para la Esperanza Condicional). *Si $X_n \geq 0$ para $n \in \mathbb{N}$ entonces existe*

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \,\middle|\, \mathscr{G} \right)$$

y

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \,\middle|\, \mathscr{G} \right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n \mid \mathscr{G}).$$

Propiedad 10 (Teorema de Convergencia Dominada para la Esperanza Condicional). *Si*

$$(X_n)_{n\in\mathbb{N}} \subset \mathscr{L}_1(\mathbb{P})$$

es puntualmente convergente y existe $Y \in \mathscr{L}_1(\mathbb{P})$ tal que $|X_n| \leq Y$ para $n \in \mathbb{N}$, entonces

$$\mathbb{E}\left(\lim_{n\to\infty} X_n \,\middle|\, \mathscr{G} \right)$$

existe y es igual a $\lim_{n\to\infty} \mathbb{E}(X_n \mid \mathscr{G})$ (donde la existencia de este último límite solo se asegura $\mathbb{P}|_{\mathscr{G}}$-p.s.).

Propiedad 11 ($\mathscr{G}$-homogeneidad). *Si X_1 y X_2 son variables aleatorias integrables tales que X_1X_2 es integrable y X_1 es $\mathscr{G}$-medible entonces*

$$\mathbb{E}(X_1X_2 \mid \mathscr{G}) = X_1\mathbb{E}(X_2 \mid \mathscr{G}).$$

(Note que la hipótesis de integrabilidad del producto quedaría garantizada si X_1 y X_2 pertenecen a $\mathscr{L}_2(\mathbb{P})$).

Propiedad 12. *Si $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ es boreliana, la integral de $f(X, Y)$ existe, $Y \perp \mathscr{G}$ y X es $\mathscr{G}$-medible entonces $\mathbb{E}(f(X, Y) \mid \mathscr{G}) = \int_{\mathbb{R}} f(X, y)\, \mathbb{P}_Y(dy)$.*

Como $Y \perp \mathscr{G}$, entonces $Y \perp X$ y si $f(x,y) = \mathbf{1}_A(x)\mathbf{1}_B(y)$, se puede asegurar que

$$\mathbb{E}(\, f(X,Y) \mid \mathscr{G}) = \mathbb{E}(\, \mathbf{1}_A(X)\mathbf{1}_B(Y) \mid \mathscr{G})$$

y como la integral de este producto de funciones existe, entonces

$$\mathbb{E}(\, \mathbf{1}_A(X)\mathbf{1}_B(Y) \mid \mathscr{G}) = \mathbf{1}_A(X)\mathbb{E}(\, \mathbf{1}_B(Y) \mid \mathscr{G})\,.$$

Dado que $\mathbf{1}_B(Y) \perp \mathscr{G}$, entonces

$$\mathbb{E}(\, \mathbf{1}_A(X)\mathbf{1}_B(Y) \mid \mathscr{G}) = \mathbf{1}_A(X)\mathbb{P}(Y \in B) = \mathbf{1}_A(X) \int_{\mathbb{R}} \mathbf{1}_B \, \mathbb{P}_Y(d,)$$

por lo que el teorema se cumple para las indicadoras de productos de borelianos. Sea $\mathscr{C}$ la clase de subconjuntos borelianos de $\mathbb{R}^2$ para los cuales

$$\mathbb{E}(\, \mathbf{1}_C(X,Y) \mid \mathscr{G}) = \int \mathbf{1}_C(X,t) \; \mathbb{P}_Y(dy)\,.$$

Hemos visto que $\mathscr{B}_{\mathbb{R}} \times \mathscr{B}_{\mathbb{R}} = \{A \times B : A, B \in \mathscr{B}_{\mathbb{R}}\} \subset \mathscr{C}$ y además, si $A, B \in \mathscr{C}$ y $A \subset B$, entonces

$$\begin{aligned}
\mathbb{E}\big(\, \mathbf{1}_{B \setminus A}(X,Y) \,\big|\, \mathscr{G}\big) &= \mathbb{E}(\, \mathbf{1}_B(X,Y) \mid \mathscr{G}) - \mathbb{E}(\, \mathbf{1}_A(X,Y) \mid \mathscr{G}) \\
&= \int \mathbf{1}_B(X,y) \; \mathbb{P}_Y(dy) - \int \mathbf{1}_A(X,y) \; \mathbb{P}_Y(dy) \\
&= \int \mathbf{1}_{B \setminus A} X, y \, \mathbb{P}_Y(dy)\,,
\end{aligned}$$

por lo que $B \setminus A \in \mathscr{C}$. Si $(A_n)_{n\in\mathbb{N}} \subset \mathscr{C}$ es creciente, entonces $(\mathbf{1}_{A_n})_{n\in\mathbb{N}}$ es una sucesión creciente de funciones no negativas que convergen puntualmente a $\mathbf{1}_{\cup_{n\in\mathbb{N}} A_n}$. Por el teorema de convergencia monótona para la esperanza condicional,

$$\mathbb{E}(\, \mathbf{1}_{\cup_{n\in\mathbb{N}} A_n}(X,Y) \mid \mathscr{G}) = \lim_{n\to\infty} \mathbb{E}(\, \mathbf{1}_{A_n}(X,Y) \mid \mathscr{G}) = \lim_{n\to\infty} \int_{\mathbb{R}} \mathbf{1}_{A_n}(X,y) \, \mathbb{P}_Y(d,)$$

que es igual a

$$\int \mathbf{1}\,(\cup_n A_n)\, X, y \, \mathbb{P}_Y(dy)\,,$$

por el teorema de convergencia monótona usual, de donde $\cup_{n\in\mathbb{N}} A_n \in \mathscr{C}$, por lo que $\mathscr{C}$ es un λ-sistema que contiene al π-sistema $\mathscr{B}_{\mathbb{R}} \times \mathscr{B}_{\mathbb{R}}$ y por el Lema de Clases de Sierpinski, $\sigma(\mathscr{B}_{\mathbb{R}} \times \mathscr{B}_{\mathbb{R}}) = \mathscr{B}_{\mathbb{R}^2} \subset \mathscr{C}$, esto es, el teorema es cierto para $\mathbf{1}_C(X,Y)$ cuando $C \in \mathscr{B}_{\mathbb{R}^2}$. Al aplicar el procedimiento estándar y el teorema de convergencia monótona para la esperanza condicional , obtenemos el resultado deseado.

Nota. En la demostración anterior se utilizó la igualdad $\sigma(\mathscr{B}_{\mathbb{R}} \times \mathscr{B}_{\mathbb{R}}) = \mathscr{B}_{\mathbb{R}^2}$, que se puede verificar al utilizar la separabilidad de $\mathbb{R}^2$ con la topología producto.

Propiedad 13. *Si $\mathscr{H}, \mathscr{G} \subset \mathscr{F}$ son σ-álgebras y $\mathscr{H} \perp \sigma(\mathscr{G}, \sigma(X))$, entonces*

$$\mathbb{E}(\, X \mid \sigma(\mathscr{H}, \mathscr{G})) = \mathbb{E}(\, X \mid \mathscr{G})\,.$$

Sean $Y = \mathbb{E}(X \mid \mathscr{G})$,

$$\mathscr{C} = \{G \cap H : G \in \mathscr{G}, H \in \mathscr{H}\} \quad \text{y} \quad \mathscr{I} = \{A \in \sigma(\mathscr{H}, \mathscr{G}) : \mathbb{E}(\mathbf{1}_A X) = \mathbb{E}(\mathbf{1}_A Y)\}.$$

Nuestro objetivo es demostrar que $\mathscr{I} = \sigma(\mathscr{H}, \mathscr{G})$, en donde solo falta mostrar la contención $\sigma(\mathscr{H}, \mathscr{G}) \subset \mathscr{I}$. Notemos lo siguiente:

$\mathscr{I}$ **es un λ-sistema:** Esto sucede ya que si $A, B \in \mathscr{I}$ y $A \subset B$, entonces

$$\int_{B \setminus A} Y \, d\mathbb{P} = \int_B Y \, d\mathbb{P} - \int_A Y \, d\mathbb{P} = \int_B X \, d\mathbb{P} - \int_A X \, d\mathbb{P} = \int_{B \setminus A} X \, d\mathbb{P}.$$

Además, si $(A_i)_{i \in \mathbb{N}} \subset \mathscr{I}$ es creciente, entonces

$$\int_{\cup_{i \in \mathbb{N}} A_i} Y \, d\mathbb{P} = \lim_{i \to \infty} \int_{A_i} Y \, d\mathbb{P} = \lim_{i \to \infty} \int_{A_i} X \, d\mathbb{P} = \int_{\cup_{i \in \mathbb{N}} A_i} X \, d\mathbb{P}.$$

$\mathscr{C}$ **es π-sistema y $\mathscr{C} \subset \mathscr{I}$:** Si $A, B \in \mathscr{C}$, $A = G_1 \cup H_1$ y $B = G_2 \cup H_2$, entonces $A \cap B = (G_1 \cap G_2) \cup (H_1 \cap H_2) \in \mathscr{C}$. Además, si $A = G \cup H \in \mathscr{C}$, entonces

$$\int_A Y \, d\mathbb{P} = \int_G Y \mathbf{1}_H \, d\mathbb{P} = \mathbb{P}(H) \int_G Y \, d\mathbb{P},$$

puesto que $\mathscr{G} \perp \mathscr{H}$ y Y es $\mathscr{G}$-medible. Por otro lado, tenemos que

$$\int_A X \, d\mathbb{P} = \int_G X \mathbf{1}_H \, d\mathbb{P} = \mathbb{P}(H) \int_G X \, d\mathbb{P},$$

puesto que $\mathscr{H} \perp \sigma(X)$ y entonces $\int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P}$ para $A \in \mathscr{C}$.

$\sigma(\mathscr{C}) = \sigma(\mathscr{G}, \mathscr{H})$**:** Esto es inmediato pues $\mathscr{G} \cup \mathscr{H} \subset \mathscr{C} \subset \sigma(\mathscr{G}, \mathscr{H})$.

El teorema de Clases de Sierpinski asegura que $\sigma(\mathscr{G}, \mathscr{H}) \subset \mathscr{I}$, por lo que hemos demostrado el resultado deseado.

PROPIEDAD 14 (Desigualdad de Jensen para la Esperanza Condicional). *Si $\varphi : \mathbb{R} \to \mathbb{R}$ es una función convexa, la integral de X existe y la integral de $\varphi \circ X$ está definida, entonces*

$$\varphi\left(\mathbb{E}(X \mid \mathscr{G})\right) \leq \mathbb{E}(\varphi \circ X \mid \mathscr{G}) \quad \mathbb{P}|_{\mathscr{G}} - p.s.$$

Si $\psi : \mathbb{R} \to \mathbb{R}$ es una función cóncava, X es integrable y la integral de $\psi \circ X$ está definida, entonces

$$\psi\left(\mathbb{E}(X \mid \mathscr{G})\right) \geq \mathbb{E}(\psi \circ X \mid \mathscr{G}) \, \mathbb{P}|_{\mathscr{G}} - p.s..$$

2. Martingalas

Estudiaremos ahora una familia de procesos estocásticos que es importante dentro de la teoría de la probabilidad, principalmente por sus aplicaciones teóricas, tanto así, que su estudio resulta imprescindible para la teoría moderna de la probabilidad. Mediante su uso, verificaremos ciertos teoremas clásicos para caminatas aleatorias, como la ley $0-1$ de Kolmogorov y la ley fuerte de los grandes números.

En este capítulo solamente consideraremos procesos estocásticos indicados por un subconjunto de $\mathbb{Z}$.

Consideremos la siguiente situación: jugamos una serie de volados, obteniendo 1 si ganamos el n-ésimo y -1 si lo perdemos. El modelo matemático que consideraremos está conformado por una sucesión de variables aleatorias independientes e idénticamente distribuidas $(X_i)_{i=1}^{\infty}$, donde X_i representa el resultado del i-ésimo volado. Nuestra fortuna al tiempo n, S_n, está dada por

$$S_n = \sum_{i=1}^{n} X_i$$

para $n \geq 1$ y definiremos $S_0 = 0$. Para que el juego resulte justo para las dos personas que lo juegan, debemos pedir que $\mathbb{P}(X_i = 1) = 1/2$. Si este es el caso, podemos preguntarnos por la mejor aproximación a S_{n+1} que podemos dar al utilizar la información sobre el juego que conocemos hasta el tiempo n. La información al tiempo n la interpretaremos como

$$\mathscr{F}_n = \sigma(X_1, \ldots, X_n),$$

puesto que esta σ-álgebra contiene a todos los conjuntos de la forma

$$\{X_1 = i_1, \ldots, X_n = i_n\},$$

y así, lo que realmente buscamos es la esperanza condicional de S_{n+1} dada $\mathscr{F}_n$, que es sencilla de calcular, pues

$$\mathbb{E}(S_{n+1} \mid \mathscr{F}_n) = \mathbb{E}(S_n + X_{n+1} \mid \mathscr{F}_n) = S_n + \mathbb{E}(X_{n+1}) = S_n.$$

Como un ejercicio, el lector puede verificar que de hecho,

$$\mathbb{E}(S_{n+m} \mid \mathscr{F}_n) = S_n, \quad \forall m \geq 0,$$

por lo que al conocer la información hasta el tiempo n, solamente podemos afirmar que nos quedaremos con lo que tenemos, y como lo mismo sucede con el jugador contra el cual competimos, el juego resulta ser justo.

Informalmente, podemos definir una martingala como un proceso estocástico $(X_n)_{n\in\mathbb{N}}$ tal que X_n representa la ganancia al tiempo n de un jugador involucrado en un juego justo respecto a cierta información. Para precisar esta idea, necesitamos un ingrediente extra:

Definición. Sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad y $(\mathscr{F}_n)_{n\in\mathbb{N}}$ una colección de σ-álgebras contenidas cada una en $\mathscr{F}$. Decimos que dicha familia es una **filtración** si $\mathscr{F}_n \subset \mathscr{F}_m$ cuando $n \leq m$.

Si $(\mathscr{F}_n)_{n\in\mathbb{N}}$ es una filtración, interpretaremos a $\mathscr{F}_n$ como la información acumulada al tiempo n.

Definición. Sean $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad y $(\mathscr{F}_n)_{n\in\mathbb{N}}$ una filtración en dicho espacio. Una colección de variables aleatorias reales $(X_n)_{n\in\mathbb{N}}$ es una **martingala** respecto a la filtración considerada si

(1) X_n es $\mathscr{F}_n$-medible.
(2) $X_n \in L_1$.
(3) $\mathbb{E}(X_{n+1} \mid \mathscr{F}_n) = X_n$ para cualquier $n \in \mathbb{N}$.

La primera propiedad nos dice que conocemos la ganancia al tiempo n a partir de la información que se nos proporciona hasta ese instante, generalmente se dice que la sucesión de variables aleatorias es adaptada a la filtración. La segunda es una hipótesis técnica que nos permite utilizar a la esperanza condicional como un operador lineal y la tercera nos dice que el juego es justo respecto a la información proporcionada.

Definición. Supongamos ahora que $(X_n)_{n\in\mathbb{N}}$ satisface (1) y (2), pero en vez de tener una igualdad en (3), observamos una desigualdad:

$$\mathbb{E}(X_{n+1} \mid \mathscr{F}_n) \leq X_n.$$

Entonces le llamaremos a la sucesión una **supermartingala**. (Note que de acuerdo a nuestra interpretación de X_n como evolución de nuestra fortuna, una supermartingala no tiene nada de super...) Si se da la desigualdad contraria, esto es,

$$\mathbb{E}(X_{n+1} \mid \mathscr{F}_n) \geq X_n,$$

entonces a $(X_n)_{n\in\mathbb{N}}$ le llamamos **submartingala**.

Notemos que si $(X_n)_{n\in\mathbb{N}}$ es una martingala, entonces la sucesión $(\mathbb{E}(X_n))_{n\in\mathbb{N}}$ es constante. Para una supermartingala o una submartingala, la palabra constante se debe substituir por decreciente o por creciente. Además, podemos inferir una propiedad más fuerte a partir de (3), a saber, que

$$\mathbb{E}(X_{n+m} \mid \mathscr{F}_n) = X_n$$

si $n, m \in \mathbb{N}$. Esto se sigue de las propiedades de la esperanza condicional, puesto que

$$\mathbb{E}(X_{n+m+1} \mid \mathscr{F}_n) = \mathbb{E}(\mathbb{E}(X_{n+m+1} \mid \mathscr{F}_{n+m}) \mid \mathscr{F}_n) = \mathbb{E}(X_{n+m} \mid \mathscr{F}_n).$$

Una afirmación similar es válida para supermartingalas o submartingalas al cambiar la igualdad por una desigualdad. Para concluir esta sección, veamos un método para construir submartingalas a partir de una martingala dada.

Teorema 1.2. *Si $(X_n)_{n\in\mathbb{N}}$ es una martingala respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$ y $\varphi : \mathbb{R} \to \mathbb{R}$ es una función convexa tal que $\varphi(X_n) \in L_1$, entonces $(\varphi(X_n))_{n\in\mathbb{N}}$ es una submartingala respecto a la misma filtración.*

Demostración. Como cualquier función convexa (sobre $\mathbb{R}$) es continua, entonces $\varphi(X_n)$ es $\mathscr{F}_n$-medible, que pertenece por hipótesis a L_1. Finalmente, por la desigualdad de Jensen para la esperanza condicional,

$$\mathbb{E}(\varphi(X_{n+1}) \mid \mathscr{F}_n) \geq \varphi(\mathbb{E}(X_{n+1} \mid \mathscr{F}_n)) = \varphi(X_n). \qquad \square$$

2.1. Ejemplos. En esta sección supondremos que $(\Omega, \mathscr{F}, \mathbb{P})$ es un espacio de probabilidad en el cual están definidas variables aleatorias con las características deseadas.

Ejemplo 1.1. Supongamos que X es una variable aleatoria que pertenece a L_1 y $(\mathscr{F}_n)_{n\in\mathbb{N}}$ una filtración en $(\Omega, \mathscr{F}, \mathbb{P})$. Entonces la sucesión de variables aleatorias $(X_n)_{n\in\mathbb{N}}$ en la cual

$$X_n = \mathbb{E}(\, X \mid \mathscr{F}_n)$$

(sin importar la versión de la esperanza condicional) es una martingala respecto a $(\mathscr{F}_n)_{n\in\mathbb{N}}$.

Para verificar la veracidad de la anterior afirmación, notemos que por definión de esperanza condicional, X_n es $\mathscr{F}_n$-medible y que $X_n \in L_1$. Finalmente, como $\mathscr{F}_n \subset \mathscr{F}_{n+1}$, entonces

$$\mathbb{E}(\, X_{n+1} \mid \mathscr{F}_n) = \mathbb{E}(\mathbb{E}(\, X \mid \mathscr{F}_{n+1}) \mid \mathscr{F}_n) = \mathbb{E}(\, X \mid \mathscr{F}_n) = X_n.$$

Esta martingala es un ejemplo bastante general y muy importante. Posteriormente podremos determinar cuando una martingala es de este tipo. A este proceso se le conoce como la **martingala cerrada**.

Ejemplo 1.2. Sean $(\xi_i)_{i=1}^{\infty}$ variables aleatorias independientes e idénticamente distribuidas, $\mathscr{F}_0 = \{\emptyset, \Omega\}$ y $\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ para $n \geq 1$. Entonces $(\mathscr{F}_n)_{n\in\mathbb{N}}$ es una filtración.. Si

$$S_0 = 0,$$

$$S_n = \sum_{i=1}^{n} \xi_i \quad n \geq 1,$$

y ξ_i tiene media finita μ, entonces $(X_n)_{n\in\mathbb{N}}$ es una martingala respecto a $(\mathscr{F}_n)_{n\in\mathbb{N}}$, donde

$$X_n = S_n - n\mu.$$

Si además, $\sigma^2 = \mathrm{Var}(\xi_i) < \infty$, entonces $(Y_n)_{n\in\mathbb{N}}$ es martingala respecto a la misma filtración, donde

$$Y_n = (S_n - n\mu)^2 - n\sigma^2.$$

Por otro lado, si

$$\varphi(\theta) = \mathbb{E}\big(e^{\theta\xi_i}\big) < \infty$$

para $\theta \in \mathbb{R}$, definimos $Z_0 = 1$ y para $n \geq 1$:

$$Z_n = \frac{e^{\theta S_n}}{(\varphi(\theta))^n},$$

entonces $(Z_n)_{n\in\mathbb{N}}$ es una martingala respecto a la misma filtración.

Como X_n, Y_n y Z_n están dadas por $f(\xi_1, \ldots, \xi_n)$, para una función continua $f : \mathbb{R}^n \to \mathbb{R}$ (una función distinta para cada variable aleatoria) y el vector aleatorio $(\xi_1, \ldots, \xi_n)$ es medible respecto a $\mathscr{F}_n$, se sigue que las tres variables consideradas son $\mathscr{F}_n$-medibles. Para ver que pertenecen a L_1, notemos que X_n, es la diferencia de dos funciones en L_1, por ser este último cerrado bajo la suma. Además, si ξ_i tiene momento de segundo orden finito, entonces a S_n le pasa lo mismo, por lo que $Y_n \in L_1$. Para Z_n, el argumento que utilizamos es el de independencia, puesto que esto implica que

$$\mathbb{E}(\exp(\theta S_n)) = (\varphi(\theta))^n < \infty.$$

Para verificar la última propiedad que define a las martingalas, notemos que

$$\mathbb{E}(X_{n+1} - X_n \mid \mathscr{F}_n) = \mathbb{E}(\xi_{n+1} - \mu \mid \mathscr{F}_n) = \mathbb{E}(\xi_{n+1} - \mu) = 0,$$

por lo que $(X_n)_{n\in\mathbb{N}}$ es efectivamente una martingala. Por otro lado,

$$\begin{aligned}
\mathbb{E}(Y_{n+1} - Y_n \mid \mathscr{F}_n) =&\mathbb{E}\Big((S_{n+1} - (n+1)\,\mu)^2 - (S_n - n\mu)^2 - \sigma^2 \Big| \mathscr{F}_n\Big) \\
=&\mathbb{E}(2\,(S_n - n\mu)\,(\xi_{n+1} - \mu) \mid \mathscr{F}_n) \\
&+ \mathbb{E}\Big((\xi_{n+1} - \mu)^2 - \sigma^2 \Big| \mathscr{F}_n\Big) \\
=&2\,(S_n - n\mu)\,\mathbb{E}(\xi_{n+1} - \mu) + \mathbb{E}((\xi_{n+1} - \mu)) \\
=&0,
\end{aligned}$$

por lo que $(Y_n)_{n\in\mathbb{N}}$ es una martingala. Finalmente se tiene que

$$\begin{aligned}
\mathbb{E}(Z_{n+1} \mid \mathscr{F}_n) &= \mathbb{E}\left(Z_n \frac{e^{\theta\xi_{n+1}}}{\varphi(\theta)} \middle| \mathscr{F}_n\right) \\
&= Z_n \mathbb{E}\left(\frac{e^{\theta\xi_{n+1}}}{\varphi(\theta)} \right) \\
&= Z_n,
\end{aligned}$$

por lo que $(Z_n)_{n\in\mathbb{N}}$ es una martingala.

Más adelante, utilizaremos estas martingalas para hacer ciertos cálculos referentes a caminatas aleatorias.

Ejemplo 1.3. Sea U una variable aleatoria uniforme en $(0, 1)$ y definamos a

$$X_n = 2^n \mathbf{1}_{U\leq 1/2^n}.$$

Entonces $X_0, X_1, \ldots$ es una martingala respecto de la filtración que genera.

Ejercicio 1.4. Probar la afirmación anterior.

Notemos que en esta martingala, se tiene que $X_n \to 0$ $\mathbb{P}$-p.s., pero que sin embargo, $(X_n)_{n\in\mathbb{N}}$ no converge en L_1 a 0.

Ejemplo 1.4. Consideremos el siguiente experimento aleatorio, se tiene una urna con r bolas rojas y v bolas verdes. Extraemos una bola, la reemplazamos junto con c bolas del mismo color, revolvemos la urna y volvemos a realizar el experimento. Sea X_0 la fracción inicial de bolas rojas en la urna y X_n la fracción de bolas rojas en la urna una vez realizado el experimento n veces. Entonces

$(X_n)_{n\in\mathbb{N}}$ es una martingala con respecto a la filtración que genera esta sucesión. Antes de proceder a verificar la afirmación anterior, debemos considerar el modelo matemático preciso del experimento aleatorio en cuestión, para poder calcular las esperanzas condicionales. Notemos que al momento de la n-ésima extracción hay

$$b_n = r + v + nc$$

bolas en la urna. Sean (U_i) variables aleatorias independientes e idénticamente distribuidas, $r, v > 0$ y definamos $X_0 = r/(r+v)$ y para $n \geq 0$:

$$Y_{n+1} = \mathbf{1}_{U_{n+1}\leq X_n} \quad \text{y} \quad X_{n+1} = \frac{r+v+nc}{r+v+(n+1)\,c} X_n + \frac{c}{r+v+(n+1)\,c} Y_n.$$

Esta es la descripción matemática que utilizaremos del experimento considerado anteriormente y en él, la variable X_n es función de $X_0, U_1, \ldots, U_n$ para $n \geq 1$ (de hecho es función de X_{n-1} y U_n) y por lo tanto, U_{n+1} es independiente de $\mathscr{F}_n$, la σ-álgebra generada por $X_0, \ldots, X_n$.

Ejercicio 1.5. Verificar que la sucesión X es una martingala respecto de $(\mathscr{F}_n)$.

Ejercicio 1.6 (Descomposición de Doob para submartingalas). Sea $X = (X_n)_{n\in\mathbb{N}}$ una submartingala. Pruebe que X se puede descomponer de manera única como $X = M + A$ donde M es una martingala y A es un proceso previsible con $A_0 = 0$. (Decimos que A es previsible si A_n es $\mathscr{F}_{n-1}$-medible para toda $n \geq 1$ y A_0 es $\mathscr{F}_0$-medible. Sugerencia: Asuma que ya tiene la descomposición y calcule esperanza condicional de X_{n+1} dada X_n.

Ejercicio 1.7. Sean X y Y dos martingalas (respecto de la misma filtración) y tales que $\mathbb{E}(X_i), \mathbb{E}(Y_i) < \infty$ para toda i. Pruebe la siguiente fórmula de integración por partes:

$$\mathbb{E}(X_n Y_n) - \mathbb{E}(X_0 Y_0) = \sum_{i=1}^{n} \mathbb{E}((X_i - X_{i-1})(Y_i - Y_{i-1})).$$

2.2. El teorema de muestreo opcional de Doob. Entre las razones por las cuales las martingalas son importantes, se encuentran los teoremas de convergencia de martingalas, que bajo ciertas condiciones de acotamiento nos permiten concluir la convergencia casi segura (o de otro tipo) de una martingala. Para abordar este resultado, es importante extender la igualdad $\mathbb{E}(X_n) = \mathbb{E}(X_0)$ para abarcar no sólo a tiempos deterministas como n, sino también a ciertos tiempos aleatorios, concepto que procedemos a discutir. Consideremos $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad, $(\mathscr{F}_n)_{n\in\mathbb{N}}$ una filtración y $(X_n)_{n\in\mathbb{N}}$ una martingala respecto a la anterior filtración. Nuestro objetivo es observar a la martingala a un tiempo que a su vez es una variable aleatoria. Esto se logra como sigue: si $T : \Omega \to \mathbb{N}$ es una variable aleatoria y definimos a $X_T : \Omega \to \mathbb{R}$ por medio de

$$X_T(\omega) = X_{T(\omega)}(\omega),$$

entonces X_T resulta ser una variable aleatoria, puesto que si $B \in \mathscr{B}_{\mathbb{R}}$, entonces

$$X_T^{-1}(B) = \cup_{n\in\mathbb{N}} \{\omega \in \Omega : T(\omega) = n, X_n(\omega) \in B\}.$$

Mediante la anterior variable aleatoria, observamos a la martingala al tiempo aleatorio T. En realidad, trabajaremos con una clase más reducida de tiempos aleatorios, a saber, los tiempos de paro. Para explicarlos, pensemos que al instante n debemos decidir si parar a la martingala (definiendo a n como el valor de T) de acuerdo a la información que tenemos disponible (es decir $\mathscr{F}_n$). Esto motiva la siguiente

Definición. Sea $T : \Omega \to \mathbb{N} \cup \{\infty\}$ una variable aleatoria. Decimos que T es un tiempo de paro respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$ si

$$\{\omega \in \Omega : T(\omega) = n\} \in \mathscr{F}_n \quad \forall n \in \mathbb{N}.$$

El lector puede verificar que T es un tiempo de paro respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$ si y sólo si $\{T \leq n\} \in \mathscr{F}_n$.

Teorema 1.3. *Sea X una submartingala. Si T es tiempo de paro respecto a $(\mathscr{F}_n)_{n\in\mathbb{N}}$ y T está acotado por N entonces*

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_N).$$

Si X es una martingala entonces

$$\mathbb{E}(X_T) = \mathbb{E}(X_N).$$

Demostración. Por hipótesis, existe un natural $N > 0$ tal que $T \leq N$. Así,

$$\mathbb{E}(X_T) = \sum_{n=0}^{N} \mathbb{E}\big(X_n \mathbf{1}_{(T=n)}\big),$$

pero como el conjunto $\{T = n\}$ pertenece a $\mathscr{F}_n$

$$\mathbb{E}\big(X_n \mathbf{1}_{(T=n)}\big) \leq \mathbb{E}\big(\mathbb{E}(X_N \mid \mathscr{F}_n)\, \mathbf{1}_{(T=n)}\big) = \mathbb{E}\big(X_N \mathbf{1}_{(T=n)}\big),$$

por lo que

$$\mathbb{E}(X_T) \leq \sum_{n=0}^{N} \mathbb{E}\big(X_N \mathbf{1}_{(T=n)}\big) = \mathbb{E}(X_N).$$

Si X es una martingala, la desigualdad que utilizamos es una igualdad. □

El teorema anterior vale para tiempos de paro acotados y posteriormente, al hacer un análisis más a fondo de las martingalas y de los tiempos de paro podremos extender el teorema anterior a una familia más amplia de tiempos de paro.

Ejercicio 1.8. Sea X una supermartingala. Pruebe que si T es un tiempo de paro acotado por N entonces

$$\mathbb{E}(X_T) \geq \mathbb{E}(X_N).$$

Curiosamente, la técnica de demostración anterior no funciona para probar que si X es una supermartingala y T es un tiempo de paro acotado por N entonces $\mathbb{E}(X_0) \geq \mathbb{E}(X_T)$. Utilizaremos una segunda representación de la variable X_T para este fin. En efecto, notemos que

$$X_T - X_0 = \sum_{n=1}^{N} \mathbf{1}_{T>n-1} (X_n - X_{n-1}).$$

Puesto que $\{T > n-1\} \in \mathscr{F}_{n-1}$ y X es supermartingala

$$\mathbb{E}(\mathbf{1}_{T>n-1} (X_n - X_{n-1})) = \mathbb{E}(\mathbf{1}_{T>n-1} [\mathbb{E}(X_n \mid \mathscr{F}_{n-1}) - X_{n-1}]) \leq 0.$$

Por lo tanto, vemos que

$$\mathbb{E}(X_T) \leq \mathbb{E}(X_0).$$

La idea anterior es parte de un resultado mucho más interesante que permite obtener nuevas martingalas a partir de otras.

Definición. Sea $C = (C_n, n \geq 1)$ un proceso estocástico. Decimos que C es predecible respecto de $(\mathscr{F}_n)$ si C_n es $\mathscr{F}_{n-1}$-medible.

Si C es un proceso predecible y acotado y M es una martingala, formemos al nuevo proceso $C \cdot M$ como sigue:

$$(C \cdot M)_0 = 0 \quad \text{y} \quad (C \cdot M)_n = \sum_{i \leq n} C_i (M_i - M_{i-1}).$$

Ejercicio 1.9. Mostrar cuidadosamente que $C \cdot M$ es una martingala. Obtenga un enunciado análogo si M es una submartingala.

Ejercicio 1.10 (Extensiones del teorema de paro opcional). Sea $M = (M_n, n \in \mathbb{N})$ una (super)martingala respecto de una filtración $(\mathscr{F}_n, n \in \mathbb{N})$ y sean S y T tiempos de paro.

(1) Pruebe que $S \wedge T$, $S + T$ y $S \vee T$ son tiempos de paro.
(2) Sea

$$\mathscr{F}_T = \{A \in \mathscr{F} : A \cap \{T \leq n\} \in \mathscr{F}_n \text{ para toda } n\}$$

es una σ-álgebra, a la que nos referimos como la σ-álgebra detenida en τ. Comente qué puede fallar si T no es tiempo de paro. Pruebe que T es $\mathscr{F}_T$-medible.
(3) Pruebe que si T es finito, entonces M_T es $\mathscr{F}_T$-medible.
(4) Pruebe que si $S \leq T \leq n$ entonces $\mathscr{F}_S \subset \mathscr{F}_T$. Si además T es acotado entonces $X_S, X_T \in L_1$ y

$$\mathbb{E}(M_T \mid \mathscr{F}_S) \leq M_S.$$

(5) Si $X = (X_n, n \in \mathbb{N})$ es un proceso estocástico $(\mathscr{F}_n)$-adaptado y tal que $X_n \in L_1$ y tal que para cualesquiera tiempos de paro acotados S y T se tiene que $\mathbb{E}(X_S) = \mathbb{E}(X_T)$ entonces X es una martingala. Sugerencia: considere tiempos de paro de la forma $n\mathbf{1}_A + (n+1)\mathbf{1}_{A^c}$ con $A \in \mathscr{F}_n$.

2.3. El teorema de muestreo opcional de Doob y el problema de la ruina. En esta sección aplicaremos el teorema de muestreo opcional para tiempos de paro acotados para resolver algunas preguntas concernientes a un problema clásico dentro de la probabilidad, el problema de la ruina. Utilizaremos las martingalas del ejemplo (1.2). Supongamos que dos personas, A y B, juegan a los volados, donde A gana el n-ésimo volado con probabilidad $p \in (0,1)$. Si A cuenta con una fortuna inicial de a pesos, B una de b pesos y apuestan en una serie de volados, un peso cada volado, hasta que uno de los dos quede sin dinero, ¿Cuál es la probabilidad de que A se quede con la fortuna de B? y ¿Cuál es la duración esperada del juego?

Para responder a dichas preguntas, primero las formularemos en términos de la caminata aleatoria simple de la siguiente manera: Sean $(X_i)_{i=1}^{\infty}$ variables aleatorias independientes que toman el valor 1 con probabilidad p y -1 con probabilidad $1-p$. Así, X_i toma el valor 1 si A le gana un peso a B y el valor -1 si pierde un peso en el i-ésimo volado. Sean

$$S_0 = 1 \quad \text{y} \quad S_n = X_1 + \cdots + X_n \text{ para } n \geq 1.$$

Así, $a + S_n$ representa a la fortuna de A después de n volados, y por lo tanto, si

$$T_x = \inf\{n \geq 1 : S_n = x\},$$

A le gana su fortuna a B si $T_b < T_{-a}$. Para responder la primera pregunta, debemos calcular

$$\mathbb{P}(T_b < T_{-a}).$$

La cantidad de volados que juegan hasta que alguien se quede sin dinero es $T_b \wedge T_{-a}$, por lo que para responder a la segunda pregunta, debemos calcular

$$\mathbb{E}(T_b \wedge T_{-a}).$$

El análisis es distinto si se trata de un juego justo ($p = 1/2$) o no. Haremos el caso $p = 1/2$ y el caso $p \neq 1/2$ se dejará indicado como ejercicio.

Necesitamos un resultado preliminar.

Proposición 1.3. *Para cualquier $a, b > 0$, $\mathbb{P}(T_b \wedge T_{-a} < \infty) = 1$.*

Demostración. Sea K un entero mayor a $a+b$. Notemos que $\mathbb{P}(|S_K| \geq a \vee b) > 0$ y que, como los eventos

$$|S_K| \geq a \vee b, |S_{2K} - S_K| \geq a \vee b, \ldots$$

son independientes y tienen la misma probabilidad, el lema de Borel-Cantelli nos dice que $\left|S_{nK} - S_{(n-1)k}\right| \geq a \vee b$ para una infinidad de índices n casi seguramente. Por otra parte, vemos que si $\left|S_{nK} - S_{(n-1)k}\right| \geq a \vee b$ entonces $T_a \wedge T_b \leq nK$. □

Caso $p = 1/2$: Como $(S_n)_{n\in\mathbb{N}}$ es martingala respecto a la filtración que genera, tiene media cero y $T_{-a} \wedge T_b \wedge n$ es un tiempo de paro acotado, se tiene que

$$0 = \mathbb{E}\big(S_{T_{-a}\wedge T_b\wedge n}\big)\,.$$

Además, $\big(S_{T_{-a}\wedge T_b\wedge n}\big)_{n\geq 1}$ converge a $S_{T_{-a}\wedge T_b}$ y los elementos de dicha sucesión están acotados por $a\vee b$,. Por el teorema de convergencia acotada,

$$0 = \mathbb{E}\big(S_{T_{-a}\wedge T_b}\big) = -a\mathbb{P}(T_{-a} < T_b) + b\mathbb{P}(T_b < T_{-a})\,,$$

de donde

$$\mathbb{P}(T_b < T_{-a}) = \frac{a}{a+b}.$$

Para responder a la segunda pregunta en este caso, notemos que $\big(S_n^2 - n\big)_{n\in\mathbb{N}}$ es martingala respecto a la filtración que genera, ya que $\mathbb{E}(X_i) = 0$ y $\mathrm{Var}(X_i) = 1$. Al utilizar el tiempo de paro acotado $T_{-a} \wedge T_b \wedge n$, vemos que

$$\mathbb{E}\Big(S^2_{T_{-a}\wedge T_b\wedge n}\Big) = \mathbb{E}(T_{-a} \wedge T_b \wedge n)\,.$$

Como $\Big(S^2_{T_{-a}\wedge T_b\wedge n}\Big)_{n\geq 1}$ es una sucesión acotada por $a^2 \vee b^2$ y converge a $S^2_{T_{-a}\wedge T_b}$, podemos aplicar el teorema de convergencia acotada para concluir que

$$\mathbb{E}\Big(S^2_{T_{-a}\wedge T_b}\Big) = \lim_{n\to\infty} \mathbb{E}\Big(S^2_{T_{-a}\wedge T_b\wedge n}\Big) = \lim_{n\to\infty} \mathbb{E}(T_{-a} \wedge T_b \wedge n)\,.$$

Como $(T_{-a} \wedge T_b \wedge n)_{n\in\mathbb{N}}$ es una sucesión creciente de variables aleatorias no negativas, que converge a $T_{-a} \wedge T_b$, el teorema de convergencia monótona implica que

$$\mathbb{E}(T_{-a} \wedge T_b) = \lim_{n\to\infty} \mathbb{E}(T_{-a} \wedge T_b \wedge n) = \lim_{n\to\infty} \mathbb{E}\Big(S^2_{T_{-a}\wedge T_b\wedge n}\Big) = \mathbb{E}\Big(S^2_{T_{-a}\wedge T_b}\Big)\,.$$

Finalmente, al utilizar el valor de $\mathbb{P}(T_b < T_{-a})$, vemos que

$$\mathbb{E}(T_{-a} \wedge T_b) = \mathbb{E}\Big(S^2_{T_{-a}\wedge T_b}\Big) = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab,$$

por lo que la cantidad esperada de volados hasta la ruina de alguno de los dos jugadores es ab.

Ejercicio 1.11. Suponga que $p > 1 - p$.

(1) Sea $\phi(x) = (p/q)^x$ y pruebe que $(\phi(S_n))_{n\in\mathbb{N}}$ es martingala respecto a la filtración que genera.
(2) Note que al aplicar el teorema de muestreo opcional de Doob al tiempo de paro acotado $T_{-a} \wedge T_b \wedge n$ se obtiene

$$1 = \mathbb{E}\big(\phi\big(S_{T_{-a}\wedge T_b\wedge n}\big)\big)\,.$$

Utilice alguna propiedad de la esperanza para pasar al límite conforme $n \to \infty$ y concluir que

$$1 = \mathbb{E}\big(\phi\big(S_{T_{-a}\wedge T_b}\big)\big) = \phi(-a)\,\mathbb{P}(T_{-a} < T_b) + \phi(b)\,\mathbb{P}(T_b < T_{-a})\,.$$

Concluya con el cálculo explícito de $\mathbb{P}(T_b < T_{-a})$.

(3) Pruebe que $(S_n - n\,(2p-1))_{n\in\mathbb{N}}$ es una martingala.

(4) Note que al aplicar muestreo opcional al tiempo de paro $T_{-a} \wedge T_b \wedge n$ se obtiene

$$\mathbb{E}\big(S_{T_{-a}\wedge T_b\wedge n}\big) = (2p-1)\,\mathbb{E}(T_{-a} \wedge T_b \wedge n)\,.$$

Aplique propiedades de la esperenza al lado derecho y de la probabilidad al lado derecho que permitan pasar al límite conforme $n \to \infty$ en la expresión anterior y obtener:

$$\begin{aligned}\mathbb{E}(T_{-a} \wedge T_b) &= \frac{1}{2p-1}\mathbb{E}\big(S_{T_{-a}\wedge T_b}\big)\\ &= \frac{1}{2p-1}\left(-a\mathbb{P}(T_{-a} < T_b) + b\mathbb{P}(T_b < T_{-a})\right)\end{aligned}$$

y calcule explícitamente $\mathbb{E}(T_{-a} \wedge T_b)$.

2.4. El teorema de convergencia casi segura. Para proceder a estudiar la convergencia casi segura de las martingalas, necesitamos una caracterización de la convergencia de sucesiones. Para esto, consideremos una sucesión real $\{x_n\}_{n=0}^{\infty}$. Para verificar si esta sucesión es convergente en $\mathbb{R}$, es necesario y suficiente probar que $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$ y que esta cantidad pertenece a $\mathbb{R}$. A continuación veremos una manera de concluir que el límite superior y el límite inferior de la sucesión coinciden: si $a < b$ son dos racionales, veamos cuantas veces cruzan hacia arriba los puntos $x_0, x_1, \ldots$ a $[a,b]$, cantidad que denotamos por $U_{[a,b]}(x_0, x_1, \ldots)$ y cuyo cálculo procedemos a explicar: sean

$$A_1 = \{k \in \mathbb{N} : x_k \leq a\}\,,$$

$$T_1(x_0, x_1, \ldots) = \begin{cases} \min A_1 & \text{si } A_1 \neq \emptyset \\ \infty & \text{si } A_1 = \emptyset \end{cases}$$

y de manera recursiva, para $j \geq 1$

$$A_{2j} = \{k \in \mathbb{N} : T_{2j-1} \leq k, x_k \geq b\}\,,$$

$$T_{2j}(x_0, x_1, \ldots) = \begin{cases} \min A_{2j} & \text{si } A_{2j} \neq \emptyset \\ \infty & \text{si } A_{2j} = \emptyset \end{cases}$$

$$A_{2j+1} = \{k \in \mathbb{N} : T_{2j} \leq k, x_k \leq a\}$$

$$T_{2j+1}(x_0, x_1, \ldots) = \begin{cases} \min A_{2j+1} & \text{si } A_{2j+1} \neq \emptyset \\ \infty & \text{si } A_{2j+1} = \emptyset \end{cases}.$$

A partir de las anteriores cantidades, definimos

$$\begin{aligned} U_{[a,b]}(x_0, \ldots, x_n) &= \sup \{k \in \mathbb{N} : k \geq 1, A_{2k} \neq \emptyset\} \\ &= \sup \{k \in \mathbb{N} : k \geq 1, T_{2k} < \infty\}, \end{aligned}$$

que es la cantidad de cruces hacia arriba de la sucesión en $[a, b]$ pues si que $T_{2k} < \infty$ entonces la sucesión ha cruzado $[a, b]$ hacia arriba de menos k veces, por definición de T_{2k}.

Lema 1. *El límite inferior de la sucesión $(x_n)_{n\in\mathbb{N}}$ coincide con el límite superior de la misma si y sólo si para cualquier pareja de racionales $a < b$, la cantidad $U_{[a,b]}(x_0, x_1, \ldots)$ es finita.*

Demostración. Si el límite inferior de la sucesión, l, no coincide con el límite superior de la misma, L, entonces existen dos racionales $a < b$ tales que $l < a < b < L$. Por definición de límite superior, para cada $n \in \mathbb{N}$ existe $m_n \in \mathbb{N}$ mayor que n tal que $x_{m_n} > b$ y similarmente, existe $m'_n \in \mathbb{N}$ mayor que n tal que $x_{m'_n} < a$. De lo anterior podemos concluir que $T_k < \infty$ para cada $k \in \mathbb{N}$, puesto que $T_1 \leq m'_1$, de lo cual $T_2 \leq m_{T_1}$ y si $T_{2k} < \infty$, entonces $T_{2k+1} < m'_{T_{2k}}$ y $T_{2k+2} < m_{T_{2k+1}}$. Así, como la sucesión $(T_k)_{k\geq 1}$ es estrictamente creciente, pues $a < b$, se sigue que el conjunto cuyo supremo es $U_{[a,b]}(x_0, x_1, \ldots)$ es no acotado y por lo tanto esta última cantidad es igual a ∞.

Por otro lado, si $U_{[a,b]}(x_0, x_1, \ldots) = \infty$ para alguna pareja de racionales $a < b$, entonces los conjuntos

$$\{n \in \mathbb{N} : x_n \leq a\} \quad \text{y} \quad \{n \in \mathbb{N} : x_n \geq b\}$$

son infinitos, por lo que el límite superior de la sucesión es mayor o igual a b y el inferior, menor o igual a a, y por lo tanto el límite superior y el inferior difieren. □

Teorema 1.4. *Si $\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|) < \infty$, entonces $(X_n)_{n\in\mathbb{N}}$ converge casi seguramente a una variable aleatoria X_∞ que pertenece a L_1.*

Demostración. Por el lema anterior, notemos que

$$\left\{\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n\right\} = \bigcap_{\substack{a,b\in\mathbb{Q}\\ a<b}} \left\{U_{[a,b]}(X_0, X_1, \ldots) < \infty\right\},$$

por lo que para demostrar la afirmación del teorema, veremos primero que $U_{[a,b]}$[4] es una variable aleatoria finita casi seguramente. De esto se desprenderá que el conjunto del lado derecho de la anterior igualdad pertenece a $\mathscr{F}$ y tiene probabilidad 1, por lo que el límite superior y el inferior de la martingala coinciden casi seguramente y por el lema de Fatou, obtendremos que el valor absoluto del límite inferior (y por lo tanto el del límite de la sucesión) tiene esperanza finita, las conclusiones del teorema.

[4] Para la prueba de este teorema, las cantidades $U_{[a,b]}$ y T_k las evaluaremos en $X_0, X_1, \ldots$ sin indicarlo.

Lema 2. *para cada $k \in \mathbb{N}$ mayor o igual a 1, T_k es un tiempo de paro respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$.*

Demostración. La prueba se hará por inducción: Para $k = 1$, la afirmación se deduce de la igualdad,

$$\{T_1 \leq n\} = \cup_{i=0}^{n} \{X_i \leq a\},$$

válida para cualquier $n \in \mathbb{N}$. Si $i \leq n$, el conjunto $\{X_i \leq a\}$ pertenece a $\mathscr{F}_i$ y por lo tanto a $\mathscr{F}_n$, por lo que $\{T_1 \leq n\}$ pertenece a $\mathscr{F}_n$ y por lo tanto T_1 es tiempo de paro respecto a la filtración.

Por otro lado, si T_k es tiempo de paro y $k = 2l$ es par ($l \geq 1$), entonces para $n \geq 2$

$$\begin{aligned}\{T_{k+1} \leq n\} &= \{T_k < n\} \cap \{T_{k+1} \leq n\} \\ &= \bigcup_{i=1}^{n-1} \left(\{T_k = i\} \cap \bigcup_{j=i+1}^{n} \{X_j \leq a\} \right) \in \mathscr{F}_n.\end{aligned}$$

Como $T_{k+1} \geq 2$, entonces T_{k+1} es tiempo de paro. Por otro lado, si $k = 2l+1$ es tiempo de paro ($l \geq 0$), entonces para cada $n \geq 2$,

$$\begin{aligned}\{T_{k+1} \leq n\} &= \{T_k < n\} \cap \{T_{k+1} \leq n\} \\ &\bigcup_{i=1}^{n-1} \left(\{T_k = i\} \cap \bigcup_{j=i+1}^{n} \{X_j \geq b\} \right) \in \mathscr{F}_n.\end{aligned}$$

De nueva cuenta $T_{k+1} \geq 2$, por lo que T_{k+1} es tiempo de paro. □

En particular, el lema anterior nos permite afirmar que T_k es una variable aleatoria (posiblemente extendida, pues puede tomar el valor ∞.) Para continuar con la prueba del teorema, verifiquemos ahora que $U_{[a,b]}$ es una variable aleatoria, lo cual se desprende de manera inmediata del lema anterior pues

$$U_{[a,b]} = \sum_{k=1}^{\infty} \mathbf{1}_{(T_{2k}<\infty)}.$$

Ahora veamos que $U_{[a,b]}$ finita casi seguramente: para esto, indtroducimos a las variables aleatorias

$$U_{[a,b]}^{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \mathbf{1}_{(T_{2k}<\infty)},$$

la cantidad de cruces hacia arriba de $x_0, \ldots, x_n$ en $[a, b]$. Como $\left(U_{[a,b]}^{n}\right)_{n\geq 1}$ es una sucesión creciente de variables aleatorias no-negativas que converge a $U_{[a,b]}$, entonces se sigue que

$$\mathbb{E}\left(U_{[a,b]}^{n}\right) \to \mathbb{E}\left(U_{[a,b]}\right). \tag{1}$$

Si $S_n = T_n \wedge n$, entonces S_n es un tiempo de paro acotado puesto que

$$\{S_n \le k\} = \begin{cases} \{T_n \le k\} & \text{si } k \le n \\ \Omega & \text{si } k > n \end{cases},$$

por lo que la esperanza de la variable aleatoria

$$V_n = \sum_{k=1}^{n} X_{S_{2k}} - X_{S_{2k-1}}$$

es igual a cero. En la definición de la variable aleatoria V_n, puede haber muchos sumandos iguales a cero, puesto que para $k > \lfloor n/2 \rfloor$, $S_{2k} = S_{2k-1}$. Sabemos que (para cada $\omega \in \Omega$) existe $k \ge 1$ tal que $T_k \le n < T_{k+1}$. Si k es par, entonces $V_n \ge (b-a)\, U^n_{[a,b]}$, por lo que en este caso,

$$(b-a)\, U^n_{[a,b]} - V_n \le 0,$$

mientras que si $k = 2l+1$ es impar, entonces

$$V_n \ge (b-a)\, U^n_{[a,b]} + X_n - X_{T_{2l+1}} \ge (b-a)\, U^n_{[a,b]} + X_n - a,$$

por lo que en este caso

$$(b-a)\, U^n_{[a,b]} - V_n \le a - X_n,$$

de donde obtenemos una cota para cualquier $\omega \in \Omega$:

$$(b-a)\, U^n_{[a,b]} - V_n \le (a - X_n)^+$$

Como V_n tiene esperanza 0,

$$(b-a)\, \mathbb{E}\Big(U^n_{[a,b]}\Big) = \mathbb{E}\Big((b-a)\, U^n_{[a,b]} - V_n\Big) \le \mathbb{E}\Big((a-X_n)^+\Big),$$

por lo que

$$\mathbb{E}\Big(U^n_{[a,b]}\Big) \le \frac{1}{b-a}\mathbb{E}\Big((a-X_n)^+\Big)$$

Esta es la clásica desigualdad de Doob, que nos permitirá terminar con la prueba del teorema, puesto de acuerdo a (1),

$$\begin{aligned} \mathbb{E}\big(U_{[a,b]}\big) &= \lim_{n\to\infty} \mathbb{E}\Big(U^n_{[a,b]}\Big) \\ &\le \sup_{n\ge 1} \frac{1}{b-a}\mathbb{E}\Big((a-X_n)^+\Big) \\ &\le \frac{1}{b-a}\left(\sup_{n\ge 1}\mathbb{E}(|X_n|) + |a|\right) < \infty. \end{aligned}$$

Así, la variable $U_{[a,b]}$ es finita $\mathbb{P}$-p.s., pues pertenece a L_1.

Lo anterior muestra que existe una variable aleatoria X_∞ con valores extendidos (esto es, puede tomar los valores $-\infty$ e ∞). Sin embargo, por el lema de Fatou y la hipótesis $\sup_n \mathbb{E}(|X_n|) < \infty$, vemos que X_∞ es integrable (y por lo tanto es finita casi seguramente). $\square$

Por ejemplo, como la esperanza es constante para una martingala, entonces una martingala no-negativa satisface las condiciones del teorema anterior. En los ejemplos que hemos analizado, vemos que la martingala $(Z_n)_{n\in\mathbb{N}}$ del ejemplo (1.2) converge casi seguramente, así como las martingalas de los ejemplos (1.3) y (1.4). Veamos que la martingala del ejemplo (1.1) también converge $\mathbb{P}$-p.s.: Por la desigualdad de Jensen para la esperanza condicional,

$$\mathbb{E}(|X_n|) = \mathbb{E}(|\mathbb{E}(X \mid \mathscr{F}_n)|) \leq \mathbb{E}(\mathbb{E}(|X| \mid \mathscr{F}_n)) = \mathbb{E}(|X|),$$

de donde

$$\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|) \leq \mathbb{E}(|X|) < \infty.$$

Consideremos a la martingala X del ejemplo (1.4) es un caso particular de la del ejemplo (1.1). En efecto, (X_n) converge casi seguramente (digamos a X_∞) y es una sucesión acotada. Por lo tanto, podemos aplicar el teorema de convergencia acotada para probar que si $A \in \mathscr{F}$ entonces

$$\lim_{n\to\infty} \mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_\infty \mathbf{1}_A).$$

Por otra parte, si $A \in \mathscr{F}_m$ y $m \leq n$ entonces

$$\mathbb{E}(X_m \mathbf{1}_A) = \mathbb{E}(X_n \mathbf{1}_A)$$

puesto que X es una martingala. Así, también vemos que

$$\lim_{n\to\infty} \mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_m \mathbf{1}_A).$$

Se concluye que para todo $A \in \mathscr{F}_m$

$$\mathbb{E}(X_m \mathbf{1}_A) = \mathbb{E}(X_\infty \mathbf{1}_A)$$

y que por lo tanto

$$X_m = \mathbb{E}(X_\infty \mid \mathscr{F}_m).$$

2.5. Procesos de Galton-Watson y un criterio de extinción. Consideremos una población en la que los individuos se reproducen independientemente de los demás y en la que cada uno de ellos tiene k hijos con probabilidad μ_k, donde $\sum_{k=0}^\infty \mu_k = 1$. A la colección $\mu = (\mu_k, k \geq 0)$ le llamamos distribución de progenie. Nos interesará el tamaño de las generaciones sucesivas. Para modelar matemáticamente esta situación, consideremos a una colección $(\xi_{i,n})$ de variables aleatorias independientes con distribución μ. Sea $k \in \mathbb{N}$ y construyamos recursivamente a

$$Z_0 = k \quad \text{y} \quad Z_{n+1} = \sum_{1\leq i\leq Z_n} \xi_{i,n}.$$

Nuestra interpretación es que $\xi_{i,n}$ es la cantidad de hijos que tiene el i-ésimo individuo de la generación n, si es que existe, por lo que Z_n representa el tamaño de la generación n. A Z se le conoce como proceso de Galton-Watson de ley de reproducción (o progenie) μ. El proceso fue introducido para investigar la propensión a desaparecer que tenían los apellidos de la aristocracia en la Inglaterra victoriana y se llegó a la conclusión (erronea como veremos) de que la extinción siempre ocurre.

La extinción se puede traducir como el siguiente evento: $E = \{\exists n : Z_n = 0\}$ y ahora calcularemos la probabilidad de extinción en términos de la ley de reproducción μ. Consideremos primero a la filtración $\mathscr{F}_n = \sigma(\xi_{i,k} : i \geq 1, k < n)$ para $n \geq 1$ y $\mathscr{F}_0 = \{\Omega, \emptyset\}$. Notemos que Z es $(\mathscr{F}_n)$-adaptado. Ahora construiremos una martingala asociado al proceso de Galton-Watson. Para esto, calcularemos $\mathbb{E}(Z_{n+1} \mid \mathscr{F}_n)$ al notar que $\sigma(\xi_{i,n}, i \geq 1)$ es independiente de $\mathscr{F}_n$. Sea m la media de μ, que suponemos finita y positiva. Entonces

$$\begin{aligned}
\mathbb{E}(Z_{n+1} \mid \mathscr{F}_n) &= \mathbb{E}\left(\sum_{k\geq 1} \mathbf{1}_{Z_n=k} \left(\xi_{1,n} + \cdots + \xi_{k,n}\right) \,\middle|\, \mathscr{F}_n \right) \\
&= \sum_{k\geq 1} \mathbf{1}_{Z_n=k} \mathbb{E}(\xi_{1,n} + \cdots + \xi_{k,n} \mid \mathscr{F}_n) \\
&= \sum_{k\geq 1} \mathbf{1}_{Z_n=k} km \\
&= mZ_n.
\end{aligned}$$

Así, notamos que $(Z_n/m^n, n \geq 0)$ es una $(\mathscr{F}_n)$-martingala. A partir del teorema de convergencia casi segura, vemos que dicha martingala converge casi seguramente a una variable aleatoria M_∞. Esto nos quiere decir que sobre $M_\infty > 0$, Z_n tiene una velocidad de crecimiento (o decrecimiento si $m < 1$) exponencial, pues Z_n es asintótico a $W_\infty m^n$. Dividimos a los procesos de Galton-Watson en

Subcríticos: si $m < 1$
Críticos: si $m = 1$
Supercríticos: si $m > 1$

Teorema 1.5. *La probabilidad de extinción de procesos de Galton-Watson es igual a 1 si y sólo si el proceso es crítico o subcrítico.*

Demostración. Basta probar el teorema cuando $Z_0 = 1$ puesto que cuando $Z_0 = k$ la probabilidad de extinción será la potencia k de la probabilidad de extinción cuando $Z_0 = 1$. (Informalmente, esto sucede pues las k subpoblaciones que descienden de cada individuo de la generación k son independientes.) Supongamos pues que $Z_0 = 1$.

Cuando $m < 1$, la desigualdad de Markov implica que

$$\mathbb{P}(Z_n > 0) = \mathbb{P}(Z_n \geq 1) \leq \mathbb{E}(Z_n) = m^n \to 0.$$

Por lo tanto

$$\mathbb{P}(\exists n, Z_n = 0) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_n 1 - \mathbb{P}(Z_n > 0) = 1.$$

Cuando $m = 1$, notamos que Z es una martingala no-negativa y con valores enteros, por lo que casi seguramente es eventualmente constante. Ahora probemos que si $\mu_1 < 1$ (o en otras palabras, el proceso no es trivial) entonces para toda $k \geq 1$

$$\mathbb{P}(Z_n = k \text{ para toda } n \geq N) = 0.$$

En efecto, notemos primero que dicha hipótesis implica que $\operatorname{Var}(\xi_{1,1}) > 0$ lo cual implica que $\operatorname{Var}(\xi_{1,1} + \cdots + \xi_{1,k}) > 0$ por lo cual

$$\mathbb{P}(\xi_{1,1} + \cdots + \xi_{1,k} = k) < 1.$$

Así,

$$\begin{aligned} &\mathbb{P}(Z_N = k, Z_{N+1} = k, \ldots, Z_{N+n} = k) \\ &= \mathbb{P}(Z_N = k)\, \mathbb{P}(\xi_{1,1} + \cdots + \xi_{1,k} = k)^n \to_{n\to\infty} 0, \end{aligned}$$

por lo que para toda $k \geq 1$:

$$\mathbb{P}(Z_n = k \text{ para toda } n \geq N) = \lim_{n\to\infty} \mathbb{P}(Z_N = k, Z_{N+1} = k, \ldots, Z_{N+n} = k) = 0.$$

En este argumento no hemos utilizado que estemos en el caso crítico. La conclusión es que si Z_n converge (incluimos al infinito) entonces lo hace a 0 o a ∞. En el caso crítico, como Z_n es martingala no-negativa, converge en $\mathbb{N}$ y se concluye que $Z_n = 0$ para toda n suficientemente grande casi seguramente, por lo que la probabilidad de extinción es 1.

Finalmente, cuando $m > 1$, definamos a

$$f(s) = \sum_k \mu_k s^k.$$

Entonces $f(1) = 1$, el límite por la izquierda de f en 0, denotado $f(0+)$, es igual a μ_0,

$$f'(s) = \sum_{k\geq 1} k s^{k-1} \mu_k \geq 0 \quad \text{y} \quad f''(s) = \sum_{k\geq 2} k\,(k-1)\, s^{k-2} \mu_k \geq 0$$

por lo que f es una función convexa y creciente en $(0, 1)$. Notemos además que el límite por la izquierda de f' en 1, denotado $f'(1-)$ es entonces igual a $m > 1$, por lo que f cruza a la identidad en $[0, 1]$ exactamente en dos puntos: 1 y $\eta \in [0, 1)$. Calculemos

$$\begin{aligned} \mathbb{E}\big(\eta^{Z_{n+1}} \,\big|\, \mathscr{F}_n\big) &= \sum_{k\geq 0} \mathbf{1}_{Z_n = k} \mathbb{E}\big(\eta^{\xi_{1,n} + \cdots + \xi_{k,n}}\big) \\ &= \sum_{k\geq 0} \mathbf{1}_{Z_n = k} f(\eta)^k \\ &= \sum_{k\geq 0} \mathbf{1}_{Z_n = k} \eta^k \\ &= \eta^{Z_n}. \end{aligned}$$

Se concluye que $\left(\eta_n^Z, n \geq 0\right)$ es una martingala y puesto que es acotada, converge casi seguramente. Esto implica que $Z_n, n \geq 0$ converge casi seguramente en $\mathbb{N} \cup \{\infty\}$, por lo cual o converge a 0 o converge a ∞. Se sigue que $\eta^{Z_n} \to \mathbf{1}_{\text{Extinción}}$ casi seguramente y por el teorema de convergencia acotada

$$\mathbb{P}(\text{Extinción}) = \lim_{n\to\infty} \mathbb{E}\left(\eta^{Z_n}\right) = \mathbb{E}\left(\eta^{\zeta_0}\right) = \eta. \qquad \square$$

2.6. Desigualdades maximales de Doob. Ahora veremos un criterio sencillo para verificar que una martingala converge no sólo casi seguramente sino también en L_p para algún $p > 1$. Para esto, estudiaremos al máximo valor que toma una martingala. Con una cota adecuada para el máximo, se puede entonces simplemente aplicar convergencia dominada para verificar la convergencia en L_p. Sea $M = (M_n, n \geq 0)$ una $(\mathscr{F}_n)$-submartingala. Definamos a

$$\overline{M}_n^+ = \max_{1\leq i\leq n} M_n^+.$$

Proposición 1.4 (Desigualdad maximal de Doob). *Para toda $\lambda > 0$,*

$$\lambda \mathbb{P}\left(\overline{M}_n^+ > \lambda\right) \leq \mathbb{E}\left(M_n^+\right).$$

La cota obvia, obtenida al aplicar la desigualdad de Markov, es

$$\lambda \mathbb{P}\left(\overline{M}_n^+ > \lambda\right) \leq \mathbb{E}\left(\overline{M}_n^+\right);$$

el contenido no trivial de la desigualdad maximal de Doob es que de hecho podemos acotar la cola de la distribución del supremo de la martingala al utilizar la martingala misma.

Demostración. Recordemos que M_n^+ es una sub-martingala. Definamos a

$$A = \left\{\overline{M}_n^+ > \lambda\right\} \quad \text{y} \quad T = \min\left\{k \geq 0 : M_k^+ > \lambda\right\} \wedge n.$$

Notemos que

$$A \cap \{T = k\} = \left\{M_i^+ \leq \lambda \text{ para } i < k, M_k^+ > \lambda\right\} \in \mathscr{F}_k.$$

Por lo tanto

$$\lambda \mathbf{1}_{A\cap\{T=k\}} \leq M_k^+ \mathbf{1}_{A\cap\{T=k\}}.$$

Entonces

$$\begin{aligned}
\lambda\mathbb{P}(A) &= \sum_{k=0}^{n} \lambda\mathbb{P}(A\cap\{T=k\}) \\
&\leq \sum_{k=0}^{n} \mathbb{E}\big(M_k^+ \mathbf{1}_{A\cap\{T=k\}}\big) \\
&\leq \sum_{k=0}^{n} \mathbb{E}\big(M_n^+ \mathbf{1}_{A\cap\{T=k\}}\big) \\
&= \mathbb{E}\big(M_n^+ \mathbf{1}_{A}\big) \\
&\leq \mathbb{E}\big(M_n^+\big).
\end{aligned}$$

□

A partir de la desigualdad anterior, veremos que las normas p de $\overline{M}_n^+$ y de M_n^+ son comparables. Notemos que obviamente

$$\mathbb{E}\big(M_n^+\big) \leq \mathbb{E}\Big(\overline{M}_n^+\Big).$$

El contenido del siguiente resultado es establecer una especie de desigualdad recíproca.

Proposición 1.5 (Desigualdad L_p de Doob). *Para cualquier $p \in (1,\infty)$:*

$$\|\overline{M}_n^+\|_p \leq \frac{p}{p-1}\|M_n^+\|_p.$$

Demostración. Consideremos una constante $K>0$ y escribamos:

$$\begin{aligned}
\mathbb{E}\Big(\Big(\overline{M}_n^+ \wedge K\Big)^p\Big) &= \int_0^K p\lambda^{p-1}\mathbb{P}\Big(\overline{M}_n^+ > \lambda\Big)\, d\lambda \\
&\leq \int_0^K p\lambda^{p-2} \int M_n^+ \mathbf{1}_{\overline{M}_n^+ > \lambda}\, d\mathbb{P}\, d\lambda \\
&= \int M_n^+ \int_0^{\overline{M}_n^+\wedge K} p\lambda^{p-2}\, d\lambda\, d\mathbb{P} \\
&= \frac{p}{p-1}\int M_n^+ \Big(\overline{M}_n^+ \wedge K\Big)^{p-1}\, d\mathbb{P}.
\end{aligned}$$

Al utilizar la desigualdad de Hölder, utilizando el exponente conjugado $q = p/(p-1)$ se obtiene:

$$\mathbb{E}\Big(\Big(\overline{M}_n^+ \wedge K\Big)^p\Big) \leq \frac{p}{p-1}\|X_n^+\|_p \Big\|\Big(\overline{M}_n^+ \wedge K\Big)^q\Big\|_q,$$

por lo que despejando se obtiene

$$\|\overline{M}_n^+ \wedge K\|_p \leq \frac{p}{p-1}\|M_n^+\|_p.$$

La demostración termina al tomar el límite conforme $K \to \infty$. □

Finalmente, podemos obtener un criterio de convergencia en L_p para martingalas.

TEOREMA 1.6. *Si M_n es una martingala con $\sup_n \mathbb{E}(|M_n|^p) < \infty$ para alguna $p > 1$, X_n converge casi seguramente y en L_p a una variable M_∞ y se tiene que*

$$M_n = \mathbb{E}(\, M_\infty \mid \mathscr{F}_n)\,.$$

DEMOSTRACIÓN. La hipótesis implica que $\sup_n \mathbb{E}(|M_n|) < \infty$, por lo que el teorema de convergencia casi segura de martingalas nos permite afirmar que M_n converge casi seguramente a M_∞. Por otra parte, vemos que

$$\mathbb{E}\left(\sup_n |M_n|^{\,p}\right) = \lim_{n\to\infty} \mathbb{E}\left(\sup_{m\leq n} |M_n|^{\,p}\right) \leq \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}(|M_n|^{\,p}) < \infty.$$

Puesto que

$$|M_n - M_\infty|^p \leq \left(2\sup_n |M_n|\right)^p \in L_1,$$

podemos aplicar el teorema de convergencia dominada para ver que $M_n \to M_\infty$ en L_p conforme $n \to \infty$.

Finalmente, puesto que M_n converge a M_∞ en L_p también converge en L_1 y por lo tanto si $A \in \mathscr{F}$ entonces

$$\lim_{n\to\infty} \mathbb{E}(M_n \mathbf{1}_A) = \mathbb{E}(M_\infty \mathbf{1}_A)\,.$$

Por otra parte, si $A \in \mathscr{F}_m$ y $m \leq n$ entonces

$$\mathbb{E}(M_m \mathbf{1}_A) = \mathbb{E}(M_n \mathbf{1}_A)$$

puesto que X es una martingala. Así, también vemos que

$$\lim_{n\to\infty} \mathbb{E}(M_n \mathbf{1}_A) = \mathbb{E}(M_m \mathbf{1}_A)\,.$$

Se concluye que para todo $A \in \mathscr{F}_m$

$$\mathbb{E}(M_m \mathbf{1}_A) = \mathbb{E}(M_\infty \mathbf{1}_A)$$

y que por lo tanto

$$M_m = \mathbb{E}(\, M_\infty \mid \mathscr{F}_m)\,. \qquad \square$$

Sea Z el proceso de Galton-Watson de ley de reproducción μ que hemos considerado. Supongamos que $Z_0 = 1$ y que $m = \mathbb{E}(\xi_{1,1})$. Regresemos a la martingala $M_n = Z_n/m^n$ y a su línite casi seguro M_∞. Recordemos que el crecimiento de Z_n crece exponencialmente cuando $M_\infty > 0$. A continuación se presenta un criterio sencillo para que $M_\infty > 0$ con probabilidad positiva. El criterio insuperable en este sentido es el dado en el teorema de Kesten-Stigum y nos dice que $\mathbb{P}(M_\infty > 0) > 0$ si y sólo si $\mathbb{E}(\xi_{1,1} \log \xi_{1,1}) < \infty$.

PROPOSICIÓN 1.6. *Si $Var(\xi_{1,1}) \in (0, \infty)$ entonces $\mathbb{P}(M_\infty > 0) > 0$.*

Demostración. Sea $\sigma^2 = \mathrm{Var}(\xi_{1,1})$. Comenzamos por encontrar una relación de recurrencia para el segundo momento de M_n para lo cual calculamos

$$\begin{aligned}\mathbb{E}\big(M_{n+1}^2 \bigm| \mathscr{F}_n\big) &= M_n^2 + 2M_n\mathbb{E}(M_{n+1} - M_n \mid \mathscr{F}_n) + \mathbb{E}\Big((M_{n+1} - M_n)^2 \Bigm| \mathscr{F}_n\Big) \\ &= M_n^2 + \frac{1}{m^{2(n+1)}}\mathbb{E}\Big((Z_{n+1} - mZ_n)^2 \Bigm| \mathscr{F}_n\Big) \\ &= M_n^2 + \frac{1}{m^{2(n+1)}} Z_n\sigma^2.\end{aligned}$$

Obtenemos por lo tanto la relación de recurrencia

$$\mathbb{E}\big(M_{n+1}^2\big) = \mathbb{E}\big(M_n^2\big) + \frac{\sigma^2}{m^{n+2}}.$$

Se deduce que si $\mu > 1$:

$$\sup_n \mathbb{E}\big(M_n^2\big) = 1 + \sum_k \frac{\sigma^2}{m^{k+2}} < \infty$$

y por el teorema de convergencia en L_2 de Doob vemos que

$$\mathbb{E}\big(M_\infty^2\big) = \lim_{n\to\infty} \mathbb{E}(M_n) = \mathbb{E}(M_0) = 1 > 0.$$

Por lo tanto se concluye que $\mathbb{P}(M_\infty > 0) > 0$. □

2.7. La transformada martingalas. Sea $M = (M_n)$ una martingala. Recordemos que nuestra interpretación es que $M_n - M_{n-1}$ es la ganancia que obtenemos de apostar en un juego justo nuestra fortuna al tiempo $n-1$. Si ahora decidimos apostar la fracción C_n de nuestra fortuna, entonces nuestra ganancia será $C_n\,(M_n - M_{n-1})$. Así, a las cantidades $C_0, C_1, \ldots$ la podemos pensar como la estrategia de apuesta y C_n obviamente dependerá de la información que tengamos al tiempo $n-1$, que la habíamos interpretado como $\mathscr{F}_{n-1}$. En otras palabras, se requiere que C_n sea $\mathscr{F}_{n-1}$-medible. Esta condición define a lo que se conoce como un **proceso predecible**. Nuestra ganancia al tiempo n al seguir la estrategia de apuesta $C = C_1, C_2, \ldots$, que denotaremos por $(C \cdot M)_n$, está dada por

$$(C\cdot M)_0 = 0 \quad \text{y} \quad (C\cdot M)_n = \sum_{m=1}^{n} C_m\,(M_m - M_{m-1})\,.$$

Teorema 1.7. *Sea M una (sub)martingala y C un proceso predecible y acotado entonces $C\cdot M$ es una (sub)martingala.*

Ejercicio 1.12. Pruebe el teorema anterior.

El teorema anterior es otra manifestación del hecho de que no es posible generar ganancias en un juego justo.

Por ejemplo, consideremos la siguiente estrategia: sean $a < b$ dos reales fijos y apostaremos ya sea todo lo que tengamos o nada con las siguientes reglas. Nos fijamos en M y esperamos hasta que M se encuentre por debajo de a, ahí comenzamos a apostar, deteniéndonos cuando M se encuentre por arriba de b. Repetimos al infinito. Obviamente esta estrategia trata de utilizar a los cruces hacia arriba de la martingala en el intervalo $[a,b]$ para producir ganancias. La definición formal de la estrategia es como sigue:

$$C_1 = \mathbf{1}_{M_0 \leq a} \quad \text{y} \quad C_n = \mathbf{1}_{C_{n-1}=0}\mathbf{1}_{M_{n-1}\leq a} + \mathbf{1}_{C_{n-1}=1}\mathbf{1}_{M_{n-1}\leq b}.$$

Sea $Y = C \cdot M$.

Ejercicio 1.13. Sea U_n la cantidad de cruces hacia arriba que hace el proceso M en el intervalo $[a,b]$ antes de n. Argumente que

$$Y_n \geq (b-a)\, U_n + (M_n - a)^- .$$

Al tomar esperanzas verifique que se satisface la desigualdad de cruces de Doob

$$\mathbb{E}(U_n) \leq \frac{1}{b-a}\mathbb{E}\Big((a-M_n)^+\Big) .$$

3. Martingalas e integrabilidad uniforme

El objetivo de esta sección es analizar el concepto de integrabilidad uniforme de una familia de variables aleatorias integrables. El interés de esta noción, para el estudio de las martingalas, es que permite caracterizar a las martingalas que son de la forma $X_n = \mathbb{E}(X \mid \mathscr{F}_n)$, y con esto, permite dar una versión muy general del teorema de paro opcional de Doob.

Definición. Sea $\{X_t\}_{t\in T}$ es una familia de variables aleatorias reales. Decimos que es uniformemente integrable si

$$\lim_{c\to\infty} \sup_{t\in T} \mathbb{E}\big(|X_t|\, \mathbf{1}_{|X_t|\geq c}\big) = 0.$$

Ejemplo 1.5. La familia que consta de un sólo elemento $X \in L_1$ es uniformemente integrable. Esto se sigue de aplicar el teorema de convergencia dominada para concluir que

$$\lim_{c\to\infty} \mathbb{E}\big(|X|\, \mathbf{1}_{|X|>c}\big) = 0$$

al ser X casi seguramente finita.

Ejemplo 1.6. Si $\{X_t\}_{t\in T}$ es tal que $\sup_{t\in T}\mathbb{E}(|X_t|^{\,p}) < \infty$ para alguna $p > 1$ entonces dicha familia es uniformemente integrable. En efecto, basta notar que

$$c^{p-1}\mathbb{E}\big(|X_t|\, \mathbf{1}_{|X_t|>c}\big) \leq \mathbb{E}\big(|X_t|^{\,p}\big) .$$

Ejemplo 1.7. Para cada $X \in L_1$, a la familia

$$E = \{\mathbb{E}(X \mid \mathscr{G}) : \mathscr{G} \text{ es sub}\sigma\text{-álgebra de } \mathscr{F}\} .$$

Se afirma que E es uniformemente integrable. En efecto, la desigualdad de Jensen implica que

$$\mathbb{E}\big(|\mathbb{E}(X \mid \mathscr{G})|\, \mathbf{1}_{|\mathbb{E}(X \mid \mathscr{G})|>c}\big) \leq \mathbb{E}\big(\mathbb{E}(\,|X| \mid \mathscr{G})\, \mathbf{1}_{|\mathbb{E}(X \mid \mathscr{G})|>c}\big) = \mathbb{E}\big(X\mathbf{1}_{|\mathbb{E}(X \mid \mathscr{G})|>c}\big) .$$

Por la desigualdad de Markov, vemos que

$$\mathbb{P}(|\mathbb{E}(X \mid \mathscr{G})| > c) \leq \frac{1}{c}\mathbb{E}(|\mathbb{E}(X \mid \mathscr{G})|) \leq \frac{1}{c}\mathbb{E}(|X|),$$

por lo cual

$$\lim_{c\to\infty} \sup_{\mathscr{G}\subset\mathscr{F}} \mathbb{P}(|\mathbb{E}(X \mid \mathscr{G})| > c) = 0.$$

Finalmente, se afirma que para cada $\varepsilon > 0$ existe $\delta > 0$ tal que si $\mathbb{P}(E) < \delta$, entonces $\mathbb{E}(|X| \mathbf{1}_E) < \varepsilon$. Esto se prueba a partir de la desigualdad

$$\mathbb{E}(|X| \mathbf{1}_E) \leq c\mathbb{P}(E) + \mathbb{E}\big(|X| \mathbf{1}_{|X|>c}\big).$$

Por convergencia dominada, el segundo término del lado derecho tiende a cero conforme $c \to \infty$. Así, dada $\varepsilon > 0$, escogemos $c > 0$ tal que el segundo sumando del lado derecho sea menor a $\varepsilon/2$. Basta entonces tomar $\delta < \varepsilon/2c$. Esto termina la prueba de que E es uniformemente integrable, puesto que dada $\varepsilon > 0$, escogemos δ tal que si $\mathbb{P}(A) < \delta$ entonces $\mathbb{E}(|X| \mathbf{1}_A) < \varepsilon$ y finalmente C tal que para toda $c \geq C$ y toda $\mathscr{G}$ subσ-álgebra de $\mathscr{F}$ tengamos

$$\mathbb{P}(|\mathbb{E}(X \mid \mathscr{G})| > c) < \delta.$$

Entonces

$$\mathbb{E}\big(|\mathbb{E}(X \mid \mathscr{G})| \mathbf{1}_{|\mathbb{E}(X \mid \mathscr{G})|>c}\big) < \varepsilon$$

para toda $c > C$.

En vista del ejemplo anterior, si $X_n = \mathbb{E}(X \mid \mathscr{F}_n)$ y $(\mathscr{F}_n, n \in \mathbb{N})$ es una filtración entonces la martingala (X_n) es uniformemente integrable. Un ejemplo de una martingala que no es uniformemente integrable es el siguiente: si U es una variable uniforme en $(0,1)$ y $X_n = 2^n \mathbf{1}_{U\leq 2^n}$, entonces X_n es una martingala respecto a la filtración que genera. Puesto que $U > 0$ casi seguramente, se sigue que $X_n \to 0$ casi seguramente. Sin embargo, $\lim_{c\to\infty} \sup_n \mathbb{E}(X_n \mathbf{1}_{X_n \geq c}) = 1$, por lo que no es uniformemente integrable.

Si $\{X_t\}_{t\in T}$ es uniformemente integrable, sea $c > 0$ tal que

$$\sup_{t\in T} \mathbb{E}(|X_t| \mathbf{1}_{X_t \geq c}) \leq 1.$$

Vemos que entonces

$$\sup_{t\in T} \mathbb{E}(|X_t|) \leq c + 1 < \infty,$$

por lo que la familia $\{X_t\}_{t\in T}$ es acotada en L_1.

La importancia de la integrabilidad uniforme es que nos permite relacionar dos modos de convergencia, la casi segura y la convergencia en L_1:

Teorema 1.8. *Si $\{X_n\}_{n\in\mathbb{N}}$ y X son variables aleatorias integrables tales que $X_n \to X$ casi seguramente, entonces las siguientes condiciones son equivalentes:*

a) $\{X_n\}_{n\in\mathbb{N}}$ es uniformemente integrable.
b) $X \in L_1$ y $X_n \to X$ en L_1.

Como hemos visto anteriormente, una condición necesaria y suficiente para que una sucesión convergente casi seguramente también sea convergente en L_1 es que la sucesión sea uniformemente integrable, por lo que ahora estudiaremos martingalas uniformemente integrables para abarcar otro modo de convergencia en el estudio de las martingalas.

Si $(X_n)_{n\in\mathbb{N}}$ es una martingala uniformemente integrable (respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$) entonces el conjunto $\{\mathbb{E}(|X_n|) : n \in \mathbb{N}\}$ es acotado, por lo que se satisfacen las condiciones del teorema de convergencia de martingalas y por lo tanto existe una variable aleatoria integrable X a la que la sucesión converge casi seguramente conforme $n \to \infty$. Por ser la martingala uniformemente integrable, la convergencia también se dá en L_1. Si A es un elemento de $\mathscr{F}_n$, la tercera condición que define a las martingalas nos permite afirmar que

$$\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_m \mathbf{1}_A) \quad \forall m \geq n,$$

y como

$$\mathbb{E}(X_n \mathbf{1}_A) \to \mathbb{E}(X \mathbf{1}_A)$$

por la convergencia de $(X_n)_{n\in\mathbb{N}}$ a X en L_1, entonces

$$\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X \mathbf{1}_A) \quad \forall A \in \mathscr{F}_n,$$

de donde se concluye que $X_n = \mathbb{E}(X \mid \mathscr{F}_n)$ (pues X_n es $\mathscr{F}_n$-medible) y por lo tanto, la martingala original era una martingala cerrada. De hecho:

Teorema 1.9. *Sea $(X_n)_{n\in\mathbb{N}}$ una martingala respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$. Entonces existe una variable aleatoria integrable X tal que $X_n = \mathbb{E}(X \mid \mathscr{F}_n)$ si y sólo si $\{X_n : n \in \mathbb{N}\}$ es uniformemente integrable. Además, si se cumple alguna de las condiciones anteriores, $(X_n)_{n\in\mathbb{N}}$ converge casi seguramente y en L_1 a $\mathbb{E}(X \mid \mathscr{F}_\infty)$, donde*

$$\mathscr{F}_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathscr{F}_n\right).$$

Nota. Para una martingala cerrada, $(\mathbb{E}(X \mid \mathscr{F}_n))_{n\in\mathbb{N}}$ el límite casi seguro no tiene porque ser igual a X; sin embargo, si $\mathscr{F} = \mathscr{F}_\infty$, entonces el límite casi seguro sí es igual a X.

Demostración. En el párrafo anterior, hemos visto como para cualquier martingala uniformemente integrable existe una variable aleatoria integrable X que la convierte en una martingala cerrada. Así, sólo hace falta verificar que una martingala cerrada es uniformemente integrable. Pero esto es inmediato, pues hemos verificado que si Σ es una familia de σ-álgebras en Ω contenidas en $\mathscr{F}$, entonces la familia de variables aleatorias $\{\mathbb{E}(X \mid G) : G \in \Sigma\}$ es uniformemente integarble, por lo que cualquier martingala cerrada lo es.

Si se satisfacen alguna de las dos condiciones, sea Y el límite casi seguro y en L_1 para la martingala $(Y_n = \mathbb{E}(X \mid \mathscr{F}_n))_{n\in\mathbb{N}}$. Como Y_n es $\mathscr{F}_\infty$-medible para cada

$n \in \mathbb{N}$, se sigue que Y también lo es. Además, por la convergencia en L_1, se sigue que para todo $A \in \mathscr{F}_n$,

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y_n\mathbf{1}_A) = \lim_{m\to\infty} \mathbb{E}(Y_{n+m}\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A).$$

Sea

$$\mathscr{C} = \{A \in \mathscr{F}_\infty : \mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(Y\mathbf{1}_A)\}.$$

Hemos visto que

$$\bigcup_{n\in\mathbb{N}} \mathscr{F}_n \subset \mathscr{C} \subset \mathscr{F}_\infty$$

y la anterior unión de σ-álgebras es un álgebra. Además, $\mathscr{C}$ es una clase monótona, puesto que si $(A_n)_{n\in\mathbb{N}}$ es una sucesión creciente o decreciente de elementos de $\mathscr{C}$ y A es el límite de la anterior sucesión de conjuntos (igual al límite superior o al inferior, que coinciden) entonces el teorema de convergencia dominada nos permite afirmar que

$$\mathbb{E}(X\mathbf{1}_A) = \lim_{n\to\infty} \mathbb{E}(X\mathbf{1}_{A_n}) = \lim_{n\to\infty} \mathbb{E}(Y\mathbf{1}_{A_n}) = \mathbb{E}(Y\mathbf{1}_A),$$

por lo que A pertenece a $\mathscr{C}$. Así, por el lema de clases monótonas, $\mathscr{C} = \mathscr{F}_\infty$, lo cual nos dice que

$$Y = \mathbb{E}(X \mid \mathscr{F}_\infty). \qquad \square$$

Bajo la hipótesis de integrabilidad uniforme, también podemos dar una primera extensión del teorema (1.3):

Teorema 1.10. *Si $(X_n)_{n\in\mathbb{N}}$ es una martingala uniformemente integable respecto a la filtración $(\mathscr{F}_n)_{n\in\mathbb{N}}$ y T es un tiempo de paro respecto a la misma filtración entonces X_T es una variable aleatoria integrable y $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.*

Demostración. Puesto que X_n converge conforme $n \to \infty$, digamos a X_∞, podemos definir a X_T aún cuando T no sea finito. Para ver que X_T es integrable, notemos que para cada $A \in \mathscr{F}_n$:

$$\mathbb{E}(|X_n|\,\mathbf{1}_A) \leq \mathbb{E}(\mathbb{E}(\,|X_\infty|\mid \mathscr{F}_n)\,\mathbf{1}_A) = \mathbb{E}(|X_\infty|\,\mathbf{1}_A).$$

Dado que T es tiempo de paro, el evento $\{T = n\}$ pertenece a $\mathscr{F}_n$, por lo que de acuerdo a la desigualdad anterior

$$\begin{aligned}
\mathbb{E}(|X_T|) &= \mathbb{E}(|X_\infty|\,\mathbf{1}_{T=\infty}) + \sum_{n\in\mathbb{N}} \mathbb{E}\big(|X_n|\,\mathbf{1}_{(T=n)}\big) \\
&\leq \sum_{n\in\mathbb{N}} \mathbb{E}\big(|X|\,\mathbf{1}_{(T=n)}\big) \\
&= \mathbb{E}(|X|) < \infty.
\end{aligned}$$

Finalmente, sea

$$Y_n = X_\infty\mathbf{1}_{T=\infty} + \sum_{i=0}^{n} X_i\mathbf{1}_{(T=i)},$$

por lo que $(Y_n)_{n\in\mathbb{N}}$ converge casi seguramente a X_T. Por la las desigualdades

$$|Y_n| \leq |X_\infty| \, \mathbf{1}_{T=\infty} + \sum_{i=0}^{n} |X_i| \, \mathbf{1}_{(T=i)} \leq |X_T| \in L_1,$$

podemos aplicar el teorema de convergencia dominada para concluir que

$$\mathbb{E}(Y_n) \to \mathbb{E}(X_T)$$

y como

$$\begin{aligned}\mathbb{E}(Y_n) &= \mathbb{E}(X_\infty \mathbf{1}_{T=\infty}) + \sum_{i=0}^{n} \mathbb{E}\big(X_i \mathbf{1}_{(T=i)}\big) \\ &= \mathbb{E}(X \mathbf{1}_{T=\infty}) + \sum_{i=0}^{n} \mathbb{E}\big(X \mathbf{1}_{(T=i)}\big) \\ &= \mathbb{E}(X \mathbf{1}_{T\leq n \text{ ó } T=\infty}),\end{aligned}$$

se concluye, mediante el uso del teorema de convergencia dominada que

$$\mathbb{E}(Y_n) \to \mathbb{E}(X) = \mathbb{E}(X_0)$$

y por lo tanto

$$\mathbb{E}(X_T) = \mathbb{E}(X_0)\,. \qquad \square$$

La integrabilidad uniforme nos da un criterio importante para ver si podemos aplicar el teorema de muestreo opcional de Doob. En efecto, si $X = (X_n, n \in \mathbb{N})$ es cualquier martingala y T es un tiempo de paro finito, la integrabilidad uniforme de la martingala detenida $X^T = (X_{n\wedge T}, n \in \mathbb{N})$ implica la igualdad $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

4. La ley $0-1$ de Kolmogorov

En esta sección veremos una primera aplicación de los teoremas de convergencia de martingalas a sucesiones de variables aleatorias independientes. Sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad en el cual están definidas una sucesión de variables aleatorias independientes $(X_i)_{i\in\mathbb{N}}$. Definiremos a

$$\mathscr{F}_n = \sigma(X_i : i \leq n) \quad \text{y a} \quad \mathscr{F}_\infty = \sigma(X_i : i \in \mathbb{N})\,.$$

El resultado que probaremos, la ley $0-1$ de Kolmogorov, nos permite concluir bajo ciertas hipótesis, que un elemento de $\mathscr{F}_\infty$ tiene probabilidad 0 ó 1. Para esto, sean

$$\mathscr{G}_n = \sigma(X_i : i > n) \quad \text{y} \quad \mathscr{T} = \bigcap_{n\in\mathbb{N}} \mathscr{G}_n.$$

Recordemos que $\mathscr{T}$ es una σ-álgebra, pues es intersección de σ-álgebras, a la cual llamaremos la σ-álgebra cola. Por la hipótesis acerca de la independencia de las variables aleatorias se obtiene lo siguiente.

Proposición 1.7. *Las σ-álgebras $\mathscr{F}_n$ y $\mathscr{G}_n$ son independientes. Además, la σ-álgebra τ es independiente de $\mathscr{F}_n$ para toda $n \in \mathbb{N}$.*

Finalmente, podemos enunciar y demostrar la Ley 0-1 de Kolmogorov:

Teorema 1.11. *Si $A \in \tau$, entonces $\mathbb{P}(A) \in \{0,1\}$.*

Demostración. Como $A \in \tau$, entonces A es independiente de $\mathscr{F}_n$ para cualquier $n \in \mathbb{N}$, de donde

$$\mathbb{E}(\mathbf{1}_A \mid \mathscr{F}_n) = \mathbb{P}(A) \quad \forall n \in \mathbb{N}.$$

Por otro lado, sabemos que

$$\mathbb{E}(\mathbf{1}_A \mid \mathscr{F}_n) \to \mathbb{E}(\mathbf{1}_A \mid \mathscr{F}_\infty) \quad \mathbb{P}-p.s.$$

y como $A \in \mathscr{F}_\infty$, entonces

$$\mathbb{P}(A) = \mathbf{1}_A \quad \mathbb{P}-p.s.,$$

de donde $\mathbb{P}(A) \in \{0,1\}$. □

Veamos algunos casos particulares del resultado anterior. Si $S_n = X_1 + \cdots + X_n$ y a_n es una sucesión de reales que tiende a ∞ entonces $\limsup_{n\to\infty} S_n/a_n$ y $\liminf_{n\to\infty} S_n/a_n$ son variables aleatorias τ-medibles. Por lo tanto, son casi seguramente constantes. Vemos además que $\mathbb{P}(\limsup_{n\to\infty} S_n/a_n = \liminf_{n\to\infty} S_n/a_n) \in \{0,1\}$. Esto explica por qué, si las variables X_i son además idénticamente distribuidas, se observa que la existencia del límite de S_n/n ocurre con probabilidad cero o uno y por qué, de existir con probabilidad uno, se trata de una variable aleatoria casi seguramente constante.

5. Martingalas reversas y la ley fuerte de los grandes números

En esta sección, exploraremos una situación análoga a la contenida en el teorema de convergencia de martingalas naturales. Lo que se consideró en ese teorema fué la existencia del límite casi seguro y en L_1 de la sucesión

$$(\mathbb{E}(X \mid \mathscr{F}_n))_{n\in\mathbb{N}},$$

donde $(\mathscr{F}_n)_{n\in\mathbb{N}}$ es una subsucesión creciente de σ-álgebras de $\mathscr{F}$. Si ahora consideramos una colección $(\mathscr{F}_n)_{n\in\mathbb{Z}}$ creciente, podemos concluir la existencia casi segura y en L_1 de $\mathbb{E}(X \mid \mathscr{F}_n)$ conforme $n \to -\infty$? Si $\mathscr{G}_n = \mathscr{F}_{-n}$ para $n \in \mathbb{N}$, por lo que $\mathscr{G}_{n+1} \subset \mathscr{G}_n$, lo que quisieramos afirmar es la existencia del límite casi seguro y en L_1 de $\mathbb{E}(X \mid \mathscr{G}_n)$ conforme $n \to \infty$.

Para indicar la relevancia de tal afirmación, consideremos una sucesión $(X_i)_{i=1}^\infty$ de variables aleatorias independientes e idénticamente distribuidas con media cero, $(S_n)_{n\in\mathbb{N}}$ la sucesión de sumas parciales asociadas y para $n \in \mathbb{N}$, $\mathscr{G}_n = \sigma(S_n, S_{n+1}, \ldots)$. Se deja como ejercicio al lector comprobar que

$$\mathbb{E}(X_i \mid \mathscr{G}_n) = \frac{S_n}{n}, \quad i = 1, \ldots, n, n \geq 1$$

por lo que

$$\mathbb{E}(S_n \mid \mathscr{G}_{n+1}) = \frac{n}{n+1} S_{n+1},$$

de donde

$$\mathbb{E}\left(\frac{S_n}{n} \,\middle|\, \mathscr{G}_{n+1}\right) = \frac{S_{n+1}}{n+1}.$$

De aquí se sigue que

$$\frac{S_n}{n} = \mathbb{E}(X_1 \mid \mathscr{G}_n)$$

puesto que $\mathscr{G}_{n+1} \subset \mathscr{G}_n$, por lo cual la pregunta que formulamos anteriormente es acerca del límite casi seguro y en L_1 de S_n/n, resultado conocido como la ley fuerte de los grandes números.

Utilizaremos a continuación los resultados ya verificados sobre martingalas para atacar la pregunta que nos concierne, para lo cual necesitamos precisar cuales son los procesos con los cuales vamos a trabajar.

Definición. Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias y $(\mathscr{G}_n)_{n\in\mathbb{N}}$ una sucesión decreciente de σ-álgebras contenidas en $\mathscr{F}$. Decimos que $(X_n)_{n\in\mathbb{N}}$ es una martingala reversa respecto a $(\mathscr{G}_n)_{n\in\mathbb{N}}$ si

(1) $X_n \in L_1$ para toda $n \in \mathbb{N}$,
(2) X_n es $\mathscr{G}_n$-medible para toda $n \in \mathbb{N}$ y
(3) $\mathbb{E}(X_n \mid \mathscr{G}_{n+1}) = X_{n+1}$.

Notemos que de las propiedades de la esperanza condicional se sigue la igualdad

$$X_{n+2} = \mathbb{E}(X_{n+1} \mid \mathscr{G}_{n+2}) = \mathbb{E}(\mathbb{E}(X_n \mid \mathscr{G}_{n+1}) \mid \mathscr{G}_{n+2}) = \mathbb{E}(X_n \mid \mathscr{G}_{n+2}),$$

por lo que de manera inductiva se verifica

$$X_n = \mathbb{E}(X_0 \mid \mathscr{G}_n).$$

Así, la pregunta formulada anteriormente es simplemente verificar si una martingala reversa tiene un límite casi seguro y en L_1.

Teorema 1.12. *Sea $(X_n)_{n\in\mathbb{N}}$ una martingala reversa respecto a $(\mathscr{G}_n)_{n\in\mathbb{N}}$. Entonces $(X_n)_{n\in\mathbb{N}}$ converge casi seguramente conforme $n \to \infty$ a una variable aleatoria X que pertenece a L_1.*

Nota. Nos basaremos en el teorema de convergencia casi segura para martingalas. En dicho teorema, se vió que para demostrar la existencia del límite casi seguro, era suficiente verificar que la cantidad de cruces cruces hacia arriba de $X_0, X_1, \ldots$ en $[a,b]$, denotada por $U_{[a,b]}(X_0, X_1, \ldots)$, era finita casi seguramente para cualquier pareja de racionales a, b tal que $a < b$. En la prueba del teorema, vimos que la cantidad de cruces hacia arriba de $X_0, X_1, \ldots$ en $[a,b]$ así como la cantidad de cruces hacia arriba de $X_0, \ldots, X_n$ en $[a,b]$, $U^n_{[a,b]}(X_0, \ldots, X_n)$ eran variables aleatorias y se demostró la desigualdad clásica de Doob.

Demostración. Sea $m \in \mathbb{N}$, $Y_n = -X_{(m-n)\vee 0}$ y $\mathscr{H}_n = \mathscr{G}_{(m-n)\vee 0}$ para $n \in \mathbb{N}$. Verifiquemos que $(Y_n)_{n\in\mathbb{N}}$ es una martingala respecto a $(\mathscr{H}_n)_{n\in\mathbb{N}}$: si $n \in \mathbb{N}$, entonces

$$\begin{aligned}\mathbb{E}(Y_{n+1} \mid \mathscr{H}_n) &= \begin{cases}\mathbb{E}(-X_{m-n-1} \mid \mathscr{G}_{n-m}) & \text{si } n < m\\ \mathbb{E}(-X_0 \mid \mathscr{G}_n) & \text{si } n \geq m\end{cases}\\ &= \begin{cases}-X_{n-m} & \text{si } n < m\\ -X_0 & \text{si } n \geq m\end{cases}\\ &= Y_n.\end{aligned}$$

Así, por la desigualdad clásica de Doob, y al utilizar la igualdad

$$U^n_{[-b,-a]}(Y_0, \ldots, Y_n) = U^n_{[a,b]}(X_0, \ldots, X_n)$$

se tiene que

$$\mathbb{E}\Big(U^n_{[a,b]}(X_0, \ldots, X_n)\Big) \leq \frac{1}{b-a}\left(|b| + \mathbb{E}(|Y_n|)\right) = \frac{1}{b-a}\left(|b| + \mathbb{E}(|X_0|)\right).$$

De esta manera, vemos que

$$\mathbb{E}\big(U_{[a,b]}(X_0, X_1, \ldots)\big) \leq \frac{1}{b-a}\left(|b| + \mathbb{E}(|X_0|)\right),$$

por lo que $U_{[a,b]}(X_0, X_1, \ldots) < \infty$ $\mathbb{P}$-p.s.. Esto nos dice que existe el límite casi seguro de X_n conforme $n \to \infty$ y para ver que pertenece a L_1, aplicamos el lema de Fatou:

$$\mathbb{E}\Big(\Big|\lim_{n\to\infty} X_n\Big|\Big) \leq \liminf_{n\to\infty} \mathbb{E}(|X_n|) = \liminf_{n\to\infty} \mathbb{E}(|\mathbb{E}(X_0 \mid \mathscr{G}_n)|) \leq \liminf_{n\to\infty} \mathbb{E}(|X_0|) < \infty.$$

□

Ahora, veamos que toda martingala reversa es uniformemente integrable, por lo que la convergencia casi segura nos permitirá concluir la convergencia en L_1.

Teorema 1.13. *Si $(X_n)_{n\in\mathbb{N}}$ es una martingala reversa respecto a $(\mathscr{G}_n)_{n\in\mathbb{N}}$, entonces es uniformemente integrable.*

Demostración. Como $X_n = \mathbb{E}(X_0 \mid \mathscr{G}_n)$ y hemos visto que

$$\{\mathbb{E}(X \mid \mathscr{G}) : \mathscr{G} \in G\},$$

con G una familia de σ-álgebras contenidas en $\mathscr{F}$ y X un elemento de L_1 es uniformemente integrable, se sigue que $\{X_n\}_{n\in\mathbb{N}}$ es uniformemente integrable. □

Pasaremos a la identificación del límite:

Teorema 1.14. *Sea $(X_n)_{n\in\mathbb{N}}$ una martingala reversa respecto a $(\mathscr{G}_n)_{n\in\mathbb{N}}$. Entonces*

$$\lim_{n\to\infty} X_n = \mathbb{E}\left(X_0 \,\middle|\, \bigcap_{n\in\mathbb{N}} \mathscr{G}_n\right).$$

Demostración. Sea X el límite casi seguro y en L_1 de $(X_n)_{n\in\mathbb{N}}$. Como X_m es $\mathscr{G}_n$-medible para cualquier $m \geq n$, se sigue que

$$X = \lim_{m\to\infty} X_{n+m}$$

es $\mathscr{G}_n$-medible para toda $n \in \mathbb{N}$, por lo que es medible respecto a la intersección de dichas σ-álgebras. Por otro lado, si $A \in \mathscr{G}_n$ para toda $n \in \mathbb{N}$, entonces

$$\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_0 \mathbf{1}_A)$$

y por la convergencia de X_n a X en L_1,

$$\mathbb{E}(X \mathbf{1}_A) = \lim_{n\to\infty} \mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_0 \mathbf{1}_A) . \qquad \square$$

Para finalizar esta sección, daremos una prueba de la ley fuerte de los grandes números que utiliza las ideas que se han desarrollado.

Teorema 1.15. *Sea $(X_i)_{i\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que $X_i \in L_1$. Si $(S_n)_{n=0}^{\infty}$ denota a la sucesión de sumas parciales asociada, entonces*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}(X_1) \quad \mathbb{P}\text{-}p.s..$$

Demostración. Sea $\mathscr{G}_n = \sigma(S_k : k \geq n)$ y $\mu = \mathbb{E}(X_1)$, por lo que $(S_n/n - \mu)_{n\geq 1}$ es una martingala reversa respecto a $(\mathscr{G}_n)_{n\in\mathbb{N}}$. Esto nos dice que $S_n/n - \mu$ tiene un límite conforme $n \to \infty$ y como

$$\lim_{n\to\infty} \frac{S_n}{n} - \mu = \lim_{n\to\infty} \frac{S_n - S_m}{n} - \mu = \lim_{n\to\infty} \frac{X_{m+1} + \cdots + X_n}{n} - \mu,$$

se sigue que dicho límite es medible respecto a $\sigma(X_k : k \geq m)$ para toda $m \in \mathbb{N}$, de donde es medible respecto a la σ-álgebra cola asociada a $(X_i)_{i\in\mathbb{N}}$ y por lo tanto es constante. Para determinar la constante[5], recordemos que S_n/n también converge en L_1 y que una variable aleatoria constante es igual a su esperanza, por lo que

$$\lim_{n\to\infty} \frac{S_n}{n} - \mu = \mathbb{E}\left(\lim_{n\to\infty} \frac{S_n}{n} - \mu\right) = \lim_{n\to\infty} \mathbb{E}\left(\frac{S_n}{n} - \mu\right) = \mathbb{E}(S_1 - \mu) = 0,$$

de donde S_n/n converge a μ. $\square$

6. Urnas de Pólya y el teorema de de Finetti

Comencemos por analizar con mayor profundidad el ejemplo de las urnas de Pólya. Sean (U_i) variables aleatorias independientes e idénticamente distribuidas, $r, v > 0$ y definamos $X_0 = r/(r+v)$ y para $n \geq 0$:

$$Y_{n+1} = \mathbf{1}_{U_{n+1} \leq X_n} \quad \text{y} \quad X_{n+1} = \frac{r + v + nc}{r + v + (n+1)\,c} X_n + \frac{c}{r + v + (n+1)\,c} Y_n.$$

[5]Para no utilizar la integrabilidad uniforme en lo que sigue, se puede usar la ley débil de los grandes números.

Hemos interpretado a la sucesión $(X_n)_{n\in\mathbb{N}}$ como la fracciónes sucesivas de bolas rojas en una urna que inicialmente contiene r bolas rojas y v bolas verdes (aunque con esta construcción podemos considerar a r y a v como reales positivos) y tal que, en cada unidad de tiempo, se revuelve, se extrae una bola y se regresa con c bolas del mismo color. Además, hemos visto que si $\mathscr{F}_n = \sigma(U_1,\ldots,U_n)$, entonces (X_n) es una $(\mathscr{F}_n)$-martingala acotada. Esto implica que converge casi seguramente y en L_∞ a una variable aleatoria X_∞ que se puede interpretar como la proporción límite de la urna. Ahora determinaremos la distribución de X_∞ mediante una técnica importante que básicamente generaliza nuestra prueba de la ley fuerte de los grandes números.

Analicemos ahora a la sucesión (Y_n): la variable Y_n se interpreta como la indicadora de que en la enésima extracción so obtuvo una bola roja. Ahora calcularemos la distribución conjunta de $(Y_1,\ldots,Y_n)$, para lo cual necesitamos la notación del factorial ascendente

$$a^{(n)} = a\,(a+1)\cdots(a+n-1)\,.$$

Proposición 1.8. *Las variables aleatorias* (Y_n) *son intercambiables. De hecho, si* $i_1,\ldots,i_n\in\{0,1\}$ *y* $s_n = i_1+\cdots+i_n$ *entonces*

$$\mathbb{P}(Y_1 = i_1,\ldots,Y_n = i_n) = \frac{(r/c)^{(s_n)}(v/c)^{(n-s_n)}}{((r+v)\,/c)^{(n)}}.$$

Ejercicio 1.14. Pruebe la proposición anterior. Sugerencia, utilice el principio de inducción.

Definiremos a $S_n = Y_1+\cdots+Y_n$, por lo que

$$X_n = \frac{r+v+cS_n}{r+v+nc}$$

y por lo tanto $S_n/n\to X_\infty$. Más adelante, justificaremos el hecho de que

$$\mathbb{E}(\,Y_j \mid S_n, Y_{n+1}, Y_{n+2},\ldots) = \mathbb{E}(\,Y_1 \mid S_n, Y_{n+1}, Y_{n+2},\ldots) \text{ si } 1\le j\le n,$$

de lo cual obtendremos

$$\mathbb{E}(\,Y_1 \mid S_n, Y_{n+1},\ldots) = S_n/n.$$

Al tomar el límite conforme $n\to\infty$, vemos que

$$\frac{r}{r+v} = \mathbb{E}(Y_1) = \mathbb{E}(X_\infty)\,.$$

Procederemos análogamente para el cálculo de los momentos de X_∞. Sea $\mathscr{G}_n$ la σ-álgebra generada por las variables $f(Y_1,\ldots,Y_n)$, donde f es una función (medible y) simétrica. Sea $\mathscr{H}_n = \sigma(\mathscr{G}_n, Y_{n+1}, Y_{n+2},\ldots)$.

Proposición 1.9. *Si* π *es una permutación de los índices* 1 *al* n *entonces*

$$\mathbb{E}(\,f(Y_1,\ldots,Y_n)\mid\mathscr{H}_n) = \mathbb{E}(\,f(Y_{\pi_1},\ldots,Y_{\pi_n})\mid\mathscr{H}_n)\,.$$

La proposición anterior nos permite hacer utilizar la intercambiabilidad para hacer cálculos. Por ejemplo, al utilizar $f(y_1, \ldots, y_n) = y_1$,vemos que $\mathbb{E}(Y_1 \mid \mathscr{H}_n) = S_n/n$. Otro ejemplo interesante es

$$\mathbb{E}(Y_1 Y_2 \mid \mathscr{H}_n) = \frac{2}{n\,(n-1)} \sum_{\substack{1\leq i,j\leq n\\ i\neq j}} Y_j Y_j = \frac{1}{n\,(n-1)} \left(S_n^2 - \sum_{i=1}^n Y_i^2 \right) \to X_\infty^2.$$

Por lo tanto, vemos que

$$\mathbb{E}\big(X_\infty^2\big) = \mathbb{E}(Y_1 Y_2) = \mathbb{P}(Y_1 = 1, Y_2 = 1) = \frac{r\,(r+c)}{(r+v)\,(r+v+c)} = \frac{\mathrm{B}(r/c+1, v/c)}{\mathrm{B}(r/c, v/c)}.$$

El mismo argumento, muestra que

$$\mathbb{E}(X_\infty^n) = \frac{\mathrm{B}(r/c+n, v/c)}{\mathrm{B}(r/c, v/c)},$$

de lo cual se deduce que X_∞ tiene los mismos momentos que una variable B de parámetros r/c y v/c. Al ser la variable B acotada, cualquier variable aleatoria que tenga los mismos momentos tendrá dicha distribución. Se concluye que X_∞ tiene distribución B de parámetros r/c y v/c. Sin embargo, también obtenemos una consecuencia sorprendente: aunque las variables $Y_1, Y_2, \ldots$ disten mucho de ser iid, vemos que si $\mathscr{H}_\infty = \bigcap_n \mathscr{H}_n$, $i_1, \ldots, i_n \in \{0, 1\}$ y $s_n = i_1 + \cdots + i_n$ entonces:

$$\mathbb{P}(Y_1 = i_1, \ldots, Y_n = i_n \,|\mathscr{H}_\infty) = X_\infty^{s_n}\,(1 - X_\infty)^{n-s_n} = \prod_{j=1}^n \mathbb{P}(Y_1 = i_j \,|\mathscr{H}_\infty)\,,$$

por lo cual la sucesión $Y_1, Y_2, \ldots$ es iid, pero condicionalmente a $\mathscr{H}_\infty$ (y por lo tanto, también condicionalmente a X_∞).

Este es un caso particular del teorema de de Finetti que afirma que toda sucesión de variables intercambiables es condicionalmente iid.

7. Regularización de martingalas

Ahora daremos una extensión adicional del teorema de convergencia de martingalas y probaremos el teorema de regularización de martingalas. Este último es útil a la construcción de una gran familia de procesos estocásticos entre los cuales se encuentran los procesos de Feller y en particular los procesos de Lévy. Nos centraremos en procesos a tiempo continuo.

Definición. Una **filtración** a tiempo continuo es una colección $(\mathscr{F}_t)_{t\geq 0}$ de subσ-álgebras de $\mathscr{F}$ tales que si $s \leq t$ entonces $\mathscr{F}_s \leq \mathscr{F}_t$. Decimos que la filtración es **continua por la derecha** si, al definir

$$\mathscr{F}_{t+} = \bigcap_{u>t} \mathscr{F}_u,$$

se tiene que $\mathscr{F}_t = \mathscr{F}_{t+}$. Decimos que la filtración es **completa** si $\mathscr{F}_0$ (y por lo tanto también cada $\mathscr{F}_t$ con $t > 0$) contienen a los conjuntos $\mathbb{P}$ nulos de $\mathscr{F}_\infty$. Decimos que la filtración satisface las **hipótesis habituales** si es continua por la derecha y completa.

Una colección de variables aleatorias $(X_t)_{t\geq 0}$ es una **martingala** respecto de $(\mathscr{F}_t)_{t\geq 0}$ si

(1) X_t es $\mathscr{F}_t$-medible.
(2) X_t es integrable.
(3) Si $s \leq t$ entonces $\mathbb{E}(X_t \mid \mathscr{F}_s) = X_s$.

Análogamente se definen las nociones de supermartingala y submartingala al reemplazar la igualdad por $\leq$ y $\geq$ respectivamente.

Considere dos colecciones de variables aleatorias $(X_t, t \geq 0)$ y $(Y_t, t \geq 0)$, decimos que Y es una **modificación** de X si $\mathbb{P}(X_t = Y_t) = 1$ para toda $t \geq 0$.

El ejemplo clásico de un proceso que es modificación de otro es el siguiente: sea U una variable aleatoria uniforme en $(0, 1)$ y $X_t = \mathbf{1}_{U=t}$. Si $Y_t = 0$ para toda t, entonces Y es una modificación (con trayectorias continuas) de X. La idea es que una modificación es casi el mismo proceso estocástico; nuestro objetivo será construir modificaciones con mejores características que el proceso original.

Extenderemos ahora la noción de cantidad de cruces de una función $f : [0, \infty) \to \mathbb{R}$: recordemos que si $F \subset [0, \infty)$ es finito, ya tenemos definida la noción de la cantidad de cruces hacia arriba de $(f(t))_{t\in F}$ en el intervalo $[a, b]$, llamémosle $U_F(f, a, b)$. Si $T \subset \mathbb{R}$ es arbitrario, podemos definir

$$U_T(f, a, b) = \sup_{F\subset T, F \text{ finito}} U_F(f, a, b) .$$

Es claro que si T es numerable y X es un proceso estocástico entonces $U_T(X, a, b)$ es una variable aleatoria. Por otra parte, si $T = [u, v] \cap \mathbb{Q}$, entonces para todo $t \in [u, v]$ existen los límites

$$f(t+) = \lim_{s\downarrow t, s\in T} \qquad t \in [u, v)$$

y

$$f(t-) = \lim_{s\downarrow t, s\in T} \qquad t \in (u, v]$$

si y sólo si $U_T(f, a, b) < \infty$ para cualquier pareja de racionales $a < b$. En este caso, si f es acotada entonces los límites por la derecha y por la izquierda son finitos.

Teorema 1.16 (Desigualdad de cruces de Doob). *Si $(X_t)_{t\geq 0}$ es una $(\mathscr{F}_t)$-supermartingala y $T \subset [0, \infty)$ es numerable entonces*

$$\mathbb{E}(U_T(X)) \leq \sup_{t\in T} \mathbb{E}\big((a - X_t)^-\big) .$$

El teorema anterior se sigue de la desigualdad de cruces de Doob que ya demostramos al tomar supremos. Nuestro objetivo ahora será demostrar la siguiente proposición.

Teorema 1.17. *Sea $(X_t, t \geq 0)$ una martingala respecto a una filtración continua por la derecha y completa $(\mathscr{F}_t, t \geq 0)$. Entonces existe una modificación Y de X que también es una martingala respecto de $(\mathscr{F}_t, t \geq 0)$ y tal que Y tiene trayectorias càdlàg casi seguramente.*

Demostración. Veamos primero que $\sup_{t\in[0,n]\cap\mathbb{Q}} |X_t| < \infty$ casi seguramente. En efecto, al pasar al límite (sobre conjuntos finitos que vayan creciendo a $[0, n]\cap\mathbb{Q}$, la desigualdad maximal de Doob nos dice que

$$\mathbb{P}\left(\sup_{t\in[0,n]\cap\mathbb{Q}} |X_s| = \infty\right) = \lim_{\lambda\to\infty} \mathbb{P}\left(\sup_{t\in[0,n]\cap\mathbb{Q}} |X_s| > \lambda\right) \leq \lim_{\lambda\to\infty} \frac{\mathbb{E}(|X_n|)}{\lambda} = 0.$$

Para cualquier $n \in \mathbb{N}$ y $a < b$, la desigualdad de cruces de Doob nos dice que

$$\mathbb{E}\big(U_{[0,n]\cap\mathbb{Q}}(X, a, b)\big) \leq |a| + \mathbb{E}(|X_n|) < \infty,$$

por lo cual

$$\mathbb{P}\big(U_{[0,n]\cap\mathbb{Q}}(X, a, b) < \infty\big) = 1.$$

Por σsubaditividad, vemos que

$$\mathbb{P}\big(U_{[0,n]\cap\mathbb{Q}}(X, a, b) < \infty \text{ si } a, b \in \mathbb{Q},\ a < b \text{ y } n \in \mathbb{N}\big) = 1.$$

En dicho conjunto, que denotaremos por N^c, X admite límites por la izquierda y por la derecha en t para todo $t \geq 0$, mismos que son finitos, y por lo tanto podemos definir a

$$\tilde{X}_t(\omega) = \begin{cases} X_{t+}(\omega) & \text{si } \omega \in N^c \\ 0 & \text{si } \omega \in N \end{cases}.$$

Como X_{t+} es $\mathscr{F}_{t+}$-medible y $\mathscr{F}_{t+} = \mathscr{F}_t$ entonces X_{t+} es $\mathscr{F}_t$-medible y puesto que N pertenece a $\mathscr{F}_\infty$ y tiene probabilidad cero, entonces $N \in \mathscr{F}_t$ y por lo tanto $\tilde{X}_t$ es $\mathscr{F}_t$-medible. Además, $\tilde{X}$ es continuo por la derecha en N^c por el argumento siguiente: si $\varepsilon > 0$ entonces existe $\delta > 0$ tal que si $r \in [t, t+\delta] \cap \mathbb{Q}$ y $\omega \in N^c$ entonces $\left|\tilde{X}_t(\omega) - X_r(\omega)\right| < \varepsilon$; al tomar límite conforme $r \to s \in [t, t+\delta]$, vemos que $\left|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)\right| \leq \varepsilon$. Una argumento análogo muestra que $\tilde{X}$ admite límites por la izquierda en N^c.

Si $t_1 < t_2$ y s_n es una sucesión de racionales que decrecen a t_1, sabemos que

$$\mathbb{E}(X_{t_2} \mid \mathscr{F}_{s_n}) = X_{s_n}$$

y por el teorema de convergencia de Lévy hacia abajo, vemos que casi seguramente y en L_1:

$$\tilde{X}_{t_1} = \lim_{n\to\infty} X_{s_n} = \lim_{n\to\infty} \mathbb{E}(X_{t_2} \mid \mathscr{F}_{s_n}) = \mathbb{E}(X_{t_2} \mid \mathscr{F}_{t_1+}) = \mathbb{E}(X_{t_2} \mid \mathscr{F}_{t_1}) = X_{t_1},$$

por lo que $\tilde{X}$ es una modificación de X.

Consideremos ahora $t_1 < t_2$ y s_n una sucesión de racionales que decrezcan a t_2. Puesto que X_{s_n} converge casi seguramente y en L_1 a $\tilde{X}_{t_2}$, como vimos en el párrafo anterior, el teorema de convergencia dominada para la esperanza condicional nos dice que

$$\tilde{X}_{s_1} = \mathbb{E}(X_{s_n} \mid \mathscr{F}_{s_1}) \to \mathbb{E}\Big(\tilde{X}_{t_2} \,\Big|\, \mathscr{F}_{s_1}\Big),$$

por lo que $\tilde{X}$ es una $\mathscr{F}_t$-martingala. □

8. Martingalas y sumas de variables aleatorias independientes

En esta sección, presentaremos dos aplicaciones del teorema de convergencia casi segura de martingalas a la convergencia de series aleatorias. Como mencionamos anteriormente, las martingalas tienen muchas aplicaciones teóricas y los siguientes ejemplos presentan una pequeña muestra de dicha aseveración.

TEOREMA 1.18. *Sean $(\xi_i)_{i=1}^\infty$ una sucesión de variables aleatorias independientes tales que $\mathbb{E}(\xi_i) = 0$ y $\sum_{i=1}^\infty \mathrm{Var}(\xi_i) < \infty$. Si $X_n = \sum_{i=1}^n \xi_i$ para $n \geq 1$, entonces $(X_n)_{n=1}^\infty$ converge casi seguramente y en L_2.*

arreglar

DEMOSTRACIÓN. Sea $\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. Hemos visto anteriormente que $(X_n)_{n=1}^\infty$ es una martingala respecto a $(\mathscr{F}_n)_{n=1}^\infty$. Por otra parte, el segundo momento de X_n es su varianza, por lo que

$$\mathbb{E}\big(X_n^2\big) = \sum_{i=1}^n \mathrm{Var}(\xi_i) \leq \sum_{i=1}^\infty \mathrm{Var}(\xi_i) < \infty,$$

por lo que el teorema de convergencia de martingalas acotadas en L_2 nos permite concluir. □

La siguiente aplicación resulta sorprendente pues nos dice que para sumas parciales de variables aleatorias independientes, son equivalentes 3 modos de convergencia que usualmente no tienen porque serlo.

TEOREMA 1.19. *Sean $(\xi_i)_{i=1}^\infty$ variables aleatorias independientes y $X_n = \sum_{i=1}^n \xi_i$ para $n \geq 1$. Entonces son equivalentes*

(1) *$(X_n)_{n=1}^\infty$ converge casi seguramente.*
(2) *$(X_n)_{n=1}^\infty$ converge en probabilidad.*
(3) *$(X_n)_{n=1}^\infty$ converge en distribución.*

DEMOSTRACIÓN. Sabemos que en general $1 \Rightarrow 2 \Rightarrow 3$, por lo que solamente debemos demostrar que $3 \Rightarrow 1$.

Supongamos que $(X_n)_{n=1}^\infty$ converge en distrbución. Sea $\psi_n : \mathbb{R} \to \mathbb{C}$ la función característica de X_n, dada por

$$\psi_n(t) = \mathbb{E}\big(e^{itX_n}\big).$$

La hipótesis de independencia de las variables implica que

$$\psi_{n+1}(t) = \psi_n(t)\, \mathbb{E}\big(e^{it\xi_{n+1}}\big)\,. \tag{2}$$

Además, la hipótesis sobre la convergencia en distribución de $(X_n)_{n=1}^{\infty}$ implica que ψ_n converge puntualmente a una función ψ que es una función característica, por lo que ψ es continua en 0 y como $\psi(0) = 1$, entonces para cada $\delta \in (0,1)$ existe $\alpha > 0$ tal que si $|t| \leq \alpha$ entonces $|\psi(t)| > \delta$. Esto nos dice que para cada $t \in [-\alpha, \alpha]$ existe $N(t) \in \mathbb{N}$ tal que para toda $n \geq N(t)$, $|\psi_n(t)| > \delta > 0$. Así, estamos en posición de considerar a las variables

$$Y_n^t = \frac{e^{itX_n}}{\psi_n(t)}, \quad n \geq N(t)\,, t \in [-\alpha, \alpha],$$

que satisfacen

$$\left|Y_n^t\right| = \frac{\left|e^{itX_n}\right|}{|\psi_n(t)|} < \frac{1}{\delta}.$$

Si definimos $\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, dicha sucesión de variables aleatorias es una martingala compleja[6] respecto a $(\mathscr{F}_n)_{n \geq N(t)}$ ya que si $n \geq N(t)$ entonces al utilizar (2) y la definición de Y_n^t,

$$\mathbb{E}\big(Y_{n+1}^t \,\big|\, \mathscr{F}_n\big) = Y_n^t \mathbb{E}\left(\frac{e^{it\xi_{n+1}}}{\mathbb{E}(e^{it\xi_{n+1}})} \,\middle|\, \mathscr{F}_n \right)$$

y dado que ξ_{n+1} es independiende de $\mathscr{F}_n$, obtenemos

$$\mathbb{E}\big(Y_{n+1}^t \,\big|\, \mathscr{F}_n\big) = Y_n^t \mathbb{E}\left(\frac{e^{it\xi_{n+1}}}{\mathbb{E}(e^{it\xi_{n+1}})} \right) = Y_n^t.$$

Como la sucesión de partes reales y la de partes imaginarias están acotadas en valor absoluto por $1/\delta$, podemos utilizar el teorema de convergencia casi segura para martingalas y concluir que $(Y_n^t)_{n \geq N(t)}$ converge casi seguramente a una variable aleatoria Y^t. Por otra parte, como $\psi_n \to \psi$ puntualmente, se sigue que

$$e^{itX_n} = Y_n^t \psi_n(t) \to Y^t \psi(t) \quad \mathbb{P} - p.s. \quad \forall t \in [-\alpha, \alpha]. \tag{3}$$

Esto significa que para cada $t \in [-\alpha, \alpha]$ existe $\Omega_t \in \mathscr{F}$ tal que $\mathbb{P}(\Omega_t) = 1$ y $(\exp itX_n(\omega))_{n=1}^{\infty}$ converge para toda $\omega \in \Omega_t$. De hecho, podemos mejorar un poco la anterior afirmación: Antes que nada, notemos que si $t \in \mathbb{R}$, entonces existe $m \in \mathbb{N}$ tal que $t/m \in [-\alpha, \alpha]$, por lo que $(\exp iX_n(\omega)\, t/m)_{n=1}^{\infty}$ converge casi seguramente y por lo tanto, $(\exp itX_n(\omega))_{n=1}^{\infty}$ converge casi seguramente. Por otro lado, sea

$$A = \left\{ (t, \omega) \in \mathbb{R} \times \Omega : \left(e^{itX_n(\omega)} \right) \text{ no converge} \right\}.$$

[6] Esto quiere decir una sucesión de variables aleatorias con valores en $\mathbb{C}$ que satisface las tres propiedades que definen a las martingalas reales. De manera alternativa, una sucesión de variables aleatorias complejas para la cual las sucesiónes que constan de las partes reales y la de las partes imaginarias respectivamente son martingalas.

Como la asignaciónes

$$(t,\omega) \mapsto (t, X_n(\omega)), (s,t) \mapsto st \text{ y } t \mapsto e^{it}$$

son

$$(\mathscr{B}_{\mathbb{R}} \otimes \mathscr{F}, \mathscr{B}_{\mathbb{R}^2}), (\mathscr{B}_{\mathbb{R}^2}, \mathscr{B}_{\mathbb{R}}) \text{ y } (\mathscr{B}_{\mathbb{R}}, \mathscr{B}_{\mathbb{C}}) - \textit{medibles}$$

respectivamente, entonces $H_n : \mathbb{R} \times \Omega \to \mathbb{C}$ dada por $H_n(t,\omega) = \exp itX_n(\omega)$ es medible como composición de las anteriores asignaciones. Como A es el conjunto en el cual la sucesión de funciones medibles $(H_n)_{n=1}^{\infty}$ es convergente, se sigue que pertenece a $\mathscr{B}_{\mathbb{R}} \otimes \mathscr{F}$. Por lo que demostramos anteriormente, obtenemos

$$\mathbb{E}(\mathbf{1}_A(t,\omega)) = 0 \quad \forall t \in \mathbb{R},$$

por lo cual

$$\int_{\mathbb{R}} \mathbb{E}(\mathbf{1}_A(t,\omega)) \, d\lambda = 0,$$

donde λ es la medida de Lebesgue. Por el teorema de Tonelli,

$$\mathbb{E}\left(\int_{\mathbb{R}} \mathbf{1}_A(t,\omega)\right) = 0$$

y dado que la asignación

$$\omega \mapsto \int_{\mathbb{R}} \mathbf{1}_A(t,\omega) \, d\lambda$$

es no negativa, entonces

$$\int_{\mathbb{R}} \mathbf{1}_A(t,\omega) \, d\lambda = 0 \quad \mathbb{P} - \text{c.s.}.$$

Esto quiere decir que existe un conjunto $\Omega' \in \mathscr{F}$ tal que $\mathbb{P}(\Omega') = 1$ y para todo $\omega \in \Omega'$, existe un conjunto $S_\omega \in \mathscr{B}_{\mathbb{R}}$ tal que $\lambda(S_\omega) = 0$ y $(\exp itX_n)_{n=1}^{\infty}$ converge para toda $t \in \mathbb{R} \setminus S_\omega$.

Estamos ahora en posición de extraer un límite casi seguro para $(X_n)_{n=1}^{\infty}$ mediante el uso del siguiente lema:

LEMA 3. *Sea $(a_n)_{n\in\mathbb{N}}$ una sucesión de reales tal que $(\exp ita_n)_{n\in\mathbb{N}}$ es convergente para toda $t \in \mathbb{R} \setminus S$, donde $\lambda(S) = 0$. Entonces $(a_n)_{n\in\mathbb{N}}$ es convergente.*

DEMOSTRACIÓN. Aunque el anterior resultado se puede probar mediante técnicas analítcas, a continuación presentamos una prueba probabilística; sin embargo, probaremos que si $(a_{n_k})_{k\in\mathbb{N}}$ y $(a_{m_k})_{k\in\mathbb{N}}$ son dos subsucesiones de $(a_n)_{n\in\mathbb{N}}$ entonces $\lim_{k\to\infty} a_{n_k} - a_{m_k} = 0$. Queda como ejercicio al lector verificar que esto implica la convergencia de la sucesión $(a_n)_{n\in\mathbb{N}}$.

Sea U una variable aleatoria condistribución uniforme en $(0,1)$ definida en un espacio de probabilidad $(\Omega', \mathscr{F}', \mathbb{P}')$. Esto quiere decir que

$$\mathbb{P}'(U \in A) = \lambda(A \cap (0,1)), \quad \forall A \in \mathscr{B}_{\mathbb{R}}$$

donde λ es la medida de Lebesgue. Notemos que si $\omega \in \Omega \setminus U^{-1}(S/t)$, donde $(\exp isa_n)_{n\in\mathbb{N}}$ converge para toda s en el complemento de S y $\lambda(S) = 0$, entonces $(\exp ita_nU)_{n\in\mathbb{N}}$ converge y como

$$\mathbb{P}'\big(U^{-1}(S/t)\big) = \lambda(S/t) = \lambda(S)\,/t = 0$$

entonces $(\exp ita_nU)_{n\in\mathbb{N}}$ converge casi seguramente para toda $t \in \mathbb{R}$. Esto nos dice que $(\exp it\,(a_{n_k} - a_{m_k})\,U)_{k\in\mathbb{N}}$ converge a 1 casi seguramente por lo que el teorema de convergencia acotada nos permite concluir que

$$\mathbb{E}\Big(e^{it\left(a_{n_k}-a_{m_k}\right)U}\Big) \to 1 = \mathbb{E}\big(e^{it0}\big)\,,$$

de donde $(a_{n_k} - a_{m_k})\,U$ converge en distribución a 0 y por lo tanto $|a_{n_k} - a_{m_k}|\,U$ también. Sea $\varepsilon > 0$, por lo que ε es un punto de continuidad de la distribución de la variable aleatoria degenerada que toma el valor cero. Como dicha distribución le asigna el valor 1 a $\varepsilon > 0$ entonces la distribución de $|a_{n_k} - a_{m_k}|\,U$ valuada en ε converge a 1, por lo que podemos afirmar la existencia de $K \in \mathbb{N}$ tal que para toda $k \geq K$, se tiene que

$$\mathbb{P}(|a_{n_k} - a_{m_k}|\,U < \varepsilon) > 1/2.$$

Como la distribución de $|a_{n_k} - a_{m_k}|\,U$ valuada en $|a_{n_k} - a_{m_k}|\,/2$ es igual a $1/2$ cuando $|a_{n_k} - a_{m_k}| > 0$, se sigue que $|a_{n_k} - a_{m_k}|\,/2 < \varepsilon$ para $k \geq K$, por lo que $|a_{n_k} - a_{m_k}| \to 0$. □

Al utilizar el lema anterior, vemos que para cada $\omega \in \Omega'$, $(X_n(\omega))_{n=1}^{\infty}$ es convergente y por lo tanto $(X_n)_{n=1}^{\infty}$ converge casi seguramente. □

CAPÍTULO 2

Cadenas de Markov

En este capítulo se estudiará una clase de procesos estocásticos a tiempo y espacio discreto conocodos como cadenas de Markov. Se estudiarán aspectos asintóticos (a tiempos grandes) de estos procesos así como su liga íntima con las relaciones de recurrencia.

Un procesos estocástico es simplemente una colección de variables aleatorias indexadas usualmente por un conjunto ordenado. En nuestro caso, indexaremos a las variables ya sea con $\mathbb{N}$ (o por algúno de sus subconjuntos), que es cuando diremos que se trata de un proceso a tiempo discreto, o por $[0,\infty)$ o alguno de sus subconjuntos (y hablaremos entonces de un proceso a tiempo continuo). Conenzaremos con la construcción básica de una sucesión de variables aleatorias independientes para continuar con la construcción de cadenas de Markov y más generalmente, con una versión a tiempo discreto del teorema de Kolmogorov

1. Una sucesión de variables aleatorias Bernoulli independientes

Primero se ejemplificará la construcción de una sucesión de variables aleatorias independientes a partir de una sola variable.

Consideremos al espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ en el que $\Omega = (0,1]$, $\mathscr{F} = \mathscr{B}_{(0,1]}$ y $\mathbb{P}$ es la medida de Lebesgue restringida a Ω. Definamos a las variables aleatorias $d_n : \Omega \to \mathbb{R}$ como sigue: a cada $\omega \in \Omega$, se le puede asignar su expansión diádica con colas infinitas:

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n},$$

donde cada d_n es cero o uno. Aunque la expansión diádica todavía no esté bien definida, puesto que por ejemplo a $1/2$ se le podría asociar ya sea $(1,0,0,\ldots)$ o $(0,1,1,\ldots)$, la expansión diádica con colas infinitas, que es la segunda en nuestro ejemplo, sí lo está. Más formalmente, definamos

$$d_1(\omega) = \begin{cases} 0 & \text{si } \omega \in (0,1/2] \\ 1 & \text{si } \omega \in (1/2,1] \end{cases}.$$

Notemos que si $\omega_1 = 2\omega - d_1(\omega)$, entonces $\omega_1 \in (0,1]$; recursivamente, definimos

$$d_{n+1}(\omega) = d_1(\omega_n) \quad \text{y} \quad \omega_{n+1} = 2\omega_n - d_1(\omega_n) \in (0,1].$$

Se prueba entonces por inducción que $d_1(\omega), \ldots, d_n(\omega)$ son los únicos elementos de $\{0,1\}$ tales que

$$\sum_{n=1}^{n} \frac{d_i(\omega)}{2^n} < \omega \leq \sum_{n=1}^{n} \frac{d_i(\omega)}{2^n} + \frac{1}{2^n},$$

por lo que

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n}. \tag{4}$$

Además, la sucesión $(d_n(\omega))_{n\geq 1}$ no tiene una cola de ceros. La unicidad es evidente pues los intervalos $[\sum_{i=1}^{n} u_i/2^n, \sum_{i=1}^{n} u_i/2^n + 1/2^n)$ son ajenos si los u_i son diferentes. Por otra parte, las cotas para ω se siguen de notar que

$$\omega_n = 2^n \left(\omega - \sum_{n=1}^{n} \frac{d_i(\omega)}{2^i}\right),$$

puesto que $\omega_n \in (0,1]$. La fórmula anterior es cierta para $n=1$ y si es válida para n entonces:

$$\begin{aligned}
\omega_{n+1} &= 2\omega_n - d_{n+1}(\omega) \\
&= 2^{n+1}\left(\omega - \sum_{i=1}^{n} \frac{d_i(\omega)}{2^i}\right) - d_{n+1}(\omega) \\
&= 2^{n+1}\left(\omega - \sum_{n=1}^{n+1} \frac{d_i(\omega)}{2^i}\right).
\end{aligned}$$

Afirmamos que $(d_n)_{n\geq 1}$ son variables aleatorias independientes e idénticamente con distribución Bernoulli de parámetro $1/2$. Para verificar esto, notemos que si $u_1, \ldots, u_n \in \{0,1\}$ y $d_1(\omega) = u_1, \ldots, d_n(\omega) = u_n$ entonces

$$\sum_{i=1}^{n} \frac{u_i}{2^n} < \omega \leq \sum_{i=1}^{n} \frac{u_i}{2^n} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \sum_{i=1}^{n} \frac{u_i}{2^n} + \frac{1}{2^n}.$$

La aserción recíproca es también verdadera por la propiedad carácterística de $d_1, \ldots, d_n$; se tiene entonces la igualdad entre conjuntos

$$\{\omega \in \Omega : d_1(\omega) = u_1, \ldots, d_n(\omega) = u_n\} = (\sum_{i=1}^{n} \frac{u_i}{2^n}, \sum_{i=1}^{n} \frac{u_i}{2^n} + \frac{1}{2^n}]$$

y es claro que el conjunto del lado derecho pertenece a $\mathscr{F}$. Por lo tanto las $d_i, i \geq 1$ son variables aleatoras y

$$\mathbb{P}(d_1 = u_1, \ldots, d_n = u_n) = \frac{1}{2^n},$$

por lo cual son independientes y de distribución Bernoulli de parámetro $1/2$.

2. El espacio canónico

Consideraremos ahora otro punto de vista para analizar la construcción de una sucesión de variables aleatorias independientes con Bernoulli de parámetro $1/2$.

Sea $\Omega = \{0,1\}^{\mathbb{N}}$ y definamos a $B_i : \Omega \to \{0,1\}$ por medio de

$$B_i(\omega) = \omega_i \quad \text{si} \quad \omega = (\omega_i, i \in \mathbb{N}) .$$

También consideramos a

$$\mathscr{F}_n = \sigma(B_0, \ldots, B_n) ,$$

que consta precisamente de los conjuntos de la forma $A \times \Omega$ donde $A \subset \{0,1\}^{n+1}$, y a

$$\mathscr{F} = \sigma(\mathscr{A}) \quad \text{donde} \quad \mathscr{A} = \bigcup_n \mathscr{F}_n .$$

Para constuir una sucesión de variables aleatorias Bernoulli independientes, podríamos intentar construir una medida de probabilidad $\mathbb{P}$ sobre $(\Omega, \mathscr{F})$ tal que para todo $A \subset \{0,1\}^{n+1}$ tengamos que

$$\mathbb{P}(A \times \Omega) = \frac{|A|}{2^{n+1}}, \tag{5}$$

donde $|A|$ denota a la cardinalidad de A.

En efecto, vemos que entonces, para toda n, $B_0, \ldots, B_n$ son variables independientes con distribución Bernoulli de parámetro $1/2$. Notemos que la ecuación (5) nos define a una función $\mathbb{P}$ que no está definida en $\mathscr{F}$ sino tan sólo en $\bigcup_n \mathscr{F}_n$ y que está bien definida, pues si a un mismo conjunto lo escribo como $A \times \Omega$ y como $B \times \Omega$ donde $A \subset \{0,1\}^{m+1}$ y $B \subset \{0,1\}^{n+1}$ con $n < m$ entonces $B = A \times \{0,1\}^{n-m}$ y

$$\frac{|A|}{2^{m+1}} = \frac{|A| 2^{n-m}}{2^{m+1} 2^{n-m}} = \frac{|B|}{2^{n+1}}.$$

Así, la construcción de la medida $\mathbb{P}$ se reduce a un problema de extensión de una función aditiva definida en el álgebra $\mathscr{A}$ a una función σ-aditiva definida en toda la σ-álgebra $\mathscr{F}$. Sabemos que para que la función $\mathbb{P} : \mathscr{A} \to [0,1]$ dada por (5) pueda extenderse a una medida de probabilidad sobre $\mathscr{F}$, sólo debemos verificar que $\mathbb{P}$ sea σ-aditiva sobre $\mathscr{A}$. Comentario personal: me parece un camino complicado.

Para construir la medida $\mathbb{P}$ también se puede razonar como sigue: consideremos la función de $(0,1]$ en Ω que a cada x le asigna su expansión binaria con colas infinitas, digamos $x_1, x_2, \ldots$. Esta función es medible pues su composición con cada B_i resulta en la función $x \mapsto x_i$ que es medible. Sea $\mathbb{P}$ la medida imagen de la medida de Lebesgue bajo esta asignación. Por lo que ya hemos probado en la Seccion 1, bajo la medida $\mathbb{P}$, las proyecciones $(B_i, i \in \mathbb{N})$ son independientes y con distribución Bernoulli de parámetro $1/2$.

En general trabajamos sin especificar al espacio de probabilidad. A veces, sin embargo, resulta conveniente utilizar algún espacio de probabilidad específico que tenga un poco más de estructura. Cuando la elección es natural (como lo es para una sucesión de variables Bernoulli) hablamos del espacio canónico. Un ejemplo de estructura que tenemos en este espacio canónico es la posibilidad de operar con los elementos del espacio de eventos mediante los operadores de traslación: si $\omega \in \Omega$ se escribe como $\omega = (\omega_0, \omega_1, \ldots)$, podemos definir $\theta_i(\omega)$ como $(\omega_i, \omega_{i+1}, \ldots)$. Es fácil ver que la medida $\mathbb{P}$ es la medida imagen de $\mathbb{P}$ bajo θ_i para cualquier i. (Por ejemplo, pues ambas medidas coinciden sobre $\mathscr{A}$.) Esto puede ser el inicio del estudio de propiedades de ergodicidad de la medida $\mathbb{P}$, mediante las cuales se puede obtener una prueba la ley fuerte de los grandes números para estas variables Bernoulli.

3. Una sucesión de variables aleatorias uniformes independientes

Ahora demostraremos que si $(X_i)_{i\geq 1}$ son variables aleatorias independientes con distribución Bernoulli de parámetro $1/2$ (definidas en otro espacio de probabilidad) entonces $U = \sum_{i\geq 1} X_i/2^i$ tiene distribución uniforme. En efecto, puesto que

$$\mathbb{P}(X_1 = u_1, \ldots, X_n = u_n) = \mathbb{P}(d_1 = u_1, \ldots, d_n = u_n)\,,$$

vemos que

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i/2^i < x\right) = \mathbb{P}\left(\sum_{i=1}^{n} d_i/2^i < x\right).$$

Por otro lado,

$$U = \lim_n \sum_{i=1}^{n} X_i/2^i,$$

y de hecho la sucesión de variables aleatorias es creciente. Por lo tanto U es variable aleatoria y

$$\mathbb{P}(U < x) = \lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n} X_i/2^i < x\right) = \lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n} d_i/2^i < x\right) = \mathbb{P}((0, x]) = x.$$

(La penúltima igualdad se sigue de (4).) Así, vemos que U es una variable uniforme.

Ahora utilizaremos lo anterior para mostrar que existe un espacio de probabilidad en el que están definidas una sucesión de variables aleatorias uniformes independientes. De hecho el espacio de probabilidad que consideraremos es el mismo $(\Omega, \mathscr{F}, \mathbb{P})$ que en la Sección 1. Como $\mathbb{Z}_+$ y $\mathbb{Z}_+^2$ tienen la misma cardinalidad, consideremos una biyección de $\phi : \mathbb{Z}_+^2 \to \mathbb{Z}_+$. Definamos $d_i^n = d_{\phi(n,i)}$ y para cada $n \in \mathbb{Z}_+$, sea

$$U_n = \sum_{i\geq 1} \frac{d_i^n}{2^i}.$$

Como $(d_i^n)_{i\geq 1}$ son variables aleatorias independientes de distribución Bernoulli de parámetro $1/2$, se sique que U_n tiene distribución uniforme para cada $n \in \mathbb{N}$. Se afirma ahora que las variables $(U_n)_{n\geq 1}$ son independientes. En efecto, esto es consecuencia del siguiente lema, un tanto más general. Notemos que U_n es medible respecto de la σ-álgebra generada por $(d_i^n)_{i\geq 1}$, a la cual llamaremos $\mathscr{F}_n$.

LEMA 4. *Sean $\mathscr{F}_{n,i}, i \geq 1, n \geq 1$ σ-álgebras independientes y definamos*

$$\mathscr{F}_n = \sigma(\mathscr{F}_{n,1}, \mathscr{F}_{n,2}, \ldots)\,.$$

Entonces $\mathscr{F}_n, n \geq 1$ son σ-álgebras independientes.

DEMOSTRACIÓN. Debemos mostrar que para todo $A_1 \in \mathscr{F}_1, \ldots, A_n \in \mathscr{F}_n$, se tiene que

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)\,. \tag{6}$$

Sea

$$\mathscr{C}_n = \{A_1 \cap \cdots \cap A_m : m \geq 1 \text{ y } A_j \in \cup_{i\geq 1} \mathscr{F}_{n,i} \text{ para } j = 1, \ldots, m\}\,.$$

Puesto que

$$\bigcup_{i\geq 1} \mathscr{F}_{n,i} \subset \mathscr{C}_n \subset \mathscr{F}_n,$$

vemos que

$$\sigma(\mathscr{C}_n) = \mathscr{F}_n.$$

Por otra parte, es fácil ver que $\mathscr{C}_n$ es un π-sistema.

Consideremos ahora la clase

$$\mathscr{M}_1 = \{A \in \mathscr{F}_k : \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \text{ si } B = B_1 \cap \cdots \cap B_n \text{ con } B_j \in \mathscr{C}_j\}\,.$$

Es fácil ver que $\mathscr{M}_k$ es un λ-sistema que contiene, por hipótesis a $\mathscr{C}_1$. Por lo tanto

$$\mathscr{M}_1 = \mathscr{F}_1.$$

Ahora consideramos a

$$\begin{aligned}\mathscr{M}_2 = \{&A \in \mathscr{F}_2 : \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \\ &\text{si } B = B_1 \cap B_3 \cap \cdots \cap B_n \text{ con } B_1 \in \mathscr{F}_1 \text{ y } B_j \in \mathscr{C}_j \text{ para } j \geq 3\}\,.\end{aligned}$$

Se prueba entonces que $\mathscr{M}_2$ es un λ-sistema que por hipótesis y la igualdad $\mathscr{M}_1 = \mathscr{F}_1$ contiene a $\mathscr{C}_2$. Al aplicar este razonamiento sucesivamente, obtenemos la igualdad (6). □

4. Una sucesión de variables aleatorias independientes con distribuciones arbitrarias

Ahora utilizaremos la construcción de la sucesión de variables aleatorias uniformes independientes para demostrar el siguiente resultado:

TEOREMA 2.1. *Sean μ_n, $n \geq 1$ medidas de probabilidad en $\mathbb{R}$. Entonces existe un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ y una sucesión de variables aleatorias independientes $X_n : \Omega \to \mathbb{R}$, $n \geq 1$ tales que la distribución de X_n es μ_n.*

La herramienta principal de la construcción será el siguiente lema: (tomado de Billingsley p. 190). Recordemos que una función $F : \mathbb{R} \to [0,1]$ es la función de distribución de una variable aleatoria real si y sólo si es no decreciente, continua por la derecha (y con límites por la izquierda) tal que $\lim_{x\to-\infty} F(x) = 0$ y $\lim_{x\to\infty} F(x) = 1$.

Definición. La **función de cuantiles** de una función de distribución F es la función $\phi : (0,1) \to \mathbb{R}$ dada por

$$\phi(u) = \inf\left\{x \in \mathbb{R} : u \leq F(x)\right\}.$$

La función de cuantiles satisface la igualdad

$$\phi(u) \leq x \Leftrightarrow u \leq F(x)$$

que se demostrará posteriormente. De esta igualdad se deduce la medibilidad de ϕ.

Prueba del Teorema 2.1. Sabemos que existe un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ en el que existe una sucesión de variables aleatorias independientes $(U_n)_{n\geq 1}$ uniformes en $(0,1)$. Sea F_n la función de distribución asociada a la medida de probabilidad μ_n y ϕ_n la función de cuantiles de $\mathscr{F}_n$. Como ϕ_n es una función medible, $X_n = \phi_n(U_n)$ es una variable aleatoria. Además, como las variables $U_n, \neq 1$ son independientes, también lo son las variables $X_n, n \geq 1$:

$$\{X_1 \in A_1, \ldots, X_n \in A_n\} = \left\{U_1, \in \phi_1^{-1}(A_1), \ldots, U_n, \in \phi_n^{-1}(A_n)\right\}.$$

Finalmente:

$$\mathbb{P}\big(X_i^{-1}((-\infty,x])\big) = \mathbb{P}\big(U_i^{-1}\big(\phi_i^{-1}((-\infty,x])\big)\big) = \mathbb{P}\big(U_i^{-1}((0,F_i(x)])\big) = F_i(x),$$

por lo que X_i tiene distribución μ_i. □

Ahora demostremos las propiedades de la función de cuantiles ϕ asociada a la función de distribución F. Sea $u \in (0,1)$; entonces el conjunto $\{x \in \mathbb{R} : u \leq F(x)\}$ es no vacío y como F es no decreciente, es un intervalo ya sea de la forma $[\phi(u), \infty)$ o $(\phi(u), \infty)$, ya que $\phi(u)$ es el ínfimo del conjunto considerado. La segunda opción se descarta al notar que F es continua por la derecha. Por lo tanto, $u \leq F(x)$ si y sólo si $\phi(u) \leq x$.

5. Cadenas de Markov

Una cadena de Markov es un proceso estocástico de un tipo especial; se encuentra caracterizado por satisfacer la propiedad de Markov. Una manera coloquial de expresar la propiedad de Markov es que el futuro del proceso es independiente del pasado cuando se conoce el presente. Esto se puede traducir matemáticamente mediante el concepto de independencia condicional.

Definición. Sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad y $\mathscr{H}_1, \mathscr{H}_2, \mathscr{G}$ subσ-álgebras de $\mathscr{F}$. Decimos que $\mathscr{H}_1$ y $\mathscr{H}_2$ son condicionalmente independientes dada $\mathscr{G}$, denotado por $\mathscr{H}_1 \perp_{\mathscr{G}} \mathscr{H}_2$, si para cualesquiera variables aleatorias acotadas H_1 y H_2 tal que H_i es $\mathscr{H}_i$-medible se tiene que

$$\mathbb{E}(\,H_1 H_2 \,|\mathscr{G}) = \mathbb{E}(\,H_1 \,|\mathscr{G})\,\mathbb{E}(\,H_2 \,|\mathscr{G})\,.$$

Si $X = (X_n)_{n\in\mathbb{N}}$ es un proceso estocástico, $(\mathscr{F}_n)_{n\in\mathbb{N}}$ es su filtración canónica asociada y

$$\mathscr{F}^n = \sigma(X_n, X_{n+1}, \ldots)\,,$$

diremos que X satisface la propiedad de Markov (inhomogénea) si $\mathscr{F}_n$ y $\mathscr{F}^n$ son condicionalmente independientes dada X_n. A través de una equivalencia del concepto de independencia condicional, podremos obtener la expresión usual de la propiedad de Markov.

Proposición 2.1. *Las subσ-álgebras $\mathscr{H}_1$ y $\mathscr{H}_2$ son condicionalmente independientes dada $\mathscr{G}$ si y sólo si para cualquier variable aleatoria H_1 que sea $\mathscr{H}_1$-medible y acotada:*

$$\mathbb{E}(\,H_1 \,|\mathscr{G}, \mathscr{H}_1) = \mathbb{E}(\,H_1 \,|\mathscr{G})\,.$$

Ejercicio 2.1. Probar la proposición anterior.

Lo curioso de la anterior proposición es que establece la equivalencia entre una propiedad a priori asimétrica y la propiedad simétrica de independencia condicional. Vemos por lo tanto que un proceso X satisface la propiedad de Markov si para cualquier variable $\mathscr{F}^n$ medible y acotada F se tiene que

$$\mathbb{E}(\,F \,|\,\mathscr{F}_n, X_n) = \mathbb{E}(\,F \,|\,X_n)\,.$$

Supongamos ahora que X_n toma valores en un conjunto a lo más numerable E. Entonces la propiedad de Markov se puede escribir de una manera distinta. Notemos que la familia

$$\mathscr{C}_n = \{\{X_0 = i_0, \ldots, X_n = i_n\} : i_0, \ldots, i_n \in E\}$$

es un π-sistema que genera a $\mathscr{F}_n$ mientras que la familia

$$\mathscr{C}^n = \{\{X_n = i_n, \ldots, X_{n+m} = i_{n+m}\} : m \geq 0 \text{ y } i_n, \ldots, i_{n+m} \in E\}$$

es un π-sistema que genera a $\mathscr{F}^n$.

Proposición 2.2. *Un proceso estocástico $X = (X_n)_{n\in\mathbb{N}}$ con valores en un conjunto a lo más numerable E satisface la propiedad de Markov si y sólo si cada vez que $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) > 0$, se tiene que*

$$\mathbb{P}(X_{n+1} = i_{n+1}\,|X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1}\,|X_n = i_n)\,. \tag{7}$$

Demostración. Recordemos que, puesto que X_n toma valores en el conjunto a lo más numerable E entonces

$$\mathbb{P}(A \,|X_n) = \sum_{i_n \in E} \mathbb{P}(A \,|X_n = i_n)\, \mathbf{1}_{X_n = i_n}.$$

Si X satisface la propiedad de Markov, notamos que al ser $\mathbf{1}_{X_{n+1}=i_{n+1}}$ $\mathscr{F}^n$-medible, tenemos que para todo $A \in \mathscr{C}_n$

$$\mathbb{E}\big(\mathbf{1}_{X_{n+1}=i_{n+1}} \mathbf{1}_A \mathbf{1}_{X_n=i_n}\big) = \mathbb{P}(X_{n+1} = i_{n+1} \,|X_n = i_n)\, \mathbb{P}(A, X_n = i_n)\,.$$

Por lo tanto, obtenemos (7).

Por otra parte, supongamos que X es un proceso estocástico que satisface (7). Sea

$$\mathscr{M}^n = \{B \in \mathscr{F}^n : \mathbb{P}(A \cap \{X_n = i_n\} \cap B) = \mathbb{P}(A \cap \{X_n = i_n\})\, \mathbb{P}(B \,|X_n = i_n) \quad \text{para todo } A \in \mathscr{C}_n\}\,.$$

Entonces $\mathscr{M}^n$ es un λ-sistema que contiene al π-sistema $\mathscr{C}^n$, por lo que es igual a $\mathscr{F}^n$. Podemos entonces definir a

$$\mathscr{M}_n = \{A \in \mathscr{F}_n : \mathbb{P}(A \cap \{X_n = i_n\} \cap B) = \mathbb{P}(A \cap \{X_n = i_n\})\, \mathbb{P}(B \,|X_n = i_n) \quad \text{para todo } B \in \mathscr{F}^n\}\,,$$

que es un λ-sistema que contiene a $\mathscr{C}_n$ y que por lo tanto contiene a $\mathscr{F}_n$. Se deduce que X satisface la propiedad de Markov. □

Una **cadena de Markov** (inhomogénea, a tiempo y espacio discreto) es un proceso estocástico $X = (X_n)_{n\in\mathbb{N}}$ con valores en un conjunto a lo más numerable E que satisface la propiedad de Markov. Si $\mathbb{P}(X_n = i_n) > 0$ para toda n, escribamos

$$P^n_{i_n,i_{n+1}} = \mathbb{P}(X_{n+1} = i_{n+1} \,|X_n = i_n)\,.$$

La propiedad de Markov implica, a través de una utilización iterada de (7) que

$$\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = P^0_{i_0,i_1} \cdots P^{n-1}_{i_{n-1},i_n}.$$

Diremos que la cadena de Markov es **homogénea** si $P^1_{i,j} = P^n_{i,j}$ para cualquier n; en este caso escribimos simplemente $P_{i,j}$ y se satisface:

$$\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = P_{i_0,i_1} \cdots P_{i_{n-1},i_n}.$$

A la colección numérica $(P_{i,j})_{i,j\in E}$, que usualmente nos imaginamos como una matriz cuyas entradas están indexadas por E, se le conoce como **matriz de transición**. Esta colección satisface:

$$P_{i,j} \geq 0 \quad \text{y} \quad \sum_{j\in E} P_{i,j} = 1.$$

Vale la pena preguntarse si dada cualquier colección numérica P que satisface lo anterior, le podemos asignar una cadena de Markov con matriz de transición P. El teorema de consistencia de Kolmogorov nos permite responder afirmativamente; sin embargo, daremos la idea de la prueba del teorema de consistencia de Kolmogorov al demostrar directamente la existencia de las cadenas de Markov.

6. Construcción de una cadena de Markov

Aplicaremos ahora la idea detrás de la función de cuantiles para dar una construcción de una cadena de Markov asociada a una matriz de transición dada.

Una cadena de Markov con espacio discreto es un proceso estocástico $(X_n)_{n\in\mathbb{N}}$, donde X_n toma valores en un conjunto E a lo más numerable (dotado de la σ-álgebra total); dicha cadena está especificada por dos parámetros: la distribución inicial y la matriz de transición. Supondremos que $E = \{1, \ldots, \mathbb{N}\}$ o que $E = \mathbb{N}$, para simplificar los argumentos. Sea π una medida de probabilidad en E, a la cual llamaremos distribución inicial, y $P = (P_{i,j}, i, j \in \mathbb{N})$ una colección de reales no-negativos tal que

$$\sum_{j\in E} P_{i,j} = 1.$$

A P la imaginamos como conformando una matriz (posiblemente infinita) a la cual llamamos matriz de transición.

Definición. Una cadena de Markov con distribución inicial π y matriz de transición P es un proceso estocástico $(X_n)_{n\in\mathbb{N}}$ con valores en $\mathbb{N}$ tal que para $i_0, \ldots, i_n \in E$ se tiene que

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \nu(\{i_0\})\, P_{i_0,i_1} \cdots P_{i_{n-1},i_n}.$$

Para construir a las cadenas de Markov, sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad en el que están definidas una sucesión de variables aleatorias uniformes independientes $(U_n, n \in \mathbb{N})$. Sea ϕ_i la función de cuantiles asociada a la medida de probabilidad que asigna masa $P_{i,j}$ a cada $j \in E$, ϕ la función de cuantiles asociada a π y consideremos a la sucesión de variables aleatorias (Y_n) donde

$$Y_0 = \phi(U_0) \quad \text{y} \quad Y_{n+1} = \phi_{Y_n}(U_{n+1})\,.$$

Por inducción es fácil ver que Y_n es medible respecto de $\mathscr{F}_n = \sigma(U_0, \ldots, U_n)$. Se tiene entonces que

$$\begin{aligned}
\mathbb{P}(Y_0 = i_0, \ldots, Y_n = i_n) &= \mathbb{P}\big(\phi(U_0) = i_0, \phi_{i_0}(U_1) = i_1, \ldots, \phi_{i_{n-1}}(U_n) = i_n\big) \\
&= \mathbb{P}(\phi(U_0) = i_0) \prod_{i=1}^{n} \mathbb{P}\big(\phi_{i_{j-1}}(U_j) = i_j\big) \\
&= \nu(\{i_0\})\, P_{i_0,i_1} \cdots P_{i_{n-1},i_n}.
\end{aligned}$$

Consideremos a la función que asigna a ω la sucesión $(Y_n(\omega)\,, n \in \mathbb{N})$. Dicha sucesión es medible si en $E^{\mathbb{N}}$ utilizamos la σ-álgebra generada por las proyecciones X_i (que a $(e_j, j \in \mathbb{N})$ le asocian e_i. Denotaremos por $\mathbb{P}_\nu$ a la medida imagen bajo dicha aplicación y notamos que bajo ella, el proceso canónico X es una cadena de

Markov con distribución inicial ν y matriz de transición P. Notemos que si δ_i es la masa puntual en i, entonces podemos escribir

$$\mathbb{P}_\nu = \sum_{e \in E} \nu(\{e\})\, \mathbb{P}_{\delta_e}.$$

A veces se escribe simplemente $\mathbb{P}_e$ en vez de $\mathbb{P}_{\delta_e}$.

7. Ejemplos de cadenas de Markov

Ejemplo 2.1 (Caminatas aleatorias discretas). Sea μ una medida de probabilidad concentrada en un conjunto a lo más numerable A de $\mathbb{Z}^n$. Si $P_{i,j} = \mu(j-i)$, la cadena de Markov resultante se conoce como caminata aleatoria de distribución de salto μ.

El caso particular en el que μ está concentrada en los vecinos de cero dentro de la retícula $\mathbb{Z}^n$ (por ejemplo $\{-1,1\}$ en dimensión 1 ó $\{(1,0)\,,(0,1)\,,(-1,0)\,,(0,-1)\}$ en dimensión 2) hablamos de una caminata aleatoria simple. Ésta será simétrica si a cada vecino se le asigna la misma probabilidad.

Ejemplo 2.2 (La cadena de nacimiento y muerte). Son cadenas de Markov en $\mathbb{N}$ tales que $P_{i,i+1} = p_i$ y $P_{i,i-1} = 1 - p_i$ $(= q_i)$ para $i \geq 1$ y tal que $P_{0,0} = 1$. La interpretación de la cadena es que representa el tamaño de una población tal que si hay i individuos, p_i es la probabilidad de que haya un nacimiento antes que una muerte.

Ejemplo 2.3 (Procesos de Galton-Watson). Si μ es una distribución en $\mathbb{N}$, el proceso de Galton-Watson con distribución de progenie μ es una cadena de Markov con espacio de estados $\mathbb{N}$ y probabilidades de transición $P_{i,j} = \mu^{*i}(j)$, donde $\mu^{*0} = \delta_0$ y para $i \geq 1$, μ^{*i} es la i-ésima convolución de μ consigo misma. En otras palabras:

$$\mu^{*i}(j) = \sum_{l_1 + \cdots + l_i = j} \mu(l_1) \cdots \mu(l_i)\,.$$

Este proceso puede ser realizado en el espacio canónico como ya hemos hecho o mediante la construcción usual: si $\xi_{n,i}$ son variables aleatorias independientes con distribución μ, y definimos

$$Z_0 = i \quad \text{y} \quad Z_{n+1} = \sum_{1 \leq i \leq Z_n} \xi_{n,i}. \tag{8}$$

El proceso de Galton-Watson de ley de reproducción μ tiene una propiedad especial que lo caracteriza dentro de la clase de cadenas de Markov: la propiedad de ramificación. Esta afirma que la suma de dos procesos de Galton-Watson independientes con la misma ley de reproducción μ es un proceso de Galton-Watson con ley de reproducción μ. La propiedad de ramificación puede probarse a partir de la representación (8). Sin embargo, nosotros utilizaremos el espacio canónico para ilustrarlo.

Sea $(\mathbb{P}_i)_{i\in\mathbb{N}}$ cualquier familia markoviana en el espacio canónico $\mathbb{N}^{\mathbb{N}}$ asociada a la matriz de transición P. La medida de probabilidad producto $\mathbb{P}_{i_1} \otimes \mathbb{P}_{i_2}$ está definida en $(\mathbb{N} \times \mathbb{N})^{\mathbb{N}}$; si (X_n) es el proceso canónico en dicho espacio, se puede escribir como $\left(X^1, X^2\right)$ y notemos que las distribuciones finito-dimensionales están determinadas por

$$\mathbb{P}_{i_0^1} \times \mathbb{P}_{i_0^2}\left(X_1^1 = i_1^1, X_1^2 = i_1^2, \ldots, X_n^1 = i_1^1, X_n^2 = i_n^2\right) = P_{i_0^1, i_1^1} P_{i_0^2, i_1^2} \cdots P_{i_{n-1}^1, i_n^1} P_{i_{n-1}^2, i_n^2};$$

esto significa no sólo que bajo la medida de probabilidad $\mathbb{P}_{i_0^1, i_0^2}$ las coordenadas del proceso canónico X^1 y X^2 son cadenas de Markov independientes, sino que también X es una cadena de Markov con matriz de transición dada por

$$\tilde{P}_{(i_1,i_2),(j_1,j_2)} = P_{i_1,j_1} P_{i_2,j_2}.$$

Ahora definamos $\mathbb{P}_i * \mathbb{P}_j$ como la imagen de $\mathbb{P}_i \otimes \mathbb{P}_j$ bajo $X^1 + X^2 = \left(X_n^1 + X_n^2, n \in \mathbb{N}\right)$. Dedimos que la familia markoviana $(\mathbb{P}_i, i \in \mathbb{N})$ satisface la **propiedad de ramificación** si $\mathbb{P}_i * \mathbb{P}_j = \mathbb{P}_{i+j}$.

Veamos ahora que si $(\mathbb{P}_i, i \in \mathbb{N})$ tiene la propiedad de ramificación entonces es la familia Markoviana asociada a un proceso de Galton-Watson. En efecto, si $\mu(j) = P_{1,j}$, la propiedad de ramificación nos dice inductivamente que $P_{i,j} = \mu^{*i}(j)$.

Ejercicio 2.2. Pruebe que la familia Markoviana asociada a un proceso de Galton-Watson tiene la propiedad de ramificación.

Ejemplo 2.4 (La cadena de Ehrenfest). Consideremos dos urnas con un total de n bolas numeradas del 1 al n. En cada instante de tiempo se escoje aleatoriamente un entero entre 1 y n y se cambia de urna la bola con dicha etiqueta. Si X_n denota la cantidad de bolas en la primera urna al instante n, entonces (X_n) es una cadena de Markov con espacio de estados $E = \{0, \ldots, n\}$ tal que

$$P_{i,i+1} = 1 - \frac{i}{n}, \quad P_{i,i-1} = \frac{i}{n} \quad \text{si } i \in \{1, \ldots, n-1\} \quad P_{0,1} = 1 \quad \text{y} \quad P_{n,n-1} = 1.$$

Ejemplo 2.5 (El modelo de Wright-Fisher). Es un modelo de herencia genética dentro de una población de m individuos con un gen con alelos A y a. Si suponemos que la generación $n+1$ se obtiene de la generación n haciendo que cada individuo escoja (aleatoriamente e independientemente de los otros individuos) a sus progenitores dentro de la generación n heredando un alelo de cada uno y nos interesa la cantidad de alelos a en la población conforme pasa el tiempo (en particular para saber si ocurre una fijación...). Formalmente, se trata de una cadena de Markov con espacio de estados $\{0, \ldots, m\}$ y matriz de transición P dada por

$$P_{i,j} = \binom{m}{j} \left(\frac{i}{m}\right)^j \left(1 - \frac{i}{m}\right)^{m-j}.$$

8. Análisis básico de cadenas de Markov

Sea P una matriz de transición sobre E. Puesto que

$$\mathbb{P}_i(X_n = j) = \sum_{i_1,\ldots,i_{n-1}\in E} \mathbb{P}_i(X_1 = i_1, \ldots, X_{n-1} = i_{n-1}),$$

vemos que

$$\mathbb{P}_i(X_n = j) = \sum_{i_1,\ldots,i_n} P_{i,i_1} \cdots P_{i_{n-1},j}.$$

Por lo tanto, si se introducen a las potencias de la matriz de transición $P^n, n \geq 1$ (y se define $P^0_{i,j} = \delta_{i,j}$) vemos que

$$\mathbb{P}_i(X_n = j) = P^n_{i,j}.$$

Sean i y j dos estados de E. Diremos que i conduce a j si existe $n \geq 0$ tal que $P^n_{i,j} > 0$. Claramente esto ocurre si y sólo si existen $i_0, \ldots, i_n$ con $i_0 = i$ y $i_n = j$ tales que $P_{i_{k-1},i_k} > 0$. Cuando i conduce a j y j conduce a i, diremos que i y j se comunican y lo denotaremos mediante $i \sim j$.

Proposición 2.3. *La relación $i \sim j$ es una relación de equivalencia en E.*

A las clases de equivalencia inducidas por la relación $\sim$ les llamaremos clases de comunicación.

Demostración.

Reflexividad: Puesto que $P^0_{i,i} = 1$, vemos que $i \sim i$.
Simetría: Por definición $i \sim j$ si y sólo si $j \sim i$.
Transitividad: Si $i \sim j$ y $j \sim k$, sean m y n en $\mathbb{N}$ tales que $P^m_{i,j} > 0$ y $P^n_{j,k} > 0$. Puesto que

$$P^{n+m}_{i,k} \geq P^n_{i,j} P^m_{j,k} > 0,$$

vemos que i conduce a k y un argumento análogo muestra que entonces $i \sim k$. □

Se dice que una cadena de Markov es **irreducible** si tiene una sola clase de comunicación.

9. La propiedad de Markov en el espacio canónico

Sea P una matriz de transición en el conjunto a lo más numerable E. Sea $\mathbb{P}_i$ la medida de probabilidad, en el espacio canónico $E^{\mathbb{N}}$, de una cadena de Markov con medida de transición P y distribución inicial δ_i. Ahora veremos una forma distinta de expresar la propiedad de Markov. Sea $(X_n)_{n\in\mathbb{N}}$ el proceso canónico, $\mathscr{F}_n$ la filtración canónica. Definiremos los operadores de traslación $\theta_n : E^{\mathbb{N}} \to E^{\mathbb{N}}$ dados por:

$$\theta_n\big((e_m)_{m\in\mathbb{N}}\big) = (e_{m+n})_{m\in\mathbb{N}}.$$

Notemos que θ_n tiene el efecto de *adelantar el tiempo* pues

$$X_m \circ \theta_n = X_{n+m};$$

así:

$$\theta_n^{-1}(\{X_m \in A\}) = \{X_{n+m\in A}\} .$$

Así, vemos que si $F : E^{\mathbb{N}} \to \mathbb{R}$ es medible, $F \circ \theta_n$ también lo es.

Proposición 2.4 (Propiedad de Markov). *Si $F : E^{\mathbb{N}} \to \mathbb{R}$ es medible y acotada:*

$$\mathbb{E}(F \circ \theta_n \mid X_n) = \mathbb{E}_{X_n}(F) .$$

Demostración. Como E es a lo más numerable

$$\mathbb{E}(F \circ \theta_m \mid X_0, \ldots, X_m) = \sum_{i_0,\ldots,i_m \in E} \mathbb{E}(F \circ \theta_m \mid X_0 = i_0, \ldots, X_m = i_m) \mathbf{1}_{X_0=i_0,\ldots,X_m=i_m} .$$

Por lo tanto, basta mostrar que

$$\mathbb{E}(F \circ \theta_m \mid X_0 = i_0, \ldots, X_m = i_m) = \mathbb{E}_{i_m}(F) .$$

Esto se hace mediante el procedimiento estándar, al probarlo primero para funciones indicadoras de elementos de $\mathscr{F} = \sigma(X_n, n \in \mathbb{N})$. Para esto, primero lo probamos para elementos del π-sistema

$$\mathscr{C} = \{\emptyset\} \cup \{\{X_0 = j_0, \ldots, X_n = j_n\} : n \in \mathbb{N}, j_0, \ldots, j_n \in E\} .$$

Sin embargo:

$$\mathbf{1}_{X_0=j_0,\ldots,X_n=j_n} \circ \theta_m = \{X_n = j_0, \ldots, X_{n+m} = j_n\}$$

por lo que:

$$\begin{aligned}
&\mathbb{E}(\mathbf{1}_{X_0=j_0,\ldots,X_n=j_n} \circ \theta_m \mid X_0 = i_0, \ldots, X_m = i_m) \\
&= \mathbf{1}_{j_0=i_n} P_{i_n,j_1} \cdots P_{j_{n-1},j_n} \\
&= \mathbf{1}_{j_0=i_n} \mathbb{P}_{i_n}(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n) .
\end{aligned}$$

Por otra parte, la familia

$$\mathscr{M} = \{A \in \mathscr{F} : \mathbb{E}(\mathbf{1}_A \circ \theta_n \mid X_n = i) = \mathbb{P}_i(A)\}$$

es un λ-sistema (contenido en $\mathscr{F}$) que por el párrafo anterior contiene a $\mathscr{C}$. Puesto que $\sigma(\mathscr{C}) = \mathscr{F}$, se sigue que $\mathscr{M} = \mathscr{F}$. □

Ahora daremos algunos ejemplos de aplicación de la propiedad de Markov.

Ejemplo 2.6 (El problema de la ruina). Consideremos la matriz de transición $P_{i,i+1} = 1 - P_{i,i-1} = p \in (0, 1)$ que da lugar a una caminata aleatoria simple en $\mathbb{Z}$ y a sus familia markoviana $\mathbb{P}_i, i \in \mathbb{Z}$ en el espacio canónico. Sea X el proceso canónico. Consideremos $n \geq 2$ y $m \in \{1, \ldots, n-1\}$. Definamos al siguiente tiempo de paro:

$$T = \min \{n \geq 0 : X_n \in \{0, n\}\} .$$

El objetivo del ejemplo es utilizar la propiedad de Markov, mediante el llamado análisis del primer paso para calcular la probabilidad siguiente:

$$\mathbb{P}_m(X_T = n)\,. \tag{9}$$

La interpretación es la siguiente: un jugador tiene un capital inicial de m pesos y se enrola en un juego de volados en el que gana un peso con probabilidad p (y los juegos son independientes). No deja de jugar hasta que pasan dos cosas: acumula un capital objetivo de n pesos o pierde toda su fortuna. Lo que se quiere determinar es la probabilidad de que termine de jugar con n pesos.

Antes que nada, probaremos que $T < \infty$ $\mathbb{P}_m$ casi seguramente. En efecto, notemos que

$$\begin{aligned}
&\mathbb{P}_m(T \leq k + n \mid \mathscr{F}_k) \\
&= \mathbb{P}_m(T \leq k \mid \mathscr{F}_k) + \mathbb{P}_m(k < T \leq k + n \mid \mathscr{F}_k) \\
&\geq \mathbf{1}_{T \leq k} + \mathbb{P}_m(k < T, X_{k+1} - X_k = 1, \ldots, X_{k+n} - X_{k+n-1} = 1 \mid \mathscr{F}_k)\,.
\end{aligned}$$

Al utilizar la propiedad de Markov obtenemos

$$\mathbb{P}_m(T \leq k + n \mid \mathscr{F}_k) \geq \mathbf{1}_{T \leq k} + \mathbf{1}_{k<T} p^n \geq p^n,$$

por lo cual, por el ejercicio E10.5 del libro de Williams, vemos que T es finito $\mathbb{P}_m$ casi seguramente. Otra forma más intuitiva de probar que $T < \infty$ $\mathbb{P}_m$ casi seguramente es notar que si hay n incrementos de X igual a 1 (lo cual sucede con probabilidad p^n, entonces $T < \infty$). Sin embargo, como bajo $\mathbb{P}_m$ los incrementos son independientes y toman el valor 1 con probabilidad p, esto sucedera casi seguramente (como se ve por ejemplo al utilizar Borel-Cantelli).

Escribamos

$$q_m = \mathbb{P}_m(X_T = n)$$

y determinemos ahora el valor de q_m. Hay dos casos sencillos:

$$q_0 = 0 \quad \text{y} \quad q_n = 1.$$

Por otro lado, observamos que si $m \in \{1, \ldots, n-1\}$ entonces $1 \leq T$ y $X_T = X_T \circ \theta_1$ $\mathbb{P}_m$ casi seguramente, por lo que

$$q_m = p q_{m+1} + (1 - p)\, q_{m-1}.$$

Así, la probabilidad de interés queda determinada por una relación de recurrencia con valores de frontera. Afortunadamente, la solución se conoce. En efecto, escribamos la relación de recurrencia en la forma

$$p q_m + q q_m = p q_{m+1} + q q_{m-1} \quad \text{ó} \quad p q_m + q q_m = p q_{m+1} + q q_{m-1}$$

ó inclusive

$$(q_m - q_{m+1}) = (q/p)\,(q_{m-1} - q_m)\,.$$

Se sigue entonces que

$$q_m - q_{m+1} = -q_1\,(q/p)^m$$

y por lo tanto

$$q_m = \begin{cases} q_1 \frac{1-(q/p)^m}{1-(q/p)} & \text{si } q \neq p \\ q_1 m & \text{si } q = p = 1/2 \end{cases}.$$

Al utilizar la igualdad $q_n = 1$ obtenemos finalmente

$$q_m = \begin{cases} \frac{1-(q/p)^m}{1-(q/p)^n} & \text{si } q \neq p \\ m/n & \text{si } q = p \end{cases}.$$

Cuando $p = q = 1/2$, podemos adicionalmente calcular $v_m = \mathbb{E}_m(T)$. En efecto, por la propiedad de Markov y el hecho de que $T \circ \theta_1 = \mathbf{1}_{T \geq 1} + T$, vemos que

$$v_0 = v_n = 0 \quad \text{y} \quad 2v_m = 2 + v_{m+1} + v_{m-1}.$$

La última ecuación se puede escribir, definiendo $d_m = v_m - v_{m-1}$ como una igualdad matricial:

$$\begin{pmatrix} d_{m+1} \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_m \\ -2 \end{pmatrix},$$

por lo que

$$\begin{pmatrix} d_{m+1} \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^m \begin{pmatrix} d_1 \\ -2 \end{pmatrix}.$$

La potencia de la matriz resultante se puede calcular y nos devuelve

$$d_{m+1} = d_1 - 2m.$$

Puesto que $d_1 = v_1$ y $v_0 = 0$, vemos que

$$v_m = d_1 + \cdots + d_m = v_1 - v_1 - 2 - \cdots - v_1 - 2(m-1) = mv_1 - m(m-1).$$

Al utilizar $v_n = 0$, vemos que

$$v_1 = n - 1$$

y que por lo tanto

$$v_m = m(n-m).$$

Ejemplo 2.7 (Probabilidad de superviviencia para procesos de Galton-Watson). Sea μ una medida de probabilidad en $\mathbb{N}$ y $(\mathbb{P}_k)_{,k\in\mathbb{N}}$ la familia Markoviana asociada a un proceso de Galton-Watson de distribución de progenie ν. Sea f la función generadora de μ, dada por

$$f(s) = \sum_n s^n \mu_n.$$

Sea

$$\rho_i = \mathbb{P}_i(T_0 < \infty).$$

Al utilizar la propiedad de ramificación, vemos que $\rho_i = \rho_1^i$ puesto que la suma de i procesos GW con ley de reporducción μ se anula si y sólo si se anulan todos los sumandos. Por otra parte, al utilizar la propiedad de Markov al instante uno, vemos que

$$\rho_1 = \mathbb{P}_1(T_0 < \infty) = \mathbb{E}_1(\mathbb{E}_{X_1}(T_0 < \infty)) = \mathbb{E}_1\left(\rho_1^{X_1}\right) = f(\rho_1)\,,$$

donde la última igualdad se justifica pues bajo $\mathbb{P}_1$, X_1 tiene distribución μ. Así, ρ_1 es un punto fijo de la función generadora f. Dicha función es log-convexa, y toma los valores μ_0 en 0 y 1 en 1. Por lo tanto, vemos que habrá exáctamente un punto fijo cuando $m = f'(1-) \leq 1$, mientras que hay dos puntos fijos si $\mu > 1$. Así, vemos que cuando $m \leq 1$, el proceso de Galton-Watson se extingue casi seguramente.

Por otra parte, cuando $m > 1$, cabe la pregunta de cuál de los dos puntos fijos de f es igual a la probabilidad de superviviencia del proceso de Galton-Watson. Debe ser la mínima. En efecto, notemos que si ρ es punto fijo de f entonces

$$\rho = \mathbb{E}_1\left(\rho^{X_n}\right)$$

y puesto que $X_n = 0$ si $T \leq n$ entonces

$$\rho \geq \mathbb{E}_1\left(\rho^{X_n}\mathbf{1}_{T_0\leq n}\right) = \mathbb{P}_1(T_0 \leq n)\,.$$

Al tomar el límite conforme $n \to \infty$ obtenemos

$$\rho \geq \mathbb{P}_1(T_0 < \infty) = \rho_1.$$

Por lo tanto, ρ_1 es el punto fijo más pequeño de f y como hemos argumentado, este pertenece a $(0,1)$ cuando $m > 1$. Así, un proceso de Galton-Watson supercrítico sobrevive con probabilidad positiva.

Ejemplo 2.8 (Cadenas de nacimiento y muerte). Para una sucesión de probabilidades de éxito $p_i \in (0,1)$ (con $q_i = 1 - p_i$), sea $(\mathbb{P}_i, i \geq 0)$ la familia markoviana asociada a un proceso de nacimiento y muerte con matriz de transición deterninada por $P_{i,i+1} = p_i$. Haremos un análisis de primer paso para calcular $h_i = \mathbb{P}_i(T_0 < \infty)$. Notemos que $h_0 = 1$.

De nuevo, la propiedad de Markov nos permite obtener la siguiente relación de recurrencia para $i \geq 1$:

$$h_i = p_i h_{i+1} + q_i h_{i-1}.$$

De nueva cuenta, si escribimos $d_i = h_i - h_{i+1}$, se tiene que

$$p_i d_i = q_i d_{i-1}$$

y sabemos que $d_0 = 1 - h_1$. Vemos que entonces

$$d_i = \gamma_i d_0 \quad \text{donde} \quad \gamma_i = \frac{p_i \cdots p_1}{q_i \cdots q_1}.$$

Así, podemos escribir

$$d_0\,(1 + \gamma_1 + \cdots + \gamma_{i-1}) = d_0 + \cdots + d_{i-1} = 1 - h_i.$$

Ahora debemos hacer un análisis más preciso para determinar a la constante faltante d_0.

Si $\sum_i \gamma_i = \infty$, puesto que $1 - h_i \in [0,1]$, vemos que $d_0 = 1 - h_1 = 0$ y por lo tanto $h_i = 1$ para toda $i \geq 1$.

Si $\sum_i \gamma_i < \infty$, definamos

$$\tilde{h}_0 = 1 \quad \text{y} \quad \tilde{h}_i = 1 - a\,(1 + \gamma_0 + \cdots + \gamma_i)\,,$$

por lo que $\tilde{h}_i \in [0,1]$ y

$$\tilde{h}_i = p_i \tilde{h}_{i+1} + q_i \tilde{h}_{i-1} \quad \text{si } i \geq 1.$$

Vemos que entonces

$$\tilde{h}_i = \mathbb{E}_i\Big(\tilde{h}_{X_n}\Big)$$

para toda $n \geq 1$. Sin embargo:

$$\mathbb{E}_i\Big(\tilde{h}_{X_n}\Big) = \mathbb{P}_i(T_0 \leq n) + \mathbb{E}_i\Big(\tilde{h}_{X_n} \mathbf{1}_{n<T_0}\Big) \geq \mathbb{P}_i(T_0 \leq n)\,.$$

Como lo anterior es válido para toda n, vemos que

$$h_i \leq \tilde{h}_i.$$

Al minizar sobre a en la definición de $\tilde{h}_i$, vemos que $d_0 = 1/(1 + \sum_i \gamma_i)$. Por lo tanto $h_i < 1$ para toda $i \geq 1$.

Ejemplo 2.9 (Caminatas aleatorias y funciones armónicas discretas). El problema de Dirichlet para la ecuación de Laplace es un problema clásico importante dentro de la física y la matemática. Sea $U \subset \mathbb{R}^n$ un abierto y conexo y δU su frontera, y sea $f : \delta U \to \mathbb{R}$. El problema es encontrar una función $u : \overline{U} \to \mathbb{R}$ tal que

$$\Delta u = 0 \text{ en } U \quad \text{y} \quad u = f \text{ en } \partial U.$$

Una función u que satisface $\Delta u = 0$ se llama función armónica. De hecho, dichas funciones satisfacen la propiedad del promedio

$$u(x) = \int_{\partial B_\delta(x)} u(y)\ \sigma_r(dy)\,,$$

donde σ_r es la distribución de una variable uniforme sobre $\partial B_\delta(x)$. Esta propiedad también admite una forma en términos de volumen en vez de medidas superficiales:

$$u(x) = \frac{1}{\mathrm{Vol}(B_\delta(x))} \int_{B_\delta(x)} u(y)\ dy.$$

Recíprocamente, una función es armónica si es localmente integrable y satisface la propiedad del promedio en su versión de volumen. En base a esta equivalencia, la solución al problema de Dirichlet para la ecuación de Laplace se puede interpretar de la siguiente manera: si la frontera de la región U se mantiene a temperatura constante $f(x)$ en el punto $x \in \delta U$, entonces $u(x)$ representará la temperatura en el punto $x \in U$ una vez que se haya alcanzado el equilibrio térmico.

Discreticemos ahora el problema para encontrar una solución por medio de caminatas aleatorias. Definamos, para $\delta > 0$ a la discretización U_δ de U mediante la retícula $\delta\mathbb{Z}^n$:

$$U_\delta = \{i \in \delta\mathbb{Z}^n \cap U : \text{los vecinos de } i \text{ en } \delta\mathbb{Z}^n \text{ están en } U\}$$

y a su frontera

$$\delta U_\delta = \{i \in \delta\mathbb{Z}^n \cap U : \text{algún vecino de } i \text{ en } \delta\mathbb{Z}^n \text{ no está en } U\}.$$

Nuestra discretización del problema de Dirichlet de la ecuación de Laplace se discretiza de la siguiente manera: dada una función $f : \partial U_\delta \to \mathbb{R}$, encontrar una función $u : U_\delta \cup \partial U_\delta \to \mathbb{R}$ tal que

$$u(x) = \frac{1}{2n} \sum_{y \text{ vecino de } x} u(y) \text{ si } x \in U_\delta \quad \text{y} \quad u(x) = f(x) \text{ si } x \in \partial U_\delta.$$

Ahora resolveremos este problema discreto.

Sea $(\mathbb{P}_x, x \in \delta\mathbb{Z}^n)$ la familia markoviana asociada a una caminata aleatoria que salta de x a cualquiera de sus vecinos con igual probabilidad. Si X es el proceso canónico, consideremos al tiempo de paro

$$\tau = \min\{n \in \mathbb{N} : X_m \in \partial U_\delta\}.$$

Supondremos ahora que U es acotado, lo cual tiene el efecto de que para cualquier $x \in U_\delta$, $\tau < \infty$ $\mathbb{P}_x$-casi seguramente. En efecto, puesto que U_δ es acotado, notamos que existe $N \in \mathbb{N}$ tal que si la caminata sube N veces entonces sale de U_δ (necesariamente por ∂U_δ) para cualquier punto inicial $x \in U_\delta$. Como los incrementos de la caminata son variables aleatorias independientes bajo $\mathbb{P}_x$, vemos que esto sucederá eventualmente. Por otra parte, por definición τ podemos entonces definir a la función $u : U_\delta \cup \partial U_\delta \to \mathbb{R}$ por medio de

$$u(x) = \mathbb{E}_x(f(X_\tau)).$$

Notemos que si $x \in U_\delta$, entonces $\tau \geq 1$ y $X_\tau \circ \theta_1 = X_\tau$ $\mathbb{P}_x$ casi seguramente, por lo cual la propiedad de Markov nos permite ver que

$$u(x) = \mathbb{E}_x(f(X_\tau)) = \mathbb{E}_x(u(X_1)) = \frac{1}{2n} \sum_{y \text{ vecino de } x} u(y).$$

Ahora veamos que la solución al problema de Dirichlet para la ecuación de Laplace discreta admite una única solución. Esto se basa en el llamado principio del máximo que nos dice lo siguiente: si una función u es armónica discreta y alcanza su máximo en un elemento x de U_δ, entonces todos sus vecimos también deben tomar el valor $u(x)$, pues de otra manera el promedio no puede ser igual a $u(x)$. Así, vemos que de hecho u es constante. Otra forma de interpretar este argumento es que una función armónica (discreta) alcanza su máximo y su mínimo en la frontera ∂U_δ. Por otra parte, si u_1 y u_2 son armónicas y tienen los mismos valores a la frontera entonces $u = u_2 - u_1$ es una función armónica que toma los valores cero en la frontera, por lo que es igual a cero en $U_\delta \cup \partial U_\delta$. Así, las caminatas aleatorias nos permiten obtener la única función armónica con valores a la frontera dados.

Hay una extensión muy útil de la propiedad de Markov conocida como la propiedad de Markov fuerte. Se trata de una extensión a ciertos tiempos aleatorios: a tiempos de paro. Para esto, sea $T : E^{\mathbb{N}} \to \mathbb{N}$ un tiempo de paro finito. Entonces podemos adelantarnos al tiempo T mediante el operador de traslación $\theta_T : E^{\mathbb{N}} \to E^{\mathbb{N}}$ dado por:

$$\theta_T\big((e_n)_{n\in\mathbb{N}}\big) = \big(e_{T(e)+n}\big)_{n\in\mathbb{N}}.$$

Proposición 2.5 (Propiedad de Markov fuerte). *Si $F : E^{\mathbb{N}} \to \mathbb{R}$ es medible y acotada:*

$$\mathbb{E}(F \circ \theta_T \mid \mathscr{F}_T) = \mathbb{E}_{X_T}(F).$$

Demostración. Si $A \in \mathscr{F}_T$ entonces $A \cap \{T = n\} \in \mathscr{F}_n$ por lo que

$$\mathbb{E}\big(\mathbf{1}_{A\cap\{T=n\}} F \circ \theta_T\big) = \mathbb{E}\big(\mathbf{1}_{A\cap\{T=n\}} F \circ \theta_n\big) = \mathbb{E}\big(\mathbf{1}_{A\cap\{T=n\}} \mathbb{E}_{X_n}(F)\big).$$

Al sumar sobre n (utilizando el teorema de convergencia acotada) obtenemos

$$\mathbb{E}(\mathbf{1}_A F \circ \theta_T) = \mathbb{E}(\mathbf{1}_A \mathbb{E}_{X_T}(F)).$$ □

Pasaremos ahora a algunas aplicaciones de la propiedad de Markov fuerte.

La primera es muy sencilla y es simplemente una reinterpretación de la propiedad de Markov fuerte en el caso de la caminata aleatoria.

Corolario 1. *Sea $(\mathbb{P}_k, \in \mathbb{Z})$ la familia markoviana que corresponde a una caminata aleatoria. Sea T un tiempo de paro finito $\mathbb{P}_k$-casi seguramente. Entonces la distribución condicional de $(X - X_0) \circ \theta_T$ dada $\mathscr{F}_T$ es $\mathbb{P}_0$. En particular, $(X - X_0) \circ \theta_T$ es independiente de $\mathscr{F}_T$.*

Demostración. Recordemos que $\mathbb{P}_k$ es la distribución de $X + k$ bajo $\mathbb{P}_0$. Por la propiedad de Markov fuerte, se sigue que la distribución condicional de $(X - X_0) \circ \theta_T$ dada $\mathscr{F}_T$ es $\mathbb{P}_0$. En particular, si $A \in \mathscr{F}_T$ y $F : E^{\mathbb{N}} \to \mathbb{R}$ es medible y acotada entonces

$$\mathbb{E}(\mathbf{1}_A F(X - X_0) \circ \theta_T) = \mathbb{E}(\mathbf{1}_A \mathbb{E}_0(F(X))) = \mathbb{P}(A)\, \mathbb{E}_0(F),$$

lo cual implica la independencia enunciada. □

Ejemplo 2.10 (Tiempos de arribo para la caminata aleatoria simple unidimensional). Sea $(\mathbb{P}_k, k \in \mathbb{Z})$ la familia Markoviana asociada a la caminata aleatoria simple con matriz de transición $P_{i,i+1} = 1 - P_{i,i-1} = p \in (0,1)$. Sea T_0 el primer arribo a cero dado por

$$T_0 = \min\{n : X_n = 0\}.$$

Nuestro objetivo será determinar a

$$\phi(s) = \mathbb{E}_1\big(s^{T_0}\big).$$

Comenzando en 2, la caminata aleatoria simple debe pasar por uno para llegar a cero, y la trayectoria de 2 a 1 tiene la misma distribución que la de 1 a cero. Por lo tanto:

$$\mathbb{E}_2\left(s^{T_0}\right) = \phi(s)^2 .$$

Por otra parte, la propiedad de Markov al instante 1 nos dice que

$$\phi(s) = \mathbb{E}_1\left(s^{T_0}\right) = (1-p)s + ps\mathbb{E}_2\left(s^{T_0}\right) = (1-p)\,s + ps\phi(s)^2 .$$

Así, puesto que $\phi(s) \in (0,1)$ para $s \in (0,1)$, vemos que

$$\phi(s) = \frac{1-\sqrt{1-4p\,(1-p)\,s^2}}{2ps}.$$

Esto tiene varias implicaciones. La primera es que

$$\mathbb{P}_1(T_0 < \infty) = \lim_{s\to 1} \mathbb{E}_1\left(s^{T_0}\right) = \frac{1-|1-2p|}{2p} = \begin{cases} 1 & p < 1/2 \\ \frac{q}{p} & p \geq 1/2 \end{cases} .$$

La segunda es el cálculo, para $p \leq 1/2$ de la esperanza de T_0: al derivar la ecuación que satisface ϕ vemos que

$$0 = 1 - (2p-1)\,\phi'(1-)$$

La segunda consecuencia es la determinación explícita de la distribución de T_0 bajo $\mathbb{P}_1$.

Ejercicio 2.3. Haga un desarrollo en serie de ϕ para obtener el valor exacto de $\mathbb{P}_1(T_0 = 2n+1)$. *Sugerencia:* La serie binómica será de utilidad. Pruebe que

$$\mathbb{P}(T = 2n+1) = \frac{1}{2n+1}\mathbb{P}_1(S_{2n+1} = -1) .$$

Ejemplo 2.11 (El mínimo de caminatas aleatorias skip-free). Sea $\mathbb{P}_k, k \in \mathbb{Z}$ la familia markoviana asociada a una caminata aleatoria cuyos saltos pertenecen a $\{-1, 0, 1, \ldots\}$. Dichas caminatas se llaman sin saltos a la izquierda (skip-free to the left) aunque el punto es que el mínimo acumulativo tiene como imagen un intervalo de enteros. Sea

$$-I = \min_{n \geq 0} X_n.$$

Obviamente $I = 0$ si $\mathbb{P}_0(X_1 = -1) = 0$. Supongamos por lo tanto que este no es el caso. Veamos que entonces bajo $\mathbb{P}_0$, I es una variable aleatoria geométrica (aunque posiblemente degenerada en infinito). Recordemos que las variables geométricas están caracterizadas por la propiedad de pérdida de memoria, por lo que es suficiente verificar que

$$\mathbb{P}_0(I \geq m+n) = \mathbb{P}_0(I \geq m)\,\mathbb{P}_0(I \geq m) .$$

Sin embargo, notemos primero que si

$$T_{-m} = \min \{n \in \mathbb{N} : X_n = -m\},$$

entonces $\{I \geq m\} = \{T_{-m} < \infty\}$. Segundo, sobre el conjunto $\{T_{-m} < \infty\}$ se tiene que $I = I \circ \theta_{T_{-m}}$. La propiedad de Markov fuerte nos permite afirmar que

$$\mathbb{P}_0(I \geq m + n) = \mathbb{P}_0(I \geq m)\, \mathbb{P}_{-m}(I \geq n + m).$$

Finalmente, puesto que $\mathbb{P}_0$ es la distribución de $X + m$ bajo $\mathbb{P}_{-m}$, vemos que

$$\mathbb{P}_{-m}(I \geq n + m) = \mathbb{P}_0(I \geq n).$$

Ahora determinemos el parámetro de la distribución de I, al que denotaremos por p y escribimos $q = 1 - p$. Al aplicar la propiedad de Markov al tiempo 1, vemos que

$$q = \mathbb{P}_0(I \geq 1) = \mathbb{P}_0(\mathbb{P}_{X_1}(I \geq 1)) = \mathbb{E}_0\left(q^{1+X_1}\right).$$

Si ϕ denota a la función generadora de la distribución de salto, vemos que

$$1 = \phi(q).$$

Por la desigualdad de Hölder, es fácil ver que la función $\phi\left(e^{-\lambda}\right)$ es log-convexa. Por lo tanto la ecuación $\phi(s) = 1$ sólo se satisface cuando $s = 1$ si $\phi'(1-) = \mathbb{E}_0(X_1) \leq 0$ (lo cual dice que I es casi seguramente infinito) mientras que admite dos soluciones, digamos $\tilde{q}$ y 1 si $\mathbb{E}_0(X_1) > 0$. Ahora veremos que $\tilde{q} = q$ en este último caso. Basta mostrar que $q \leq \tilde{q}$. Puesto que

$$\mathbb{E}_k\left(\tilde{q}^{X_1}\right) = \tilde{q}^k \mathbb{E}_0\left(\tilde{q}_1^X\right) = \tilde{q}^k \phi(\tilde{q}) = \tilde{q}^k,$$

la propiedad de Markov nos permite probar que

$$\mathbb{E}_k\left(\tilde{q}^{X_n}\right) = \tilde{q}^k.$$

Por lo tanto, al utilizar la propiedad de Markov

$$\begin{aligned}
\tilde{q} = \mathbb{E}_0\left(\tilde{q}^{1+X_n}\right) &= \sum_{k=0}^{n} \mathbb{E}_0\left(\mathbf{1}_{T_{-1}=k} \mathbb{E}_{-1}\left(\tilde{q}^{1+X_{n-k}}\right)\right) + \mathbb{E}_0\left(\mathbf{1}_{T_{-1}>n} \tilde{q}^{1+X_n}\right) \\
&= \mathbb{P}_0(T_1 \leq n) + \mathbb{E}_0\left(\mathbf{1}_{T_{-1}>n} \tilde{q}^{1+X_n}\right) \\
&\geq \mathbb{P}_0(T_1 \leq n).
\end{aligned}$$

Al tomar el límite conforme $n \to \infty$, vemos que $\tilde{q} \geq q$ y por lo tanto, q es la solución mínima en $[0, 1]$ a la ecuación $\phi(s) = 1$.

Ejemplo 2.12 (El principio de reflexión). Consideremos la familia markoviana $(\mathbb{P}_k)_{k\in\mathbb{Z}}$ asociada a la caminata aleatoria simple y sea

$$T_m = \min \{n \in \mathbb{N} : X_n = m\}.$$

Definamos

$$\tilde{X}_n = \begin{cases} X_n & \text{si } n \leq T_m \\ 2m - X_n & \text{si } n > T_m \end{cases}.$$

El principio de reflexión, que ahora demostaremos, afirma que la distribución de $\tilde{X}$ bajo $\mathbb{P}_0$ es precisamente $\mathbb{P}_0$. En efecto, el principio de reflexión es consecuencia de la propiedad de Markov fuerte y el hecho de que la distribución de $-X$ bajo $\mathbb{P}_0$ es también $\mathbb{P}_0$.

Como una aplicación, notemos que si $k \geq 0$ entonces

$$\mathbb{P}_0(T_m \leq n) = \mathbb{P}_0(T_m \leq n, X_n \geq m) + \mathbb{P}_0(T_m \leq n, X_n < m).$$

Puesto que $\{T_m \leq n\} \subset \{X_n \geq m\}$ y

$$\{T_m \leq n, X_n \leq m\} = \left\{\tilde{X}_n \geq m\right\},$$

se deduce que

$$\mathbb{P}_0(T_m \leq n) = \mathbb{P}_0(X_n = m) + 2\mathbb{P}_0(X_n > m).$$

Ejemplo 2.13 (El teorema de Pitman para la caminata aleatoria simple). Ahora probaremos un resultado interesante de Pitman para la caminata aleatoria simple y simétrica. Sea $(\mathbb{P}_k)$ la familia markoviana asociada a dicho proceso; por simplicidad, denotaremos por $\mathbb{P}$ a $\mathbb{P}_0$. Lo que haremos será introducir una transformación no invertible de la caminata aleatoria simple que resulta ser una cadena de Markov que podemos interpretar como la caminata aleatoria condicionada a ser no-negativa. Puesto que la caminata aleatoria oscila (esto es $\mathbb{P}_k(\limsup_n X_n = \infty \liminf_n X_n = -\infty) = 1$) vemos que no podemos simplemente condicionar. Sin embargo, en [**KS63**] se prueba que para toda $k \geq 0$ existe el límite de

$$\lim_{n\to\infty} \mathbb{P}_k(X_1 = i_1, \ldots, X_m = i_m \mid X_k \geq 0 \text{ para toda } k \in \{1, \ldots, n\}).$$

A este proceso se le conoce como caminata aleatoria (simple y simétrica) condicionada a ser positiva. Hay otra manera de entender el proceso límite. No es trivial, pero el límite anterior es igual a

$$\lim_{K\to\infty} \mathbb{P}_k(X_1 = i_1, \ldots, X_m = i_m \mid X \text{ alcanza } K \text{ antes que } -1).$$

El límite anterior lo podemos calcular explícitamente gracias a la solución al problema de la ruina. En efecto, dicho límite no es cero cuando $i_1, \ldots, i_m \geq 0$ en cuyo caso:

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}_k(X_1 = i_1, \ldots, X_m = i_m \mid X \text{ alcanza } K \text{ antes que } -1) \\
&= \lim_{K\to\infty} \mathbb{P}_k(X_1 = i_1, \ldots, X_m = i_m) \frac{i_m + 1}{K+1} \frac{K+1}{k+1} \\
&= \mathbb{P}_k(X_1 = i_1, \ldots, X_m = i_m) \frac{i_m + 1}{k+1}.
\end{aligned}$$

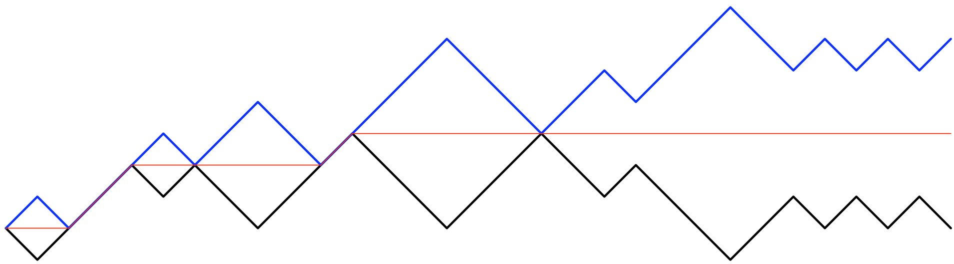

Figura 1. Reflexión de una caminata aleatoria (en negro) en su máximo acumulativo (en rojo)

Hay una manera de interpretar lo anterior en términos de cadenas de Markov. Si P es la matriz de transición de la caminata aleatoria simple y simétrica y definimos $h(i) = i$ así como a la matriz de transición P^h dada por

$$P^h_{i,j} = P_{i,j} \frac{h(j)}{h(i)},$$

entonces el cálculo anterior nos muestra que la caminata aleatoria condicionada a ser positiva tendría que ser una cadena de Markov con matriz de transición P^h.

Sea $\overline{X}_n = \max_{m \leq n} X_m$ el proceso del máximo acumulativo. Definamos a un nuevo proceso Y por medio de

$$Y_n = 2\overline{X}_n - X_n.$$

En la Figura 1 podemos ver el efecto de la transformación que lleva a X en Y. El teorema de Pitman, enunciado en [**Pit75**], afirma que Y es una cadena de Markov bajo $\mathbb{P}_0$ con matriz de transición P^h. El argumento es como sigue: nos damos de que si $i_1, \ldots, i_n$ son una trayectoria posible para Y, esto es si $i_k - i_{k=1} = \pm 1$ e $i_k \geq 0$, entonces hay exactamente $i_n + 1$ trayectorias de X que hacen que $\{Y_1 = i_1, \ldots, Y_n = i_n\}$. En efecto, cada una de dichas trayectorias está determinada por el valor de de $\overline{X}_n$, que puede tomar cualquier valor entre 0 e i_n. Por lo tanto,

$$\mathbb{P}(Y_1 = i_1, \ldots, Y_n = i_n) = \frac{1}{2^n} (i_n + 1) = \mathbb{P}^h_0(X_1 = i_1, \ldots, X_n = i_n).$$

Puesto que las distribuciones de procesos estocásticos están determinadas en los cilindros finito-dimensionales, se ha probado el teorema de Pitman.

10. Recurrencia y transitoriedad

Pasaremos ahora al análisis de dos conceptos que permiten hacer una distinción entre los estados de una cadena de Markov, de acuerdo a si siempre los revisitaré o no.

Sea $i \in E$. Definamos a la cantidad de visitas al estado i como la variable aleatoria

$$V_i = \sum_{n=0}^{\infty} \mathbf{1}_{X_n = i}.$$

Esta variable aleatoria podría tomar el valor infinito. Sin embargo, un resultado curioso es que si toma el valor infinito con probabilidad positiva, entonces toma el valor infinito con probabilidad 1. En caso de que V_i sea infinita con probabilidad 1 bajo $\mathbb{P}_i$ hablamos de un **estado recurrente** y en caso contrario de un **estado transitorio**. Analicemos ahora por qué el conjunto $\{V_i = \infty\}$ tiene probabilidad cero o uno. Para esto, definamos a $T_0, T_1, \ldots$ como los instantes sucesivos que X visita al estado i. Bajo la medida $\mathbb{P}_i$,

$$T_0 = 0, \quad T_1 = \min\{n > 0 : X_n = i\} \quad \text{y} \quad T_{n+1} = T_1 \circ \theta_{T_n}.$$

Se sigue entonces que $(T_n, n \in \mathbb{N})$ es una sucesión creciente de tiempos de paro. Esto se puede probar por inducción. Es claro que T_1 es un tiempo de paro y entonces

$$\begin{aligned}\{T_{n+1} = k\} &= \bigcup_{l=0}^{k-1} \{T_{n+1} = k, T_n = l\} \\ &= \bigcup_{l=0}^{k-1} \{T_n = l, X_{l+1}, \cdots X_{k-1} \neq i, X_k = i\} \in \mathscr{F}_k.\end{aligned}$$

Notemos que

$$V_i = \sum_{n=1}^{\infty} \mathbf{1}_{T_n < \infty}.$$

Se afirma ahora que bajo $\mathbb{P}_i$, V_i es una variable aleatoria geométrica de parámetro $\mathbb{P}_i(T_1 < \infty)$. En efecto, por una parte se tiene que

$$\{V_i \geq n\} = \{T_n < \infty\}$$

y por otra, la propiedad de Markov fuerte nos permite afirmar que para cada $n \geq 1$:

$$\mathbb{P}_i(T_{n+1} < \infty) = \mathbb{P}_i(T_{n+1} < \infty, T_n < \infty) = \mathbb{E}_i(\mathbf{1}_{T_n < \infty} \mathbb{P}_i(T_1 < \infty)),$$

por lo cual

$$\mathbb{P}_i(T_n < \infty) = \mathbb{P}_i(T_1 < \infty)^n.$$

El caso en que $\mathbb{P}_i(T_1 < \infty) = 1$ ocurre si y sólo si V_i es infinita $\mathbb{P}_i$ casi seguramente. Si no, V_i es geométrica de parámetro $\mathbb{P}_i(T_1 < \infty)$ y por lo tanto su esperanza es finita. Esto nos proporciona una equivalencia, en términos de la matriz de transición, para que un estado sea recurrente.

Proposición 2.6. *El estado i es recurrente si y sólo si $\sum_i P^n_{i,i} < \infty$.*

Demostración. La afirmación se sigue de notar que

$$\mathbb{E}_i(V_i) = \sum_n \mathbb{E}_i(\mathbf{1}_{X_n=i}) = \sum_n P^n_{i,i}.$$

□

Ahora veremos que la transitoriedad o recurrencia es de hecho una propiedad de clase.

Proposición 2.7. *Si i y j se comunican entre si e i es transitorio entonces j es transitorio.*

Demostración. Sean m y n tales que $P^m_{i,j} > 0$ y $P^n_{j,i} > 0$. Entonces

$$P^{m+l+n}_{i,i} \geq P^m_{i,j} P^l_{j,j} P^n_{j,i}.$$

Por lo tanto:

$$\text{si } \sum_n P^{m+l+n}_{i,i} < \infty \text{ entonces } \sum_n P^l_{j,j} < \infty.$$

□

La conclusión que obtenemos es que en una clase o todos los estados son recurrentes o todos son transitorios y que por lo tanto podemos hablar de **clases recurrentes** y de **clases transitorias**. Hay una forma fácil de saber si una clase es transitoria.

Proposición 2.8. *Sea $C \subset E$ una clase abierta. Entonces C es transitoria.*

Demostración. En efecto, puesto que C es una clase abierta, existe $i \in C$, $j \in E \setminus C$ y $m \geq 0$ tal que $P^m_{i,j} > 0$ mientras que $P^n_{j,i} = 0$ para toda $n \geq 0$. Por lo tanto

$$\mathbb{P}_j(V_i = 0) = 1.$$

Así, vemos que

$$\mathbb{P}_i(V_i < \infty) \geq \mathbb{P}_i(V_i \circ \theta_n < \infty, X_m = j) = P^m_{i,j} > 0$$

por lo que i es transitorio. □

Veremos ahora que la conclusiónes anteriores nos permiten clasificar a las clases de cadenas de Markov con espacio de estados finito. En efecto,

Proposición 2.9. *Si el espacio de estados es finito, una clase es recurrente si y sólo si es cerrada.*

Demostración. Sólo hace falta verificar que si C es cerrada entonces es recurrente. Puesto que C es cerrada, vemos que para cualquier $i \in C$,

$$1 = \pi_i(X_n \in C \text{ para toda } n \geq 0).$$

Por otra parte, al ser E finito, lo anterior forza a que exista $j \in C$ que se visita infinitas veces bajo $\mathbb{P}_i$:

$$0 < \mathbb{P}_i(V_j = \infty).$$

Si T denota a la primera visita a j, vemos que

$$0 < \mathbb{P}_i(T_j < \infty)\, \mathbb{P}_j(V_j = \infty)$$

de acuerdo a la propiedad de Markov fuerte. Por lo tanto, vemos que j es recurrente y que así la clase C es recurrente. □

11. Distribuciones invariantes

En esta seccin analizaremos el concepto de distribuciones invariantes que es fundamental para estudiar cmo se comporta una cadena de Markov en tiempos grandes.

Sea P una matriz de transición en el conjunto a lo más numerable E. Sea $(\mathbb{P}_x)_{x\in E}$ su familia Markoviana asociada y recordemos que entonces, si ν es una medida de probabilidad en E (que es determinada por los valores ν_x que asocia a los conjuntos $\{x\}$), hemos definido

$$\mathbb{P}_\nu = \sum_{x\in E} \nu(\{x\})\, \mathbb{P}_x$$

como la distribución de una cadena de Markov de matriz de transición P y distribución inicial ν.

Recordemos que bajo $\mathbb{P}_x$, la distribución de X_n es $P^n_{x,\cdot}$, por lo cual:

$$\mathbb{P}_\nu(X_n = y) = \sum_{x\in E} \nu_x P^n_{x,y}.$$

Si nos imaginamos a la distribución ν como un vector renglón, la ecuación anterior se puede interpretar como la multiplicación de matrices νP.

Diremos que una distribución ν sobre E es invariante si

$$\sum_{x\in E} \nu_x P_{x,y} = \nu_y.$$

Equivalentemente, ν es invariante si $\mathbb{P}_\nu \circ X_n^{-1} = \nu$ para toda $n \geq 0$. Obviamente, si ν es invariante, se tiene que $\mathbb{P}_\nu(X_n = y) \to \nu_y$ conforme $n \to \infty$. Sorprendentemente, bajo ciertas condiciones técnicas que garantizan en particular existencia y unicidad de la distribución invariante, se tiene que $\mathbb{P}_x(X_n = y) \to \nu_y$ conforme $n \to \infty$. Es decir, la distribución de la cadena se estabiliza. Este es el contenido del célebre teorema de Doeblin, que admite una prueba probabilística basada en una idea importante y fértil: el acoplamiento.

Esta sección se divide en dos partes: en la primera se utiliza la propiedad de Markov fuerte para estudiar la existencia y unicidad de la distribución estacionaria y en la segunda se utilizan estos resultados para estudiar el comportamiento límite de ciertas cadenas de Markov.

11.1. Existencia y unicidad. Más generalmente, si ν es una medida, no necesariamente de probabilidad, en E, decimos que ν es invariante si

$$\sum_{x\in E}\nu_x P_{x,y}=\nu_y.$$

Examinaremos la existencia de una medida invariante en cada clase de comunicación de la cadena; esto es suficiente pues si ν es cualquier medida invariante y C es una clase de comunicación, la medida $\nu^C_x=\nu_x\mathbf{1}_{x\in C}$ es una nueva medida invariante concentrada en C e inversamente, si para cada clase C tenemos una medida invariante ν^C concentrada en C (y la medida cero también está incluida) entonces $\nu=\sum\nu^C$ es una medida invariante. Notemos además que si ν es una medida invariante y $l>0$ entonces $l\nu$ también es una medida invariante. Puesto que además la medida cero es invariante, cuando estudiemos existencia tenemos que enfocarnos en medidas no triviales y cuando estudiemos unicidad, tendrá que ser salvo múltiplos escalares. Por otro lado, si ν es una medida invariante y $\sum_x\nu_x<\infty$ entonces podemos definir una distribución invariante dada por $\nu/\sum_x\nu_x$. Sin embargo, veremos que aunque existan medidas invariantes, no siempre existen distribuciones invariantes, lo cual nos dice que no siempre podremos encontrar una medida invariante ν tal que $\sum_x\nu_x<\infty$.

Consideremos de nuevo a los tiempos de primera visita

$$T_y=\min\{n\in\mathbb{N}:X_n=y\}.$$

Fijemos a $x\in E$ y definamos

$$\nu^x_x=1\quad\text{y}\quad\nu^x_y=\mathbb{E}_x\left(\sum_{i=0}^{T_x}\mathbf{1}_{X_n=y}\right)\quad\text{para } y\neq x.$$

Teorema 2.2. *Para una cadena irreducible y recurrente, ν^x es una medida invariante que satisface $\nu^x_x=1$ y $0<\nu^x_y<\infty$.*

Demostración. Claramente $\nu^x_x=1$. Para ver que ν^x es invariante, utilizamos la propiedad de Markov de manera similar al análisis del primer paso. Supongamos primero que $y\neq x$. Entonces, puesto que T_x es finito $\mathbb{P}_x$ casi seguramente

$$\mathbb{E}_x\left(\sum_{n<T_x}\mathbf{1}_{X_n=y}\right)=\mathbb{E}_x\left(\sum_{n\leq T_x}\mathbf{1}_{X_n=y}\right)=\sum_{n=1}^{\infty}\mathbb{E}_x(\mathbf{1}_{X_n=y}\mathbf{1}_{T_x\leq n}).$$

El sumando $n=1$ es fácil de calcular puesto que bajo $\mathbb{P}_x$, $T_x\geq 1$:

$$\mathbb{E}_x(\mathbf{1}_{X_1=y}\mathbf{1}_{T_x\leq 1})=\mathbb{E}_x(\mathbf{1}_{X_1=y})=P_{x,y}.$$

Para los sumandos con $n \geq 2$, aplicamos la propiedad de Markov al instante $n-1$ (haciendo una especie de análisis de último paso), notando que $\{T \geq n\} \in \mathscr{F}_{n-1}$:

$$\mathbb{E}_x(\mathbf{1}_{X_n=y}\mathbf{1}_{T_x\leq n}) = \sum_{z\neq x} \mathbb{E}_x\big(\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{X_n=y}\mathbf{1}_{T_x\leq n}\big)$$

$$\sum_{z\neq x} \mathbb{E}_x\big(\mathbf{1}_{X_{n-1}=z}\mathbf{1}_{T_x\leq n}\big)\, P_{z,y}.$$

Así, vemos que

$$\begin{aligned}\nu^x_y = \mathbb{E}_x\left(\sum_{n<T_x} \mathbf{1}_{X_n=y}\right) &= P_{x,y} + \sum_{z\neq x}\sum_{n=1}^{\infty} \mathbb{E}_x(\mathbf{1}_{X_n=y}\mathbf{1}_{T_x>n})\, P_{z,y} \\ &= \nu^x_x P_{x,y} + \sum_{z\neq x} \mathbb{E}_x\left(\sum_{n=0}^{T_x-1} \mathbf{1}_{X_n=y}\right) P_{z,y} \\ &= \sum_{z\in E} \nu^x_z P_{z,y}.\end{aligned}$$

Ahora veremos que $1 = \nu^x_x = \sum \nu^x_y P_{y,x}$. En efecto, basta descomponer respecto al valor de T_x y de X_{T_x-1}, recordando que T_x es finito $\mathbb{P}_x$ casi seguramente:

$$\begin{aligned}1 &= \sum_{n=1}^{\infty} \mathbb{P}_x(T_x = n) \\ &= P_{x,x} + \sum_{n=2}^{\infty}\sum_{y\neq x} \mathbb{P}_x(T_x = n, X_{n-1} = y) \\ &= P_{x,x} + \sum_{n=2}^{\infty}\sum_{y\neq x} \mathbb{P}_x(T_x \geq n-1, X_{n-1} = y)\, P_{x,y} \\ &= P_{x,x} + \sum_{y\neq x} \nu^x_y P_{x,y}.\end{aligned}$$

Finalmente, si consideramos m tal que $P^m_{x,y} > 0$ y n tal que $P^n_{y,x} > 0$ entonces por una parte

$$\nu^x_y = \sum_z \nu^x_z P^m_{z,y} \geq P^m_{x,y} > 0$$

y por otra

$$1 = \nu^x_x = \sum_z \nu^x_z P^n_{x,z} \geq \nu^x_y P^n_{x,y}$$

implica que $\nu^x_y < \infty$. □

El teorema anterior es la pieza principal en la construcción de medidas invariantes. Para esto, es necesario saber cuándo podemos asegurar que $\sum_y \nu^x_y < \infty$, pues en este caso podemos construir una distribución invariante al normalizar ν^x. Por otra parte, notemos que

$$\sum_y \nu_y^x = \sum_y \mathbb{E}_x \left(\sum_{0 \leq n < T_x} \mathbf{1}_{X_n = y} \right) = \mathbb{E}_x(T_x) \, .$$

Definición. Un estado x de una cadena de Markov es **positivo recurrente** si $\mathbb{E}_x(T_x)$. Denotamos por m_x a dicha cantidad, a la que nos referiremos como tiempo medio de recurrencia de x.

Así, vemos que asociado a un estado positivo recurrente en una cadena de Markov irreducible podemos definir una distribución invariante ν que satisface $\nu_x = m_x$. De hecho, veremos un recíproco que afirma que si existe una distribución invariante en una cadena irreducible entonces todos los estados son positivo recurrentes. Antes de eso, notemos que el teorema anterior nos permite la siguiente conclusón:

Corolario 2. *En una cadena irreducible con espacio de estados finito, todos los estados son positivo recurrentes y existe una distribución estacionaria.*

Demostración. Sea x cualquier estado de la cadena. Por el teorema anterior $\nu_y^x < \infty$ para toda y y como E es finito, entonces

$$\mathbb{E}_x(T_x) = \sum_{y \in E} \nu_y^x < \infty.$$

$\square$

Ahora analizaremos la unicidad de las medidas invariantes.

Teorema 2.3. *Si ν es una medida invariante para una matriz de transición irreducible P y $\nu_x = 1$ entonces $\nu \geq \nu^x$. Si además P recurrente entonces $\nu = \nu^x$.*

Demostración. Al aplicar la invariancia de ν se obtiene

$$\nu_y = \sum_{y_1 \in E} \nu_{y_1} P_{y_1,y} = \nu_x P_{x,y} + \sum_{y_1 \neq x} \nu_{y_1} P_{y_1,x} = P_{x,y} + \sum_{y_1 \neq x} \nu_{y_1} P_{y_1,y}.$$

Al volverla a aplicar se obtiene

$$\nu_y = P_{x,y} + \sum_{y_1 \neq x} P_{x,y_1} P_{y_1,y} + \sum_{y_1, y_2 \neq x} \nu_{y_2} P_{y_2,y_1} P_{y_1,y}$$

y al continuar repetidamente, vemos que

$$\nu_y = P_{x,y} + \sum_{y_1 \neq x} P_{x,y_1} P_{y_1,y} + \cdots + \sum_{y_1,\ldots,y_n \neq x} P_{x,y_n} P_{y_n,y_{n-1}} \cdots P_{y_2,y_1} P_{y_1,y}$$
$$+ \sum_{y_1,\cdots,y_{n+1} \neq x} \nu_{y_{n+1}} P_{y_{n+1},y_n} \cdots P_{y_2,y_1} P_{y_1,x}.$$

Si $y \neq x$, encontramos la cota

$$\nu_y \geq \sum_{m=0}^{n} \mathbb{P}_x(X_m = y, T_x \geq m)$$

y el lado derecho converge conforme $n \to \infty$ a ν_y^x, por lo que

$$\nu_y \geq \nu_y^x.$$

Por otra parte, si la cadena es recurrente además de irreducible entonces ν^x es invariante, por lo cual $\mu = \nu - \nu^x$ es una medida invariante con $\mu_x = 0$. Por irreducibilidad, para toda $y \in E$ existe $n \geq 0$ tal que $P^n y, x > 0$ y entonces

$$0 = \mu_x = \sum_z \mu_z P_{z,x}^n \geq \mu_y P_{x,y}^n,$$

por lo que $\mu_y = 0$. □

Finalmente, damos el resultado fundamental de existencia y unicidad de distribuciones invariantes.

Teorema 2.4. *Para una cadena irreducible las siguientes condiciones son equivalentes.*

1. *Todos los estados son positivo recurrentes*
2. *Algún estado es positivo recurrente*
3. *La cadena admite una distribución invariante.*

En este caso, la distribución invariante es única y asigna a x el recíproco de su tiempo medio de recurrencia.

Demostración. Primero probaremos que si la cadena admite una distribución invariante ν entonces todos los estados son positivo recurrentes. Para esto, notemos primero que $\nu_x > 0$ para toda $x \in E$. En efecto, lo debe ser para alguna x y al utilizar la irreducibilidad para encontrar n tal que $P_{y,x}^n > 0$, vemos que

$$\nu_y = \sum_{z \in E} P_{z,y}^n \nu_z \geq \nu_x P_{x,y}^n > 0.$$

Si $x \in E$, entonces ν/ν_x es una medida invariante que asigna 1 a x, por lo cual $\nu/\nu_x \geq \nu^x$. Puesto que ν es una distribución:

$$m_x = \mathbb{E}_x(T_x) = \sum_y \nu_y^x \leq \sum_y \frac{\nu_y}{\nu_x} = \frac{1}{\nu_x} < \infty.$$

Así, todos los estados son positivo recurrentes. Al ser en particular recurrentes, hemos visto que $\nu_x = \nu_x^x = m_x$ y como ν_x es una distribución entonces $m_x < \infty$. Esto termina la demostración de las implicaciones y además nos produce la fórmula requerida para la distribución invariante. □

11.2. Teorema fundamental de convergencia. En esta subsección demostraremos un resultado fundamental sobre cadenas de Markov: el teorema de convergencia a la distribución estacionaria.

Sea P una matriz de transición sobre un espacio de estados a lo más numerable E. Estudiaremos la convergencia de $P^n_{x,y}$ conforme $n \to \infty$ y veremos que bajo ciertas condiciones existe un límite. Para motivar los supuestos que haremos sobre P, consideremos primero a la matriz

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Notemos que $P^{2n+1}_{1,1} = 0$ y que $P^{2n}_{1,1} = 1$. Por lo tanto $P^n_{1,1}$ no puede converger. El problema reside en el carácter periódico del estado 1.

Definición. Decimos que $x \in E$ es aperiódico si existe $N \in \mathbb{N}$ tal que $P^n_{x,x} > 0$ para toda $n \geq N$.

Claramente, si x es aperiódico y $y \sim x$ entonces y es aperiódico, puesto que si $P^m_{x,y}, P^n_{y,x} > 0$ y $N \in \mathbb{N}$ es tal que $P^l_{x,x} > 0$ para toda $l \geq N$ entonces

$$P^{m+n+l}_{y,y} \geq P^n_{y,x} P^l_{x,x} P^m_{x,y} > 0.$$

Así, la aperiodicidad es una propiedad que comparte, o no, cada clase de comunicación. Además, en este caso, también existe N tal que para toda $n \geq N$ se tiene que $P^n_{x,y} > 0$.

Teorema 2.5. *Si P es irreducible, aperiódica y positivo recurrente y ν es su única distribución invariante entonces*

$$\lim_{n \to \infty} P^n_{x,y} = \nu_y.$$

La prueba que daremos se basa en una idea muy fructífera de W. Doeblin: la utilización del acoplamiento. Primero daremos la idea de la demostración y luego verificaremos los detalles. Sean X y Y dos cadenas de Markov independientes con matriz de transición P. Supondremos que $X_0 = x$ y que Y_0 tiene distribución ν. Definamos a

$$W_n = (X_n, Y_n).$$

Se afirma que W es una cadena de Markov irreducible y positivo recurrente. Para cualquier $y \in E$ fijo definamos al tiempo aleatorio

$$T = \min \{n \in \mathbb{N} : X_n = y = Y_n\} = \min \{n \in \mathbb{N} : W_n = (y, y)\}.$$

Puesto que W es irreducible y recurrente, T es finito casi seguramente. Esto nos permite definir al proceso

$$Z_n = \begin{cases} X_n & n \leq T \\ Y_n & n \geq T \end{cases}.$$

Gracias a la propiedad de Markov fuerte, no es descabellado suponer que Z tiene la misma distribución que X. Así, se tendrá que

$$
\begin{aligned}
\left|P^n_{x,y} - \nu_y\right| &= \left|\mathbb{P}(X_n = y) - \mathbb{P}(Y_n = y)\right| \\
&= \left|\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)\right| \\
&= \left|\mathbb{P}(X_n = y, n \leq T) + \mathbb{P}(Y_n = y, T < n) - \mathbb{P}(Y_n = y)\right| \\
&= \left|\mathbb{P}(X_n = y, n \leq T) - \mathbb{P}(Y_n = y, n \leq T)\right| \\
&\leq \mathbb{P}(n \leq T, \{X_n = y\} \, \delta \, \{Y_n = y\}) \\
&\leq \mathbb{P}(n \leq T) .
\end{aligned}
$$

Puesto que T es finito casi seguramente, $\mathbb{P}(n \leq T) \to 0$ conforme $n \to \infty$ y por lo tanto $P^n_{x,y} \to \nu_y$.

Demostración. Finalizaremos la prueba del teorema fundamental de convergencia al verificar los detalles pendientes.

W **es una cadena de Markov:** Puesto que X y Y son independientes entonces

$$
\begin{aligned}
&\mathbb{P}(W_0 = (x_0, y_0), \ldots, W_n = (X_n, y_n)) \\
&= \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n)\, \mathbb{P}(Y_0 = y_0, \ldots, Y_n = y_n) \\
&= \mathbf{1}_{x_0 = x} P_{x_0,x_1} \cdots P_{x_{n-1},x_n} \nu_{y_0} P_{y_0,y_1} \cdots P_{y_{n-1},y_n} .
\end{aligned}
$$

Si definimos a $\tilde{\pi}_{(u,v)} = \mathbf{1}_{u=x}\nu_y$ y a

$$
\tilde{P}_{(u,v),(\tilde{u},\tilde{v})} = P_{u,\tilde{u}} P_{v,\tilde{v}}
$$

entonces $\tilde{\pi}$ es una distribución inicial en E y $\tilde{P}$ es una matriz de transición sobre $E \times E$. Además

$$
\begin{aligned}
&\mathbb{P}(W_0 = (x_0, y_0), \ldots, W_n = (X_n, y_n)) \\
&= \tilde{\pi}_{(x_0,y_0)} P_{(x_0,y_0),(x_1,y_1)} \cdots P_{(x_{n-1},y_{n-1}),(x_n,y_n)},
\end{aligned}
$$

por lo que W es una cadena de Markov con distribución inicial $\tilde{\pi}$ y matriz de transición $\tilde{P}$.

W **es irreducible:** Aquí es donde utilizamos la hipótesis de aperiodicidad de P. En efecto, sean $w_i = (x_i, y_i)$, $i = 1, 2$ elementos de $E \times E$. Sea N tal que $P^n_{x_1,x_2} > 0$ y $P^n_{y_1,y_2} > 0$ para cualquier $n \geq N$. Entonces

$$
\tilde{P}^N_{w_1,w_2} = P^N_{x_1,x_2} P^n_{y_1,y_2} > 0.
$$

W **es positivo recurrente:** Sea $\tilde{\nu}_{(x,y)} = \nu_x \nu_y$. Entonces $\tilde{\nu}$ es una medida de probabilidad y como

$$\begin{aligned}
&\sum_{(x_0,y_0)\in E\times E} \tilde{\nu}_{(x_0,y_0)} P_{(x_0,y_0),(x_1,y_1)} \\
&= \sum_{x_0} \nu_{x_0} P_{x_0,x_1} \sum_{y_0} \nu_{y_0} P_{y_0,y_1} \\
&= \nu_{x_1} \nu_{y_1} \\
&= \tilde{\nu}_{(x_1,y_1)}
\end{aligned}$$

se sigue que $\tilde{\nu}$ es $\tilde{P}$-invariante. Por lo tanto W es positivo recurrente.

Z **tiene la misma distribución que** X**:** Por la propiedad de Markov fuerte, vemos que $(W_{T+n}, n \in \mathbb{N})$ es independiente de W^T y es una cadena de Markov con matriz de transición $\tilde{P}$ que comienza en (y,y). Esto implica que $X_{T+\cdot}$ y $Y_{T+\cdot}$ son independientes entre si y de de X^T, Y^T y además, son cadenas de Markov con matriz de transición P que comienzan en y. Se sigue entonces que $(X_{T+\cdot}, Y_{T+\cdot})$ tiene la misma distribución que $(Y_{T+\cdot}, X_{T+\cdot})$. Vemos así que W tiene la misma distribución que el proceso $\tilde{W}$ definido por

$$\tilde{W}_n = \begin{cases} (X_n, Y_n) & n \leq T \\ (Y_n, X_n) & n \geq T \end{cases}.$$

Al comparar la primera coordenada, se deduce que Z tiene la misma distribución que X. $\square$

12. Probabilidad condicional regular

Ahora utilizaremos de manera más elaborada la idea de la función de cuantiles para obtener un teorema general de existencia de procesos estocásticos.

Supongamos que se tiene una variable aleatoria X definida en un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$. Sea $\mathscr{G}$ una subσ-álgebra de $\mathscr{F}$. La idea es construir una medida de probabilidad aleatoria $\mu(\omega, A)$ que represente a la distribución condicional de X dada $\mathscr{G}$. Formalmente, quisiéramos que para cualquier función $f : \mathbb{R} \to \mathbb{R}$ medible y acotada, la función

$$\omega \mapsto \int f(x)\ \nu(\omega, dx)$$

fuera una variable aleatoria y que fuera una versión de $\mathbb{E}(f(X) \,|\, \mathscr{G})$. Esto se puede lograr más o menos fácilmente.

Teorema 2.6. *Existe una función $\nu : \Omega \times \mathscr{B}_{\mathbb{R}} \to [0,1]$ tal que:*

1. *Para toda $\omega \in \Omega$, la función $A \mapsto \nu(\Omega, A)$ es una medida de probabilidad,*
2. *para toda $A \in \mathscr{F}$, la función $\omega \mapsto \nu(\omega, A)$ es una variable aleatoria $\mathscr{G}$-medible y es una versión de $\mathbb{P}(A \,|\, \mathscr{G})$.*

A una función como ν se le conoce como medida aleatoria en $(\Omega, \mathscr{G}, \mathbb{P})$.

Demostración. Para cada racional r, sea $F(\omega, r)$ una versión de la probabilidad condicional $\mathbb{P}(X \leq r \,|\mathscr{G})$. Por la monotonía de la esperanza condicional, existe un conjunto nulo N_{rs} gal que si $r \leq s$ son racionales entonces $F(\omega, r) \leq F(\omega, s)$ si $\omega \in N_{r,s}^c$. Por el teorema de convergencia dominada para la esperanza condicional, existe un conjunto N_r de probabilidad cero tal que si $\omega \in N_r^c$ se satisface

$$F(r, \omega) = \lim_{n \to \infty} F(\omega, r + 1/n) .$$

Finalmente, por el mismo teorema de convergencia dominada, existe un conjunto C de probabilidad cero tal que en C^c se tiene

$$0 = \lim_{r \to -\infty, r \in \mathbb{Q}} F(\omega, r) \quad \text{y} \quad \lim_{r \to \infty, r \in \mathbb{Q}} F(\omega, r) = 1.$$

Sea N la unión de los eventos N_{rs}, N_r y C sobre los racionales $r < s$. Entonces $N \in \mathscr{F}$ y $\mathbb{P}(N) = 0$. Ahora extendamos la definición de $F(\omega, t)$ para todo real t al definir

$$F(\omega, t) = \begin{cases} \inf_{r \geq t, r \in \mathbb{Q}} F(\omega, r) & \text{si } \omega \in N^c \\ F(t) & \text{si } \omega \in N \end{cases},$$

donde F es una función de distribución arbitraria. Por la elección del conjunto N, vemos que $t \mapsto F(\omega, t)$ es realmente una extensión a todo real de la función F definida sobre los racionales. Además es no-decreciente, continua por la derecha, y tiene límites 0 y 1 en $-\infty$ e ∞ respectivamente. Por lo tanto, para toda $\omega \in \Omega$ existe una medida $\mu(\omega, \cdot)$ sobre los borelianos de $\mathbb{R}$ tal que

$$\mu(\omega, (-\infty, t]) = F(\omega, t)$$

para toda $t \in \mathbb{R}$. La familia $\mathscr{C}$ de los borelianos A tal que $\mu(\cdot, A)$ es una variable aleatoria es claramente un λ-sistema que contiene al π-sistema de los conjuntos $(-\infty, t]$ con $t \in \mathbb{R}$, y la σ-álgebra generada por dichos conjuntos es $\mathscr{B}_{\mathbb{R}}$. Así, vemos que para todo $A \in \mathscr{B}_{\mathbb{R}}$, $\mu(\cdot, A)$ es una variable aleatoria.

Finalmente, por hipótesis vemos que para todo $G \in \mathscr{G}$ y todo $t \in \mathbb{R}$

$$\mathbb{E}(\nu(\cdot, (-\infty, t]) \, \mathbf{1}_G) = \mathbb{E}(\mathbf{1}_{X \leq t} \mathbf{1}_G) .$$

Por el lema de clases de Dynkin, vemos que

$$\mathbb{E}(\nu(\cdot, A)) = \mathbb{E}(\mathbf{1}_{X \in A} \mathbf{1}_G)$$

para todo $A \in \mathscr{B}_{\mathbb{R}}$. □

Ejercicio 2.4. Pruebe que para cualquier función medible y acotada f se tiene que

$$\omega \mapsto \int f(x) \, \nu(\omega, dx)$$

es una versión de $\mathbb{E}(\, f(X) \, |\mathscr{G})$.

13. El teorema de Kolmogorov para procesos estocásticos a tiempo discreto

Veamos ahora cómo probar el teorema de Kolmogorov, en el caso de sucesiones de variables aleatorias reales, al utilizar las ideas anteriores. Específicamente, consideraremos el siguiente resultado:

Teorema 2.7 (Caso particular del teorema de Kolmogorov). *Sean $\mu_1, \mu_2, \ldots$ medidas de probabilidad en $\mathbb{R}^1, \mathbb{R}^2, \ldots$ que satisfacen la condición de consistencia siguiente: para todo $A \in \mathscr{B}_{\mathbb{R}^n}$:*

$$\mu_{n+1}(A \times \mathbb{R}) = \mu_n(A)\,.$$

existe un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ y una sucesión de variables aleatorias $X_1, X_2, \ldots$ definidas en él tal que para toda n, la distribución conjunta de $X_1, \ldots, X_n$ es μ_n.

Demostración del teorema 2.7. Sabemos que existe un espacio de probabilidad y una sucesión de variables aleatorias independientes $(U_i)_{i\geq 1}$ con distribución uniforme en $(0,1)$, definidas en él. Sea F_1 la función de distribución asociada a μ_1 y ϕ_1 la función de cuantiles de F_1. Sea $X_1 = \phi_1(U_1)$, por lo que X_1 tiene distribución μ_1. Ahora procederemos de manera recursiva: supongamos que hemos definido a $(X_1, \ldots, X_n)$ como función de $U_1, \ldots, U_n$ y que la distribución de $(X_1, \ldots, X_n)$ es μ_n. Sea ν_n una probabilidad condicional regular de μ_{n+1} dadas las primeras n coordenadas. Sea $\phi_n(x_1, \ldots, x_n, u)$ la función que para $x_1, \ldots, x_n$ fijas, es igual a la función de cuantiles de la medida de probabilidad $\nu_n((x_1, \ldots, x_n), \cdot)$. Sea $X_{n+1} = \phi_n(X_1, \ldots, X_n, U_{n+1})$ y veamos que $(X_1, \ldots, X_{n+1})$ tiene distribución μ_n. Por definición de ν_n:

$$\begin{aligned}
&\mathbb{P}((X_1, \ldots, X_n) \in A, X_{n+1} \leq x) \\
&= \mathbb{P}((X_1, \ldots, X_n) \in A, \phi(X_1, \ldots, X_n, U_{n+1}) \leq x) \\
&= \mathbb{P}((X_1, \ldots, X_n) \in A, U_{n+1} \leq \nu_n(X_1, \ldots, X_n), (-\infty, x]) \\
&= \mathbb{E}\big(\mathbf{1}_{(X_1,\ldots,X_n)\in A}\nu_n(X_1, \ldots, X_n, (-\infty, x])\big) \\
&= \int_A \nu_n(x_1, \ldots, x_n, (-\infty, x])\ \mu_n(dx_1, \ldots, dx_n) = \mu_{n+1}(A \times (-\infty, x])\,.
\end{aligned}$$

Por lo tanto $(X_1, \ldots, X_{n+1})$ tiene distribución μ_n. □

14. El teorema de Kolmogorov para procesos estocásticos a tiempo continuo

En esta sección, se profundizará en el análisis de la anterior con el fin de construir procesos estocásticos cuyo índice recorra un intervalo de tiempo. Un **proceso estocástico a tiempo continuo** es una colección de variables aleatorias reales $(Y_t)_{t\in T}$ definidas en un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ donde $T \subset \mathbb{R}_+$ es conexo. Los casos que más nos interesarán son $T = [0, 1]$ y $T = [0, \infty)$ y, puesto que así lo decidió una moneda que lancé al aire, nos enfocaremos en el primero.

Lo primero que definiremos es el concepto de distribución de un proceso estocástico. Esta distribución estará definida en el espacio canónico donde está definido el proceso, que es $\mathbb{R}^T$. Informalmente, la distribución de $Y = (Y_t)_{t\in T}$, denotada μ_Y, es la imagen de $\mathbb{P}$ bajo la aplicación $\omega \mapsto Y(\omega)$ donde $Y(\omega)$ es la trayectoria $t \mapsto Y_t(\omega)$ con $t \in T$. La siguiente σ-álgebra se encarga automáticamente de todos los detalles de medibilidad que formalizan la definición de μ_Y: definamos las proyecciones $X_t : \mathbb{R}^T \to \mathbb{R}$ que mandan a $x : T \to \mathbb{R}$ en $x(t)$ (escribiremos simplemente x_t). Las proyecciones generan a la σ-álgebra $\mathscr{F} = \sigma(X_t : t \in T)$. Dicha σ-álgebra está generada por el álgebra

$$\mathscr{C} = \bigcup_{\substack{F\subset T\\ F \text{ finito}}} \sigma(X_t, t \in F)\,.$$

Es fácil ver que todo elemento de $\mathscr{C}$ es de la forma $\left\{(X_t)_{t\in F} \in B\right\}$ donde $F \subset T$ es finito y B es un boreliano de $\mathbb{R}^F$. Además, para ver que $Y : \Omega \to \mathbb{R}^T$ es medible si y sólo si $X_t \circ Y = Y_t$ es medible. Puesto que Y es un proceso estocástico, Y es medible. Por otra parte, a las medidas imagen μ_Y^F de $(Y_t)_{t\in F}$ bajo $\mathbb{P}$, donde $F \subset T$ es finito, les llamamos distribuciones finito-dimensionales de Y. Puesto que $\mathscr{C}$ es un álgebra que genera a $\mathscr{F}$, se sigue que cualquier medida en $\mathbb{R}^T$ está determinada por sus valores en $\mathscr{C}$. Dicho de otra manera, la distribución de un proceso queda determinada por sus distribuciones finito-dimensionales. El teorema de Kolmogorov se encarga del problema de, dada una familia de medidas sobre $\mathbb{R}^F$, con $F \subset T$ finito, caracterizar cuándo son las distribuciones finito-dimensionales de un proceso estocástico indexado por T. Notemos que si $F_1 \subset F_2$ son subconjuntos finitos de T y π_{F_2,F_1} es la proyección de $\mathbb{R}^{F_2}$ en $\mathbb{R}^{F_1}$, entonces se satisface la condición de consistencia

$$\mu_Y^{F_2} \circ \pi_{F_2,F_1}^{-1} = \mu_Y^{F_1}\,.$$

El teorema de Kolmogorov afirma que de hecho la condición de consistencia es necesaria y suficiente para que una colección de medidas de probabilidad $\mu^F : \mathscr{B}_{\mathbb{R}^F} \to [0, 1]$, con $F \subset T$ finito, sean las distribuciones finito-dimensionales de un proceso estocástico. Por supuesto, la parte importante de este teorema es la existencia del proceso estocástico en cuestión. La construcción se hace en el espacio canónico: sea $\mu : \mathscr{C} \to [0, 1]$ dada por

$$\mu\big(\left\{(X_t)_{t\in F} \in B\right\}\big) = \mu^F(B)\,.$$

Note que el hecho de que las medidas $\left\{\mu^F\right\}$ sean consistentes implica que μ está bien definida.

Teorema 2.8 (Teorema fundamental (o de consistencia) de Kolmogorov). *La función μ se extiende a una medida de probabilidad (también denotada μ) en $\mathscr{F}$. La distribución de $(X_t)_{t\in F}$ bajo $\mathbb{P}$ es precisamente μ^F.*

Demostración. Supongamos primero que $S \subset T$ es numerable. Por la primera versión que probamos del teorema de Kolmogorov, sabemos que existe un espacio de probabilidad y un procesos estocástico $Y_t, t \in S$ cuyas distribuciones finito dimensionales son $\{\mu_F : F \subset S, F \text{ finito}\}$. Si ahora consideramos la función $Y^S : \Omega \to \mathbb{R}^T$ que a ω le asigna la función $t \mapsto Y_t(\omega) \mathbf{1}_{t \in S}$, la imagen de $\mathbb{P}$ se convierte en una medida de probabilidad μ^S.

Notemos que si $A \in \mathscr{A}$ es de la forma $\{(X_t, t \in F) \in B\}$ con $F \subset S$ (dicho de otra manera $A \in \mathscr{F}_F$) entonces $\mu^S(A) = \mu(A)$. Así, si $S_1 \subset S_2 \subset T$ son numerables y $A \in \mathscr{F}_{S_1}$ entonces $\mu^{S_1}(A) = \mu^{S_2}(A)$. Esto implica que si S_1, S_2 son dos conjuntos numerables y $A \in \mathscr{F}_{S_1} \cap \mathscr{F}_{S_2}$ entonces

$$\mu^{S_1}(A) = \mu^{S_1 \cup S_2}(A) = \mu^{S_2}(A) .$$

Así, podemos definir a $\tilde{\mu}$ en $\cup_S \mathscr{F}_S$ mediante $\mu(A) = \mu^S(A)$ si $A \in \mathscr{F}_S$. Sin embargo, notemos que $\cup_S \mathscr{F}_S$ es una σ-álgebra que contiene a $\mathscr{C}$ y está contenida en $\mathscr{F}$, por lo que es de hecho igual a $\mathscr{F}$. La medida $\tilde{\mu}$ es la extensión buscada de μ a $\mathscr{F}$. □

CAPÍTULO 3

Procesos de renovación

Consideremos variables aleatorias independientes e idénticamente distribuidas $S_1, S_2, \ldots$ con valores estríctamente postivos. Podemos pensar en que S_i es el tiempo de vida de un componente crucial para el funcionamiento de cierto sistema y que al fallar se debe reemplazar. Los tiempo de reemplazo serán

$$T_0 = 0, \quad T_1 = S_1, \quad T_2 = S_1 + S_2, \ldots.$$

Por otra parte la cantidad de componentes que han fallado hasta el tiempo t sería

$$N_t = \min \{n : T_n > t\} = \sum_{n=0}^{\infty} n \mathbf{1}_{T_n \leq t < T_{n+1}}.$$

A este modelo general se le llama modelo de renovación. La sucesión S será la **sucesión de tiempos de vida**, la sucesión T la de **tiempos de renovación** y la sucesión N será el **proceso de contéo asociado**.

Hay un par de procesos adicionales que son útiles e interesantes: el proceso de **tiempo residual** (hasta el próximo reemplazo) es

$$R_t = T_{N_t+1} - t,$$

el **proceso de edad** es

$$A_t = t - T_{N_t}.$$

Su suma es el **proceso de tiempos totales**

$$L_t = S_{N_t+1}.$$

En la Figura 1 se ilustran las definiciones asociadas a los procesos de renovación.

Imaginemos ahora que en vez de cambiar al componente crucial en cuanto falla, se realiza una revisión diaria en la que se ve si se cambia o no. Entonces es más conveniente medir el tiempo en días en vez de continuamente. En términos del modelo, se puede imponer simplemente que el tiempo de vida S_i sea una variable aleatoria con valores en $\{1, 2, \ldots\}$ y las definiciones tienen sentido como las hemos puesto, salvo que R_n, A_n y L_n, con $n \in \mathbb{N}$, toman valores en $\mathbb{N}$ y determinan a los procesos R, A y L. En este caso hablamos de un **proceso de renovación aritmético**.

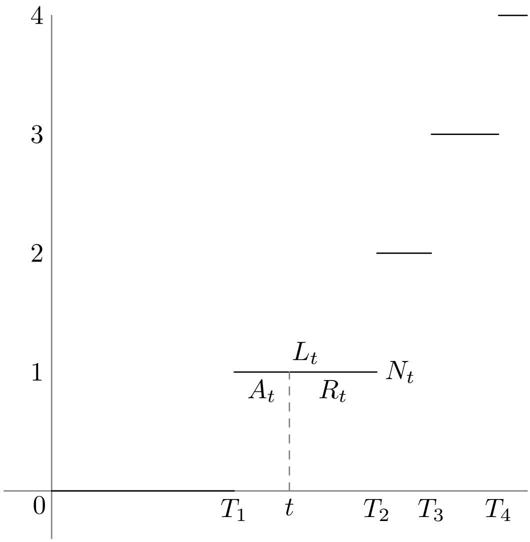

Figura 1. Ilustración de las definiciones de proceso de renovación

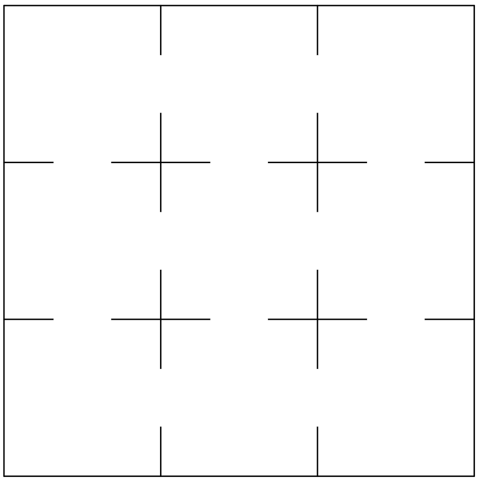

Figura 2. Laberinto para un experimento aleatorio

Ejemplo 3.1. Imaginemos la trayectoria de una rata en el laberinto de la Figura 2. Para modelar la trayectoria que sigue la rata, supongamos que cuando se encuentra en un cuarto del laberinto, la rata va a cualquier otro con la misma probabilidad. Podemos entonces modelar esta situación mediante una cadena de

Markov: enumeremos los cuartos del laberinto de izquierda a derecha, de abajo a arriba, por lo que la rata comienza en el cuarto 1 y encuentra la comida en el cuarto 9. Definamos $P_{i,j}$ como la probabilidad con que la rata pasa del cuarto i al cuarto j:

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

Supongamos que inicialmente le damos de comer a la rata en la casilla central y que, al terminar, la rata se va a dar un paseo aleatorio. Cuando regresa a la casilla central, encontrará la comida de nuevo dispuesta. En este caso, nuestros tiempos entre sucesos S_i serán la cantidad de pasos entre dos comidas de la rata y nuestros tiempos de reemplazo (o de renovación) serán los instantes de las visitas sucesivas de la rata a la casilla central. La variable R_n se puede interpretar como la cantidad de tiempo que le falta a la rata, después de n pasos, para volver a comer. La variable A_n es la cantidad de tiempo que lleva la rata sin comer al paso n mientras que L_n es la cantidad total de pasos que pasará la rata sin comer desde la última vez que comió anterior al paso n, hasta la siguiente vez que lo hará. No es completamente trivial verificar que la situación descrita corresponde a un fenómeno de renovación, pues no hemos discutido por qué los tiempos entre las visitas sucesivas de la rata a la casilla central conforman una sucesión de variables independientes e idénticamente distribuidas. Sin embargo, la propiedad de Markov fuerte nos permite ver que así es.

El tipo de preguntas a las que responde la teoría de renovación en este contexto son las siguientes: Al instante n: ¿cuántas veces ha comido la rata? ¿ Qué pasa conforme $n \to \infty$? ¿ Qué pasa en promedio? ¿ Cuál es la distribución del tiempo que le falta para volver a comer (R_n)? ¿ Qué le pasa a esta distribución conforme $n \to \infty$? ¿Cuál es la probabilidad de que al paso n la rata coma? ¿Qué pasa con dicha probabilidad conforme $n \to \infty$?

El ejemplo anterior es caso particular de uno mucho más general: si $0 = T_0 < T_1 < \cdots$ son los instantes sucesivos en que una cadena de Markov recurrente visita a su estado inicial, que fijamos igual a x, entonces las variables $T_i - T_{i-1}$ son independientes e idénticamente distribuidas por lo que conforman un fenómeno de renovación. Si $X_0 = y \neq x$, entonces la distribución de S_1 es distinta a la de $S_2, S_3, \ldots$, aunque aún así son iid. En este caso hablamos de un proceso de renovación demorado.

Ejemplo 3.2 (El proceso Bernoulli). Se trata de un proceso de renovación aritmético en el cual los tiempos entre sucesos tienen distribución geométrica:

$$\mathbb{P}(S_i = k) = p\,(1-p)^{k-1}$$

para $k = 1, 2, \ldots$. En este caso particular se pueden calcular las distribuciones de los procesos asociados al fenómeno de renovación. Además, admite una interpretación adicional: sean $B_1, B_2, \ldots$ variables aleatorias Bernoulli de parámetro p y sean

$$T_0 = 0 \quad \text{y} \quad T_{n+1} = \min\left\{i > T_n : B_i = 1\right\}.$$

Entonces T es un proceso Bernoulli en el sentido de que la sucesión $(T_{n+1} - T_n)$ es iid con distribución geométrica (concentrada en $\{1, 2, \ldots\}$) de parámetro p.

Comencemos con la distribución de T_n: de acuerdo a nuestra interpretación, T_n es el tiempo en el que ocurre el enésimo éxito de la sucesión Bernoulli B, por lo que T_n es binomial negativa de parámetros n y p con valores en $\{n, n+1, \cdots\}$, es decir:

$$\mathbb{P}(T_n = k) = \binom{k-1}{k-n} p^k\,(1-p)^{k-n}.$$

Al proceso de contéo asociado también se le puede calcular la distribución exacta: notemos que N_n es la cantidad de unos en la sucesión $B_1, \ldots, B_n$, y como estas variables toman los valores cero y uno, pues $N_n = \sum_{i=1}^{n} B_i$. Así, N_n tiene distribución binomial de parámetros n y p. Esto además implica que $\mathbb{E}(N_n) = p/n$ y por lo tanto $\mathbb{E}(N_n/n) \to p$ conforme $n \to \infty$, que es un caso particular del teorema de renovación elemental que demostraremos más adelante.

Respecto al proceso de tiempos resiguales, notemos que

$$\{R_n = k\} = \{B_{n+1} = 0, \ldots, B_{n+k-1} = 0, B_{n+k} = 1\}$$

de donde concluimos que

$$\mathbb{P}(R_n = k) = p\,(1-p)^{k-1} \quad k = 1, 2, \ldots.$$

En otras palabras, R_n es geométrica de parámetro p con valores en $\{1, 2, \ldots\}$.

Un argumento similar funciona para el proceso de edad pues

$$\{A_n = k\} = \{B_n = 0, B_{n-1} = 0, \ldots, B_{n-k+1} = 0, B_{n-k} = 1\}$$

por lo que

$$\mathbb{P}(A_n = k) = p\,(1-p)^{k} \quad k = 0, 1, \ldots, n-1.$$

El caso $k = n$ es diferente pues

$$\{A_n = n\} = \{B_n = 0, B_{n-1} = 0, \ldots, B_1 = 0\}$$

por lo que

$$\mathbb{P}(A_n = n) = (1-p)^n.$$

En otras palabras, A_n tiene la misma distribución de $S_1 \wedge n$. Cabe destacar que conforme $n \to \infty$, A_n converge en distribución a una variable geométrica. Además, las variables A_n y R_n son independientes pues al ser las B_i independientes vemos que

$$\begin{aligned}
&\mathbb{P}(A_n = j, R_n = k) \\
&= \mathbb{P}(B_n = 0, \ldots, B_{n-j-1} = 0, B_{n-j} = 1, B_{n+1} = 0, \ldots, B_{n+k-1} = 0, B_{n+k} = 1) \\
&= \mathbb{P}(B_n = 0, \ldots, B_{n-j-1} = 0, B_{n-j} = 1)\, \mathbb{P}(B_{n+1} = 0, \ldots, B_{n+k-1} = 0, B_{n+k} = 1) \\
&= \mathbb{P}(A_n = j)\, \mathbb{P}(R_n = k)
\end{aligned}$$

Finalmente, podemos ver cuál será el límite a tiempos grandes del proceso de tiempos totales. Recordemos que la suma de dos geométricas independientes es binomial negativa. (Sólo hay que tener cuidado pues este cálculo asume que ambas variables geométricas toman valores en el mismo conjunto). El resultado es que

$$\lim_{n\to\infty} \mathbb{P}(L_n = m) = mp^2 (1-p)^{m-1}. \quad m = 1, 2, \ldots$$

A continuación exploraremos otra liga entre cadenas de Markov y procesos de renovación.

Proposición 3.1. *En procesos de renovación aritméticos, el proceso de tiempos residuales es una cadena de Markov. Si M es el supremo del soporte de la distribución de S_1, la matriz de transición de R es irreducible en $\{i \in \mathbb{N} : i \leq M\}$. R es aperiódica si $\{n \geq 1 : \mathbb{P}(S_1 = n) > 0\} \not\subset h\mathbb{N}$ para toda $h \geq 2$.*

Un proceso de renovación aritmético aperiódico es aquel para el cual se satisface la anterior condición para que R sea aperiódica.

Demostración. Consideremos a la función

$$f(i, r) = \begin{cases} i - 1 & i > 1 \\ r & i = 1 \end{cases}.$$

Al definir a la sucesión $\tilde{R}$ medante

$$\tilde{R}_0 = 1 \quad \text{y} \quad \tilde{R}_{n+1} = f\left(\tilde{R}_n, S_{n+1}\right),$$

vemos que $\tilde{R}$ es una cadena de Markov con matriz de transición

$$P_{i,j} = \begin{cases} 1 & i > 1, j = i - 1 \\ \mathbb{P}(S_1 = j) & i = 1 \end{cases}.$$

Sin embargo, puesto que la sucesión S es iid, R y $\tilde{R}$ tienen la mismas distribuciones finito-dimensionales en el sentido siguiente:

$$\mathbb{P}(R_0 = i_0, \ldots, R_n = i_n) = \mathbb{P}\left(\tilde{R}_0 = i_0, \ldots, \tilde{R}_n = i_n\right).$$

En efecto, si $i_0, \ldots, i_n \in \{1, 2, \ldots\}$ e $i_l = 1$ si y sólo si $l \in I \subset \{0, \ldots, n\}$ e $i_l = i_{l-1} - 1$ si $l \notin I$ y digamos que $I = \{l_1, \ldots, l_m\}$ entonces

$$\begin{aligned}\mathbb{P}\Big(\tilde{R}_0 = i_0, \ldots, \tilde{R}_n = i_n\Big) &= \mathbb{P}\Big(S_{i_{l_k}} = i_{l_k}, k \leq m\Big) \\ &= \mathbb{P}(S_k = i_{l_k}, k \leq m) \\ &= \mathbb{P}(R_0 = i_0, \ldots, R_n = i_n)\,.\end{aligned}$$

Esto prueba que R es una cadena de Markov con matriz de transición P.

Si M es el supremo del soporte de la distribución de S_1 y $M < \infty$, entonces $\mathbb{P}(S_1 = M) > 0$ y $\mathbb{P}(S_1 = n) = 0$ para toda $n \geq M$. Entonces de 1 se accede a M, a partir del cuals e accede a $M - 1, \ldots, 1$, por lo que $\{0, \ldots, M\}$ es una clase de comunicación, que de hecho es cerrada pues $\mathbb{P}(S_1 = n) = 0$ para toda $n \geq M$. Esto hace que la cadena en $\{1, \ldots, M\}$ sea irreducible. Si $M = \infty$, entonces para toda M existe $n \geq M$ tal que $\mathbb{P}(S_1 = n) > 0$, por lo que 0 se comunica con M vía n. Es decir, la cadena es irreducible en $\{0, 1, \ldots\}$.

Si $\{n \geq 1 : \mathbb{P}(S_1 = n) > 0\} \not\subset h\mathbb{N}$ para $h \geq 2$ entonces existen n_1, n_2, k_1 y k_2 naturales tal que $n_1 k_1 = 1 + n_2 k_2$ y tales que $\mathbb{P}(S_1 = n_1)\,, \mathbb{P}(S_1 = n_2) > 0$. Vemos entonces que es posible ir de cero a cero en $n_1 k_1 = n_2 k_2 + 1$ pasos y en $n_2 k_2$ pasos, por lo que el periodo de 1 es 1 y por lo tanto la cadena es aperiódica. □

La respuesta a cuántas veces a comido la rata al paso n se puede interpretar en términos de N_n en el caso de procesos de renovación aritméticos o de N_t en el caso general. Los siguientes dos resultados nos permiten dar una posible respuesta: para tiempos grandes, la variable N_t se puede predecir determinísticamente.

Proposición 3.2 (Ley fuerte de los grandes números). *Si $\mu = \mathbb{E}(S_i) < \infty$ entonces $N_t/t \to 1/\mu$ casi seguramente*

Demostración. Por la ley fuerte de los grandes números, sabemos que $T_n/n \to \mu$ casi seguramente. Notemos que si $t \in [T_n, T_{n+1})$ entonces

$$\frac{T_n}{n}\frac{n}{n+1} \leq \frac{t}{N_t} \leq \frac{T_{n+1}}{n+1}\frac{n+1}{n}.$$

Los extremos de la desigualdad convergen al mismo límite, μ, conforme $n \to \infty$ de manera casi segura. Por lo tanto $N_t/t \to 1/\mu$. □

Nuestro siguiente resultado involucra a la llamada función de renovación. Es la función $m : [0, \infty) \to [0, \infty)$ dada por $m(t) = \mathbb{E}(N_t)$.

Proposición 3.3 (Teorema de renovación elemental). *Si $\mu = \mathbb{E}(S_1) < \infty$ entonces $m(t)\,/t \to 1/\mu$.*

La demostración utilizará un resultado sobre camintas aleatorias conocido como la identidad de Wald y que se prueba como una aplicación de la teoría de martingalas. Se trata de una generalización de la fórmula $\mathbb{E}(T_n) = n\mathbb{E}(T_1)$ al caso en que n ya no es una cantidad fija sino aleatoria. En nuestro contexto, veremos que

$$\mathbb{E}(T_{N_t+1}) = \mathbb{E}(N_t+1)\,\mathbb{E}(T_1) = \mathbb{E}(N_t+1)\,\mu. \tag{10}$$

Puesto que

$$\{N_t+1=n\} = \{N_t = n-1\} = \{T_{n-1} \leq t < T_n\},$$

vemos que N_t+1 es un tiempo de paro respecto de la filtración canónica asociada a T y por lo tanto el teorema de muestreo opcional de Doob, si se puede aplicar, nos da la ecuación (10). La prueba de su validez es sencilla y por lo tanto se hace directamente: notamos que $\{N_t+1>n\}$ pertenece a la σ-álgebra generada por $S_1,\dots S_n$ y es por lo tanto independiente de S_{n+1}. Al aplicar el teorema de Tonelli (dos veces):

$$\begin{aligned}
\mathbb{E}(T_{N_t+1}) &= \mathbb{E}\left(\sum_i \mathbf{1}_{i=N_t+1} T_i\right)\\
&= \mathbb{E}\left(\sum_i \sum_{j\leq i} \mathbf{1}_{i=N_t+1} S_j\right)\\
&= \mathbb{E}\left(\sum_j \sum_{i\geq j} \mathbf{1}_{i=N_t+1} S_j\right)\\
&= \mathbb{E}\left(\sum_j \mathbf{1}_{N_t+1\geq j} S_j\right)\\
&= \sum_j \mathbb{E}(S_j)\,\mathbb{P}(N_t+1\geq j)\\
&= \mu \sum_j \mathbb{P}(N_t+1\geq j)\\
&= \mu\mathbb{E}(N_t+1)\,.
\end{aligned}$$

Prueba del teorema de renovación elemental. Ocupémonos primero de obtener una cota inferior. A partir de la identidad de Wald, vemos que

$$t \leq \mathbb{E}(T_{N_t+1}) \leq \mathbb{E}(N_t+1)\,\mu.$$

Por lo tanto

$$\frac{1}{\mu} - \frac{1}{t} \leq \frac{\mathbb{E}(N_t)}{t}.$$

Notemos que

$$\lim_{t\to\infty} \frac{1}{\mu} - \frac{1}{t} = \frac{1}{\mu}.$$

Para obtener una cota superior, definimos a $\tilde{S}_i = S_i \wedge b$. Con la notación obvia, notemos que $\tilde{T}_i \leq T_i$ y que por lo tanto $N_t \leq \tilde{N}_t$. Además,

$$\mathbb{E}\Big(\tilde{S}_1 \wedge b\Big)\,\mathbb{E}(N_t + 1) = \mathbb{E}\Big(\tilde{T}_{\tilde{N}_t+1}\Big) \leq t + b,$$

por lo que

$$\frac{\mathbb{E}(N_t)}{t} \leq \frac{t+b}{t\mathbb{E}(S_1 \wedge b)}.$$

Al utilizar $b = \sqrt{t}$, el teorema de convergencia monótona nos dice que

$$\mathbb{E}\Big(S_1 \wedge \sqrt{t}\Big) \to \mu,$$

por lo que

$$\lim_{t\to\infty} \frac{t+\sqrt{t}}{t\mathbb{E}\big(S_1 \wedge \sqrt{t}\big)} = \frac{1}{\mu}.$$

Podemos entonces deducir que

$$\lim_{t\to\infty} \frac{\mathbb{E}(N_t)}{t} = \frac{1}{\mu}. \qquad \square$$

1. La medida de renovación en el caso aritmético

Continuaremos el estudio de procesos de renovación al introducir a la medida de renovación. En realidad se introducirá a partir de la función u dada por

$$u_n = \mathbb{P}(\exists m, T_m = n) = \sum_m \mathbb{P}(T_m = n)\,.$$

En términos de la rata en el laberinto, u_n es la probabilidad de que la rata coma al instante n. El comportamiento asintótico de u_n se conoce y es el objeto del teorema de renovación para procesos de renovación aritméticos de Erdős, Feller y Pollard.

Teorema 3.1. *Para un proceso de renovación aritmético, aperiódico y con media finita:*

$$\lim_{n\to\infty} u_n \to \frac{1}{\mu}.$$

Demostración. Primero probemos que el proceso de tiempos residuales R tiene una distribución invariante. En efecto, sean

$$\pi_i = \mathbb{P}(S_1 \geq i)\,/\mu$$

y P la matriz de transición de R. Entonces

$$\begin{aligned}
(\pi P)_j &= \sum_i \pi_i P_{i,j} \\
&= \pi_{j+1} P_{j+1,j} + \pi_1 P_{1,j} \\
&= \frac{\mathbb{P}(S_1 \geq j+1)}{\mu} P_{j+1,j} + \frac{1}{\mu}\mathbb{P}(S_1 = j) \\
&= \frac{\mathbb{P}(S_1 \geq j)}{\mu} = \pi_j .
\end{aligned}$$

Al ser la cadena irreducible y aperiódica y con una distribución invariante, es positivo recurrente y se puede aplicar el teorema fundamental de convergencia para concluir que

$$\lim_{n\to\infty} P^n_{i,1} = \frac{1}{\mu}.$$

Por otra parte, podemos calcular explícitamente $P^n_{1,1}$ pues

$$P^n_{1,1} = \mathbb{P}(R_n = 1) = \mathbb{P}(n = T_m \text{ para alguna } m) = u_n.$$

Así, vemos que $u_n \to 1/\mu$ conforme $n \to \infty$. $\square$

2. La ecuación de renovación en el caso aritmético

La ecuación de renovación aparece en el cálculo de algunas cantidades de la teoría de renovación. En general se trata de un análisis al primero instante de renovación al notar que el proceso T' dado por $0, T_2 - T_1, T_3 - T_1, \cdots$ es un proceso de renovación idéntico en distribución a $T_0, T_1, T_2, \cdots$ e independiente de T_1. Ejemplos de su utilidad es que nos permite estudiar el comportamiento asintótico de las distribuciones de los procesos de tiempo residual, de edad ó de tiempos totales. Sin embargo, también se puede entender como una relación de recurrencia para cantidades de interés en teoría de renovación. Nos concentraremos en el caso aritmético, al ser técnicamente menos complicado, y comenzaremos con algunos ejemplos. Escribiremos

$$p_k = p(k) = \mathbb{P}(S_1 = k) .$$

Ejemplo 3.3. Sea u_k la densidad de renovación para un proceso de renovación aritmético:

$$u_k = \mathbb{P}(\exists n \geq 0 \text{ tal que } T_n = k) .$$

Escribiremos a dicha cantidad como $u(k)$ cuando la tipografía lo requiera.

Notemos que $u_0 = 1$, mientras que para $k \geq 1$, claramente $\mathbb{P}(T_0 = k) = 0$ y por lo tanto:

$$\begin{aligned}
u_k &= \mathbb{P}(\exists n \geq 1 \text{ tal que } T_n = k) \\
&= \sum_{j=1}^{k} \mathbb{P}(S_1 = j, \exists n \geq 0 \text{ tal que } T_n - j = k - j) \\
&= \sum_{j=1}^{k} \mathbb{P}(S_1 = j, \exists n \geq 0 \text{ tal que } T'_n = k - j) \\
&= \sum_{j=1}^{k} p_j \mathbb{P}(\exists n \geq 0 \text{ tal que } T_n = k - j) \\
&= \sum_{j=1}^{k} p_j u_{k-j} \\
&= \mathbb{E}(u(k - S_1)),
\end{aligned}$$

donde por supuesto se utiliza el hecho que por definición $u_j = 0$ si $j < 0$.

La ecuación

$$u_k = \delta_0(k) + \mathbb{E}(u(k - S_1))$$

que también se escribe

$$u_k = \delta_0(k) + \sum p_j u_{k-j}$$

es la llamada ecuación de renovación para la densidad de renovación.

Ejemplo 3.4. Sea $L_n, n \geq 0$ el proceso de tiempos totales de un proceso de renovación aritmético. Para $r \geq 0$, sea $z(n) = \mathbb{P}(L_n = r)$ para $n \geq 0$ y $z(n) = 0$ si $n < 0$. Notemos que si $n \leq r$ entonces

$$z(n) = \mathbb{P}(S_1 = r) = p_r.$$

Por otra parte, si $n > r$, nos fijamos en la primera renovación $T_1 = S_1$, que si $L_n = r$ es necesariamente es más chica que n. Entonces se obtiene:

$$z(n) = \sum_{j \leq n} \mathbb{P}(S_1 = j, L_{n-j}(T') = r) = \sum_{j \leq n} p_j z(n-j).$$

La ecuación de renovación resultante es

$$z(n) = \mathbf{1}_{n \leq r} p_r + \sum_j p_j z(n-j).$$

Ejemplo 3.5. Sea R_n el proceso de tiempos residuales de un proceso de renovación aritmético. Sea $z(n) = \mathbb{P}(R_n = r)$ si $n \geq 0$ y $z(n) = 0$ si $n < 0$. El cálculo de R_n se puede dividir en dos casos, dependiendo de si $S_1 \leq n$ o no:

$$z(n) = \mathbb{P}(R_n = r) = p_{n+r} + \sum_{j=1}^{n} \mathbb{P}(S_1 = r)\, z(n-r)\,.$$

En general, la ecuación de renovación es la siguiente: dada una función b (tal que $b_n = 0$ si $n < 0$) se busca una función z tal que $z(n) = 0$ si $n < 0$ y

$$z(n) = b(n) + \sum_{j \leq n} p_k z(n-j)\,.$$

La solución se puede encontrar al iterar la ecuación de renovación para escribir:

$$\begin{aligned} z(n) &= b(n) + \mathbb{E}(z(n - S_1)) \\ &= b(n) + \mathbb{E}(b(n - S_1)) + \mathbb{E}(b(n - S_1 - S_2)) \\ &= b(n) + \mathbb{E}(b(n - S_1)) + \mathbb{E}(b(n - S_1 - S_2)) + \mathbb{E}(b(n - S_1 - S_2 - S_3)) = \cdots\,. \end{aligned}$$

Puesto que $b(n - T_m) = 0$ si $m \geq n$, vemos que

$$z(n) = \sum_{m} \mathbb{E}(b(n - T_m))$$

y al sumar sobre los posibles valores de T_m, se obtiene

$$z(n) = \sum_{m} \sum_{x} b(n-x)\, \mathbb{P}(T_n = x) = \sum_{x} u_x b_{n-x}.$$

Una vez establecida la relación entre soluciones a la ecuación de renovación y la densidad de renovación, se tienen los elementos clave para probar el siguiente resultado.

Teorema 3.2 (Teorema clave de renovación en el caso discreto). *Si z es solucin a la ecuación de renovación*

$$z(n) = b(n) + \sum_{j \leq n} p_k z(n-j)\,.$$

y b es sumable entonces z está dada por

$$z(n) = \sum_{x} b(x)\, u(n-x)$$

y su comportamiento asintótico está caracterizado por

$$z(n) \to \sum_{x} b(x) \,/ \mu$$

conforme $n \to \infty$.

Demostración. Sólo hace falta verificar el comportamiento asintótico de z. Sin embargo, si b es sumable, puesto que u_n es una probabilidad, se tiene la cota

$$\sum_x b(x)\, u(n-x) \leq \sum_x b(x) < \infty$$

y por lo tanto el teorema de convergencia dominada y el teorema de renovación de Erdős-Feller-Pollard (EFP) nos dicen que

$$\lim_{n\to\infty} z(n) = \lim_{n\to\infty} \sum_x b(x)\, u(n-x) = \sum_x b(x)\,/\mu. \qquad \square$$

Veamos ahora algunas aplicaciones del teorema de renovación clave. En el caso del tiempo total asintótico, vemos que

$$\lim_{n\to\infty} \mathbb{P}(R_n = r) = \sum_x p_{r+x}/\mu = \mathbb{P}(S_1 \geq r)\,/\mu,$$

aunque en realidad esto ya lo sabíamos y lo utilizamos en la prueba del teorema de renovación de EFP.

Un ejemplo más interesante es el del tiempo total:

$$\lim_{n\to\infty} \mathbb{P}(L_n = r) = \sum_{x\leq r} p_x/\mu = rp_r/\mu.$$

3. La medida de renovación en el caso no-aritmético

Consideremos un fenómeno de renovación con tiempos interarribo S_1, S_2, ... Nos concentraremos en el caso no-aritmético en el que $\mathbb{P}(S_i \in h\mathbb{Z}_+) < 1$ para toda $h > 0$.

Ejemplo 3.6. Si suponemos que S_i admite una densidad entonces claramente estamos en el caso no aritmético. Otro ejemplo es cuando $\mathbb{P}(S_i < \varepsilon)$ para toda $\varepsilon > 0$, en particular si la distribución de S_i es de la forma $\sum_i a_i \delta_{b_i}$ donde $s_i \geq 0$, $\sum a_i = 1$ y b_i decrece a cero. Un ejemplo bastante interesante del caso no-aritmético, que sirve sobre todo para comprender las dificultades en la prueba del teorema principal de esta sección, es cuando la distribución de S_i está concentrada en dos valores inconmensurables, por ejemplo 1 y $\sqrt{2}$.

Hay un proceso adicional de interés: el proceso de renovación demorado. Imaginemos que no observamos al proceso de renovación desde el instante inicial sino desde el instante $t > 0$. Entonces la primer renovación ocurrirá en $T_{N_t+1} - t$ unidades de tiempo, mientras que los siguientes tiempos interarribo serán $S_{N_t+2}, S_{N_t+3}, \ldots$. El siguiente ejercicio motiva entonces la definición de proceso de renovación demorado.

Ejercicio 3.1. Pruebe que $T_{N_t+1}-t, S_{N_t+2}, S_{N_t+3}, \ldots$ son variables aleatorias independientes y que $S_{N_t+2}, S_{N_t+3}, \ldots$ tienen la misma distribución que S_1.

Pruebe que si S_1 tiene distribución exponencial entonces $T_{N_t+1} - t$ también tiene distribución exponencial y por lo tanto los procesos N y N^t dado por $N^t_s = N_{t+s} - N_t$ tienen la misma distribución.

El proceso de contéo con tiempos entre sucesos exponenciales se conoce como **proceso de Poisson** y es un caso particular tan interesante que dedicaremos un capítulo completo a su estudio.

Definición. Un **proceso de renovación demorado** asociado a la sucesión $(S_i, i \geq 1)$ es aquél cuyos tiempos de renovación son $\tilde{T}_0 = S_0$, $\tilde{T}_n = S_0 + T_n$ y S_0 es independiente de $S_1, S_2, \ldots$. Dicho proceso de renovación tiene demorado tiene definido a su proceso de contéo $\tilde{N}$ y demás procesos asociados.

Un **proceso de renovación estacionario** es un proceso de renovación demorado para el cual $\tilde{T}_{\tilde{N}_t+1}-t, \tilde{S}_{\tilde{N}_t+2}, S_{\tilde{N}_t+3}, \ldots$ son independientes e idénticamente distribuidas.

En otras palabras, un proceso de renovación estacionario es aquél proceso de renovación demorado tal que el proceso de contéo trasladado $\tilde{N}^t$ tiene la misma distribución que $\tilde{N}$ ó aún aquél que satisface que $\tilde{R}_t$ tiene la misma distribución que S_0 para cualquier $t > 0$. Notemos además que el proceso de renovación sin demora asociado a uno demorado se obtiene del proceso demorado al trasladar el tiempo al primer tiempo de renovación.

Pasaremos ahora a una construcción general de un proceso de renovación estacionario asociado a un proceso de renovación. Está basado en el análisis que hemos realizado para el caso aritmético, en el que obtenemos una medida invariante para el proceso de tiempos residuales. Supongamos que los tiempos entre sucesos del proceso de renovación $S_1, S_2, \ldots$ son integrables y sea μ su media. Definamos al proceso de renovación demorado en el que

$$\mathbb{P}\left(\tilde{S}_0 \in dx\right) = \frac{\mathbb{P}(S_1 \geq x)}{\mu} dx, \quad \mathbb{P}\left(\tilde{S}_1 \leq x\right) = \frac{\mathbb{E}(S_1 \wedge x)}{\mu}.$$

Se afirma que el proceso de renovación demorado es estacionario. Dividiremos la verificación de este hecho en dos partes; primero mostraremos que

$$\tilde{V}(A) = \mathbb{E}\left(\sum_{n=0}^{\infty} \mathbf{1}_{T_n \in A}\right) = \sum_{n=0}^{\infty} \mathbb{P}(T_n \in A) \quad \text{y} \quad \tilde{U}_t = \tilde{U}([0,t])$$

entonces

$$\tilde{U}_t = \frac{t}{\mu}.$$

A $\tilde{U}$ se le conoce como la medida de renovación del proceso demorado. Mostraremos que es un múltiplo de la medida de Lebesgue por un argumento de renovación al introducir a la medida de renovación del proceso sin demora

$$U(A) = \tilde{V}(A) = \mathbb{E}\left(\sum_{n=0}^{\infty} \mathbf{1}_{T_n \in A}\right) \quad \text{y} \quad U_t = U([0,t]) = 1 + \mathbb{E}(N_t)\,.$$

Sea F la distribución de S_1 y G la distribución de S_0 dada por

$$G_t = \int_0^t \frac{1 - F(r)}{\mu}\, dr.$$

Al condicionar respecto del valor de S_1 y de S_0, notamos que se satisfacen las siguientes identidades de convolución que relaciona a las funciones U, $\tilde{U}$, F y G:

$$U = \mathbf{1}_{(0,\infty)} + U * F \quad \text{y} \quad \tilde{U} = U * G.$$

Si en la segunda igualdad sustituimos la primera se obtiene

$$\tilde{U} = U * G = (1 + U * F) * G = G + \tilde{U} * F,$$

por lo que $\tilde{U}$ satisface la versión continua de la ecuación de renovación que ya hemos analizado. Al iterar la ecuación obtenemos de nuevo

$$\tilde{U} = G * (1 + F + F^* + \cdots + F^{*n}) + \tilde{U} * F^{*(n+1)}.$$

Puesto que $F^{*n}(t) \to 0$ conforme $n \to \infty$, se sigue que $\tilde{U} * F^{*(n+1)} \to 0$ y por lo tanto

$$\tilde{U} = G * U.$$

Si $\tilde{\tilde{U}}$ es otra función càd, creciente y nula en cero tal que

$$\tilde{\tilde{U}} = G + \tilde{\tilde{U}} * F,$$

el mismo argumento nos dice que $\tilde{U} = \tilde{\tilde{U}}$. Finalizaremos la prueba de que $\tilde{U}_t = t/\mu$ al mostrar que

$$\frac{t}{\mu} = G_t + \int_0^t \frac{t-s}{\mu} F(ds)\,.$$

En efecto, por el teorema de Tonelli, vemos que

$$\int_0^t \frac{t-s}{\mu} F(ds) = \int_0^t \int_s^t \frac{1}{\mu}\, dr\, F(ds) = \int_0^t F_r \frac{1}{\mu}\, dr = -\frac{t}{\mu} + \int_0^t \frac{1 - F_r}{\mu}\, dr = -\frac{t}{\mu} + G_t.$$

Ahora utilizaremos la igualdad $\tilde{U}_t = t/\mu$ para mostrar que $\tilde{T}$ es un proceso de renovación estacionario al utilizar ecuaciones de renovación. Sea $x > 0$ fijo y consideremos a las funciones

$$V_t = \mathbb{P}(R_t > x) \quad \text{y} \quad \tilde{V}_t = \mathbb{P}\left(\tilde{R}_t > x\right).$$

Al condicionar por el valor de S_0 obtenemos

$$\tilde{V}_t = \mathbb{P}(S_0 > t + x) + \int_0^t V_{t-s}\, \tilde{F}(ds)\,.$$

Si F^x está dada por $F^x_t = F_{x+t}$ (y análogamente definimos a $\tilde{F}^x$) entonces la identidad anterior se escriben de manera más compacta al utilizar convolucióones:

$$\tilde{V} = \tilde{F}^x + V * \tilde{F}.$$

Sin embargo, tenemos una representación distinta para V si nos fijamos en el intervalo $[T_n, T_{n+1})$ que contiene a t:

$$V_t = \mathbb{P}(R_t > x) = \sum_{n=0}^{\infty} \mathbb{P}(T_n \leq t, t + x < T_{n+1})\,;$$

Al condicionar por el valor de S_{n+1} se obtiene

$$V_t = \sum_{n=0}^{\infty} \int_0^t F^x_{t-s}\, \mathbb{P}(T_n \in ds) = \int_0^t F^x_{t-s}\, U(ds)\,,$$

por lo que $V = F^x * U$. Así, vemos que

$$\tilde{V} = \tilde{F}^x + V * F = \tilde{F}^x + F^x * U * \tilde{F}$$

y puesto que $U * \tilde{F} = \tilde{U}$, entonces:

$$\begin{aligned}
\tilde{V}_t &= \mathbb{P}(S_0 \geq t + x) + \int_0^t F(t + x - s)\, \frac{1}{\mu}\, ds \\
&= \mathbb{P}(S_0 \geq t + x) + \int_x^{t+x} F(s)\, \frac{1}{\mu}\, ds \\
&= \mathbb{P}(S_0 \geq x)\,.
\end{aligned}$$

CAPÍTULO 4

El proceso de Poisson

Dentro de los procesos de renovación analizados en el capítulo anterior hay uno muy destacado: el proceso de Poisson. El objetivo de este capítulo es mostrar que también se le puede interpretar como una versión a tiempo continuo de una caminata aleatoria y por lo tanto como un proceso de Makov.

Definición. Un **proceso de Poisson de intensidad** λ es el proceso de contéo asociado a un fenómeno de renovación con tiempos entre sucesos exponenciales del mismo parámetro λ.

La razón por la cual se le llama proceso de Poisson es la siguiente:

Proposición 4.1. *Si N es un proceso de Poisson de parámetro λ entonces N_t tiene distribución Poisson de parámetro λt.*

Demostración. Puesto que los tiempos interarribo (S_n) son exponenciales independientes de parámetro λ, se sigue que su enésima suma parcial $T_n = S_1 + \cdots + S_n$ tiene distribución Γ de parámetro de posición n y de escala λ. Por lo tanto, la distribución conjunta de (T_n, S_{n+1}) admite la siguiente densidad:

$$f_{T_n,S_{n+1}}(t,s) = \mathbf{1}_{t>0}\frac{(\lambda t)^{n-1}\,\lambda e^{-\lambda t}}{(n-1)!}\mathbf{1}_{s>0}\lambda e^{-\lambda s}.$$

Su N es el proceso de contéo asociado, vemos que:

$$\begin{aligned}
\mathbb{P}(N_t = n) &= \mathbb{P}(T_n \leq t < T_n + S_{n+1}) \\
&= \int_0^t \int_{t-x}^{\infty} f_{T_n,S_{n+1}}(x,s)\,ds\,dx \\
&= \int_0^t \frac{(\lambda x)^{n-1}\,\lambda e^{-\lambda x}}{(n-1)!} e^{-\lambda(t-x)}\,dx \\
&= e^{-\lambda t}\frac{(\lambda t)^n}{n!}.
\end{aligned}$$

□

1. El proceso de Poisson como proceso de Lévy

Recordemos que el proceso de Poisson de intensidad λ es estacionario en el siguiente sentido: el proceso N^t dado por $N^t_s = N_{t+s} - N_t, s \geq 0$ es también un proceso de Poisson de intensidad λ. Aún más es cierto: de hecho N^t es independiente de $\mathscr{F}_t^N = \sigma(N_s : s \leq t)$. Antes de ver este resultado pasaremos por uno más sencillo.

DEFINICIÓN. Un proceso estocástico $X = (X_t, t \geq 0)$ tiene **incrementos independientes** si cuando $0 = t_0 < t_0 < \cdots < t_n$ se tiene que $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ son variables aleatorias independientes.

Un proceso estocástico $X = (X_t, t \geq 0)$ tiene **incrementos estacionarios** si $X_{t+s} - X_t$ tiene la misma distribución que X_s.

Un **proceso de Lévy** es un proceso estocástico X con trayectorias càdlàg que comienza en cero y tiene incrementos independientes y estacionarios.

En otras palabras, un proceso de Lévy es la versión a tiempo continuo de una caminata aleatoria.

EJERCICIO 4.1. Pruebe que si X tiene incrementos independientes entonces el proceso X^t dado por $X_s^t = X_{t+s} - X_t$ es independiente de $\mathscr{F}_t^X = \sigma(X_s : s \geq 0)$.

El resultado principal sobre el proceso de Poisson es el siguiente:

TEOREMA 4.1. *Un proceso de contéo es un proceso de Poisson si y sólo si es un proceso de Lévy.*

Probaremos este teorema en dos partes, pues para la segunda debemos analizar la propiedad de Markov del proceso de Poisson.

DEMOSTRACIÓN. (El proceso de Poisson es de Lévy) Claramente el proceso de Poisson comienza en cero y tiene trayectorias càdlàg.

Sean $S_1, S_2, \ldots$ exponenciales independientes de parámetro λ, T las sumas parciales asociadas dadas por $T_0 = 0$, $T_n = S_1 1 + \cdots + S_n$ y $N = (N_t, t \geq 0)$ el proceso de contéo asociado dado por

$$N_t = \sum_{n=0}^{\infty} n \mathbf{1}_{T_n \leq t < T_{n+1}}.$$

Consideremos al proceso N^t dado por $N_s^t = N_{t+s} - N_t$. Este proceso es un proceso de contéo. Sus tiempos entre sucesos son

$$S_1^t = T_{N_t+1} - t, S_{N_t+2}, S_{N_t+3}, \ldots.$$

Al descomponer respecto del valor de N_t vemos que

$$
\begin{aligned}
&\mathbb{P}\big(S^t_1 > s_1, S^t_2 > s_2, \ldots, S^t_n > s_m\big) \\
&= \sum_n \mathbb{P}(N_t = n, S_{n+1} - t > s_1, S_{n+2} > s_2, \ldots, S_{n+m} > s_m) \\
&= \sum_n \mathbb{P}(T_n \leq t < T_{n+1}, S_{n+1} - t > s_1, S_{n+2} > s_2, \ldots, S_{n+m} > s_m) \\
&= \sum_n \mathbb{P}(T_n \leq t < T_n + S_{n+1}, S_{n+1} - t > s_1)\, \mathbb{P}(S_{n+2} > s_2, \ldots, S_{n+m} > s_m)\,.
\end{aligned}
$$

Al condicionar sobre el valor de T_n y utilizando su independencia con S_{n+1} obtenemos:

$$
\mathbb{P}(T_n \leq t < T_n + S_{n+1}, T_m + S_{n+1} - t > s_1) = \mathbb{E}\Big(\mathbf{1}_{T_n \leq t} e^{-\lambda(t - T_m)}\Big)\, e^{-\lambda s_1}.
$$

Al utilizar $s_1 = 0$ vemos que

$$
\mathbb{P}(N_t = n) = \mathbb{E}\Big(\mathbf{1}_{T_n \leq t} e^{-\lambda(t - T_m)}\Big)\,,
$$

por lo cual:

$$
\begin{aligned}
&\mathbb{P}\big(S^t_1 > s_1, S^t_2 > s_2, \ldots, S^t_n > s_m\big) \\
&= \sum_n \mathbb{P}(N_t = n)\, e^{-\lambda s_1} \cdots e^{-\lambda s_m}.
\end{aligned}
$$

Se concluye que $S^t_1, \ldots, S^t_m$ son exponenciales independientes de parámetro λ, por lo que N^t es un proceso de Poisson. (Esto es consecuencia de nuestro estudio de la estacionariedad de procesos de renovación; se decidió rehacer la prueba en este capítulo para no depender de argumentos de renovación). Además, podemos deducir que entonces

$$
\mathbb{P}(N_{t_1} = k_1, \ldots, N_{t_n} = k_n) = \mathbb{P}\big(N^t_{t_1} = k_1, \ldots, N^t_{t_n} = k_n\big)
$$

al expresar los eventos anteriores en términos de tiempos entre sucesos. En particular, vemos que N tiene incrementos estacionarios.

Pasemos a la independencia de los incrementos de N. Para esto, consideremos al conjunto

$$
A = \{N_{t_1} = k_1, \ldots, N_{t_l} = k_l\}
$$

donde $t_1 \leq \cdots \leq t_l \leq t$. Al descomponer sobre el valor de N_t se obtiene:

$$
\begin{aligned}
&\mathbb{P}\big(A, S^t_1 > s_1, S^t_2 > s_2, \ldots, S^t_n > s_m\big) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(A, T_n \leq t < T_{n+1}, T_{n+1} - t > s_1)\, e^{-\lambda s_2} \cdots e^{-\lambda s_m},
\end{aligned}
$$

puesto que el evento $A \cap \{N_t = n\}$ se puede escribir como la intersección de un evento que depende de $S_1, \ldots, S_n$ intersectado con $N_t = n$. Al repetir el argumento anterior (caso $A = \Omega$), vemos que

$$\mathbb{P}\big(A, N_t = n, S_1^t > s_1\big) = \mathbb{P}(A, N_t = n)\, e^{-\lambda s_1}.$$

Se concluye entonces que $S_1^t, S_2^t, \ldots,$ son independientes de A y que por ende N^t es independiente de A. Por inducción se sigue entonces que N tiene incrementos independientes. □

Una primera consecuencia sencilla del hecho de que un proceso de Poisson sea de Lévy es que podemos calcular sus distribuciones finito-dimensionales: si $0 = t_0 < t_1 < \cdots < t_l$ y definimos $k_0 = 0$ entonces

$$\begin{aligned}
&\mathbb{P}(N_{t_1} = k_1, \ldots, N_{t_l} = k_l) \\
&= \mathbb{P}(N_{t_1} - N_{t_0} = k_1, N_{t_2} - N_{t_1} = k_2, \ldots, N_{t_l} - N_{t_l-1} = k_l - k_{l-1}) \\
&= \prod_{i=1}^{l} e^{-\lambda(t_i - t_{i-1})} \frac{\left(\lambda\,(t_i - t_{i-1})^{k_i - k_{i-1}}\right)}{(k_i - k_{i-1})!}.
\end{aligned}$$

2. El proceso de Poisson compuesto

Otra consecuencia de que el proceso de Poisson tenga incrementos independientes y estacionarios es que podemos inmediatamente construir otros ejemplos de procesos de Lévy al hacer que las caminatas aleatorias evolucionen a tiempo continuo. Sean $\xi_1, \xi_2, \ldots$ variables aleatorias independientes e idénticamente distribuidas, independientes también de un proceso de Poisson N (digamos de intensidad λ). Sea $\chi_0 = 0$ y $\chi_n = \xi_1 + \cdots + \chi_n$. Consideremos al proceso estocástico $X = (X_t, t \geq 0)$ dado por

$$X_t = \chi_{N_t}.$$

A X se le conoce como **proceso de Poisson compuesto**; su intensidad es la misma que la del proceso Poisson y su distribución de salto es la distribución de ξ_1.

Proposición 4.2. *El proceso de Poisson compuesto es un proceso de Lévy.*

Demostración. El proceso X comienza en cero y tiene trayectorias que son constantes por pedazos, además de continuas por la derecha y con límites por la izquierda.

Para ver que X tiene incrementos independientes y estacionarios, notemos que si $0 = t_0 \leq t_1 \leq \cdots \leq t_n$ y $A_i \in \mathscr{B}_{\mathbb{R}}$ para $i = 1, \ldots, n$, entonces

$$\begin{aligned}
&\mathbb{P}\big(X_{t_i} - X_{t_{i-1}} \in A_i, i = 1, \ldots, n\big) \\
&= \sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \mathbb{P}\big(\chi_{k_i} - \chi_{k_{i-1}} \in A_i, N_{t_i} = k_i, i = 1, \ldots, n\big) \\
&= \sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \prod_{i=1}^{n} \mathbb{P}\big(\chi_{k_i} - \chi_{k_{i-1}} \in A_i\big) \, \mathbb{P}(N_{t_i} = k_i, i = 1, \ldots, n) \,,
\end{aligned}$$

puesto que las variables $\{\chi_i : i = 0, 1, \cdots\}$ y $\{N_t : t \geq 0\}$ son independientes. Si $k_2 \geq k_1$, entonces $\chi_{k_2} - \chi_{k_1}$ tiene la misma distribución que $\chi_{k_1 - k_2}$, entonces

$$\begin{aligned}
&\sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \prod_{i=1}^{n} \mathbb{P}\big(\chi_{k_i} - \chi_{k_{i-1}} \in A_i\big) \, \mathbb{P}(N_{t_i} = k_i, i = 1, \ldots, n) \\
&= \sum_{0=k_0 \leq k_1 \leq \cdots \leq k_n} \prod_{i=1}^{n} \mathbb{P}\big(\chi_{k_i - k_{i-1}} \in A_i\big) \prod_{i=1}^{n} \mathbb{P}\big(N_{t_i} - N_{t_{i-1}} = k_i - k_{i-1}\big) \\
&= \sum_{0 \leq j_1, \ldots, j_n} \prod_{i=1}^{n} \mathbb{P}\big(\chi_{j_i} \in A_i, N_{t_i - t_{i-1}} = j_i\big) \\
&= \prod_{i=1}^{n} \mathbb{P}\big(X_{t_i - t_{i-1}} \in A_i\big) \,.
\end{aligned}$$

Si utilizamos el caso $n = 2$, con $A_1 = \mathbb{R}$, podemos concluir que $X_{t_2} - X_{t_1}$ y $X_{t_2 - t_1}$ tienen la misma distribución, de donde la igualdad

$$\mathbb{P}\big(X_{t_i} - X_{t_{i-1}} \in A_i, i = 1, \ldots, n\big) = \prod_{i=1}^{n} \mathbb{P}\big(X_{t_i - t_{i-1}} \in A_i\big)$$

nos permite concluir la independencia de los incrementos y por lo tanto, que el proceso Poisson compuesto es un proceso de Lévy. □

Ahora analizaremos algunas martingalas asociadas a cualquier proceso de Lévy. De hecho, son simples generalizaciones de las martingalas que hemos estudiado para caminatas aleatorias y tienen expresiones más explícitas para procesos de Poisson y de Poisson compuesto. Sea X un proceso de Lévy arbitrario. Si suponemos que X_t es integrable para alguna $t > 0$ entonces lo es para toda $t > 0$ por el siguiente argumento. Primero notamos que si $\mathbb{E}(|X_t|) < \infty$ entonces $\mathbb{E}(|X_{nt}|) \leq n\mathbb{E}(|X_t|) < \infty$. Segundo, al utilizar la descomposición X_t como suma de n variables independientes con la misma distribución que $X_{t/n}$ y al condicionar por $n - 1$ sumandos, vemos que $\mathbb{E}\big(\big|X_{t/n}\big|\big) < \infty$. Esto nos dice que $\mathbb{E}(|X_{qt}|) \leq q\mathbb{E}(|X_t|)$ para cualquier q racional. Finalmente, por continuidad por la derecha y el lema de Fatou, vemos que $\mathbb{E}(|X_{st}|) \leq s\mathbb{E}(|X_t|)$ para cualquier $s \geq 0$. Ahora notemos que por estacionariedad de los incrementos se sigue que si $\phi(s) = \mathbb{E}(X_s)$ entonces $\phi(t+s) = \phi(t) + \phi(s)$. Así, ϕ satisface la llamada ecuación funcional de Cauchy (cf. [BGT87]). Por otra parte, puesto que X tiene trayectorias continuas por la derecha y $X_0 = 0$, se sigue que

$$X_t(\omega) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbf{1}_{(k/n,(k+1)/n]}(t)\, X_{(k+1)/n}(\omega)\,.$$

Lo anterior muestra que X, considerado como una función de (ω, t), es medible de $\Omega \times [0,\infty)$ a $\mathbb{R}$ con la σ-álgebra producto $\mathscr{F} \otimes \mathscr{B}_{\mathbb{R}}$. Por el teorema de Fubini-Tonelli se sigue que $t \mapsto \mathbb{E}(X_t) = \phi(t)$ es medible. Se sabe que cualquier solución medible a la ecuación funcional de Cauchy ϕ satisface $\phi(t) = t\phi(1)$ (cf. [BGT87, Teorema 1.1.8, p. 5]). Por lo tanto:

$$\mathbb{E}(X_t) = t\mathbb{E}(X_1)\,.$$

Al utilizar que la varianza de una suma de variables aleatorias independiente es la suma de las varianzas, se verifica con argumentos análogos que si $\mathbb{E}\big(X_t^2\big) < \infty$ para alguna $t > 0$ entonces $\mathbb{E}\big(X_t^2\big) < \infty$ para toda $t \geq 0$ y

$$\mathrm{Var}(X_t) = t\mathrm{Var}(X_1)\,.$$

Al utilizar logaritmos se puede verificar que si $\mathbb{E}\big(e^{\lambda X_t}\big) < \infty$ para alguna $\lambda > 0$ y para alguna $t > 0$ entonces $\mathbb{E}\big(e^{\lambda X_t}\big) < \infty$ para toda $t > 0$ y $\mathbb{E}\big(e^{\lambda X_t}\big) = \mathbb{E}\big(e^{\lambda X_1}\big)^t$.

Sean $\mu = \mathbb{E}(X_1)$ y $\sigma^2 = \mathrm{Var}(X_1)$ y consideremos a

$$M_t = X_t - t\mu \quad \text{y} \quad N_t = M_t^2 - t\sigma^2.$$

Se afirma que tanto M como N son martingalas respecto de la filtración canónica generada por X. En efecto, simplemente utilizamos la independencia entre $X_{t+s} - X_s$ y $\mathscr{F}_s^X$ si $s \leq t$ para obtener:

$$\begin{aligned}
\mathbb{E}(\, M_t \mid \mathscr{F}_s) &= \mathbb{E}(\, X_{t+s} - X_s \mid \mathscr{F}_s) + X_s - t\mu \\
&= \mathbb{E}(X_{t+s} - X_s) + X_s - t\mu \\
&= M_s.
\end{aligned}$$

El análisis para N es similar y utiliza la ortogonalidad de los incrementos de la martingala M. Si $a = \mathbb{E}\big(e^{\lambda X_1}\big) < \infty$, podemos considerar a

$$M_t^\lambda = \frac{e^{\lambda X_t}}{a^t}$$

y notar que es martingala pues:

$$\mathbb{E}\big(\, M_t^\lambda \mid \mathscr{F}_s\big) = \mathbb{E}\left(\, M_s^\lambda \frac{e^{\lambda(X_t - X_s)}}{a^{t-s}} \,\middle|\, \mathscr{F}_s\right) = M_s^\lambda \mathbb{E}\left(\frac{e^{\lambda(X_{t-s})}}{a^{t-s}}\right) = M_s^\lambda.$$

Ejercicio 4.2. Calcular la esperanza y varianza del proceso de Poisson y de Poisson compuesto (en términos de la intensidad y la distribución de salto). Probar que si X es

$$\mathbb{E}\left(e^{iuZ_t}\right) = e^{-\lambda t(1-\psi(u))} \quad \text{donde} \quad \psi(u) = \mathbb{E}\left(e^{iu\xi_1}\right).$$

3. La propiedad de Markov fuerte para procesos de Lévy

Sea X un proceso de Lévy. Como en el caso de caminatas aleatorias, se tiene la siguiente propiedad de Markov.

Proposición 4.3. *Para cualquier $t > 0$, el proceso X^t dado por $X^t_s = X_{t+s} - X_t$ es independiente de $\mathscr{F}^X_t$ y tiene las mismas distribuciones finito dimensionales que X.*

Demostración. En el Ejercicio 4.1 se pide probar que la independencia de los incrementos de X implican que X^t es independiente de $\mathscr{F}^X_t$.

Para probar que X^t tiene las mismas distribuciones finito dimensionales que X, notamos que

$$\begin{aligned}
&\left(X^t_{t_1}, X^t_{t_2} - X^t_{t_1}, \ldots, X^t_{t_n} - X^t_{t_{n-1}}\right) \\
&= \left(X_{t+t_1} - X_t, X_{t+t_2} - X_{t+t_1}, \ldots, X_{t+t_n} - X_{t+t_{n-1}}\right) \\
&\stackrel{d}{=} \left(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}\right).
\end{aligned}$$

Al considerar $f(x_1, \ldots, x_n) = (x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n)$, vemos que entonces

$$\begin{aligned}
\left(X^t_{t_1}, X^t_{t_2}, \ldots, X^t_{t_n}\right) &= f\left(X^t_{t_1}, X^t_{t_2} - X^t_{t_1}, \ldots, X^t_{t_n} - X^t_{t_{n-1}}\right) \\
&\stackrel{d}{=} f\left(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}\right) \\
&= \left(X_{t_1}, X_{t_2}, \ldots, X_{t_n}\right).
\end{aligned}$$

□

Ahora enunciaremos una extensión mucho más útil: la propiedad de Markov fuerte. Sea T un tiempo de paro respecto de la filtración $\left(\mathscr{F}^X_t\right)$. Es decir, T es una variable aleatoria con valores $[0, \infty]$ y tal que $\{T \leq t\} \in \mathscr{F}^X_t$. Definimos a la σ-álgebra detenida en T, denotada $\mathscr{F}^X_T$, como sigue:

$$\mathscr{F}^X_T = \sigma\left(A \in \mathscr{F} : A \cap \{T \leq t\} \in \mathscr{F}^X_t \text{ para todo } t \geq 0\right).$$

(Ya se dejó al lector como ejercicio comprobar que $\mathscr{F}^X_T$ es una σ-álgebra en el caso de tiempo discreto; la prueba es muy similar).

Teorema 4.2 (Propiedad de Markov fuerte para procesos de Lévy). *Si X es un proceso de Lévy y T es un tiempo de paro entonces, condicionalmente a $T < \infty$, el proceso X^T dado por $X^T_s = X_{T+s} - X_T$ es un proceso de Lévy con las mismas distribuciones finito dimensionales que X e independiente de $\mathscr{F}^X_T$.*

Demostración. Para cada $n \geq 1$, definimos T^n igual a $(k+1)/2^n$ si $T \in [k/2^n, (k+1)/2^n)$. Formalmente, podemos escribir

$$T^n = \lceil 2^n T \rceil / 2^n.$$

Entonces T^n es una sucesión de variables aleatorias que decrecen hacia T. Además, notemos que

$$\begin{aligned}\left\{T^n = \frac{k+1}{2^n}\right\} &= \left\{\frac{k}{2^n} \leq T < \frac{k+1}{2^n}\right\} \\ &= \left\{T < \frac{k+1}{2^n}\right\} \setminus \left\{T < \frac{k}{2^n}\right\} \in \mathscr{F}_{(k+1)/2^n}.\end{aligned}$$

Por el mismo argumento, si $A \in \mathscr{F}_T$ entonces

$$A \cap \{T^n = k/2^n\} \in \mathscr{F}_{(k+1)/2^n}.$$

Así, por la propiedad de Markov, vemos que

$$\begin{aligned}&\mathbb{P}\left(A, T^n = \frac{k+1}{2^n}, X^{T^n}_{t_1} \in A_1, \ldots, X^{T^n}_{t_m} \in A_m\right) \\ &= \mathbb{P}\left(A, T^n = \frac{k+1}{2^n}\right) \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_m} \in A_m).\end{aligned}$$

Al sumar sobre k, vemos que

$$\mathbb{P}\Big(A, T < \infty, X^{T^n}_{t_1} \in A_1, \ldots, X^{T^n}_{t_m} \in A_m\Big) = \mathbb{P}(A, T < \infty)\, \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_m} \in A_m).$$

Como X es continuo por la derecha y T^n decrece a T, vemos que confome $n \to \infty$: $\mathbb{P}\big(A, T < \infty, X^T_{t_1} = k_1, \ldots, X^T_{t_m} = k_m\big) = \mathbb{P}(A, T < \infty)\mathbb{P}(X_{t_1} = k_1, \ldots, X_{t_m} = k_m)$. Se concluye que X^T es un proceso de contéo y de Lévy con las mismas distribuciones finito-dimensionales que X. □

Ya armados con la propiedad de Markov fuerte, podemos concluir la demostración del Teorema 4.1.

Conclusión de la prueba del Teorema 4.1. Sea X un proceso de contéo con incrementos independientes y estacionarios. Sea $T_n = \inf\{t \geq 0 : X_t = n\}$. Entonces T_n es un tiempo de paro para X. Puesto que X es un proceso de contéo, se sigue

$$\begin{aligned}\mathbb{P}(T_1 > t+s) &= \mathbb{P}(X_{t+s} = 0) \\ &= \mathbb{P}(X_t = 0, X_{t+s} - X_t = 0) \\ &= \mathbb{P}(X_t = 0)\, \mathbb{P}(X_s = 0) \\ &= \mathbb{P}(T_1 > t)\, \mathbb{P}(T_1 > s).\end{aligned}$$

Por lo tanto, T_1 tiene la propiedad de pérdida de memoria y por ende distribución exponencial de algún parámetro $\lambda \geq 0$. Si $\lambda > 0$, se sigue que $T_1 = \infty$ casi seguramente, por lo que $X_t = 0$ para toda $t \geq 0$ con probabilidad 1. Si $\lambda > 0$ se sigue que $T_1 < \infty$ casi seguramente. Notemos que

$$T_{n+1} - T_n = \inf \left\{ t \geq 0 : X^{T_n}_t = 1 \right\},$$

y que

$$\mathbb{P}(T_{n+1} - T_n > t) = \mathbb{P}\Big(X^{T_n}_t = 0 \Big),$$

por lo que, por inducción al utilizar la propiedad de Markov fuerte, vemos que $T_{n+1} - T_n$ tiene la misma distribución que T_1 y también es exponencial del mismo parámetro λ. Se concluye que X es un proceso de Poisson de intensidad λ. □

4. Medidas aleatorias de Poisson

Daremos ahora una generalización del proceso de Poisson mediante la cual se pueden construir procesos de Lévy más generales. Para motivarla, consideremos un proceso de Poisson N con tiempos de ocurrencia $T_1 < T_2 < \cdots$. Con estos tiempos de ocurrencia, podemos considerar a la siguiente medida:

$$\Xi = \sum_{n=1}^{\infty} \delta_{T_i}.$$

En otras palabras, para cada boreliano A de $[0, \infty)$

$$\Xi(A) = \# \{ i \geq 1 : T_i \in A \}.$$

Notemos que para cada $\omega \in \Omega$ fijo, Ξ es una medida, lo cual nos tienta a decir que Ξ es una medida aleatoria. Esta terminología de hecho es utilizada en un contexto más específico en la probabilidad, por lo que daremos la definición más adelante. Observemos que

$$N_t = \Xi([0, t])$$

y que por lo tanto $\Xi([0, t])$ es una variable aleatoria. Este hecho admite la siguiente generalización.

Proposición 4.4. *Para cada boreliano A de $[0, \infty)$, $\Xi(A)$ es una variable aleatoria.*

Demostración. Nos basamos en el lema de Dynkin al considerar al π-sistema

$$\mathscr{C} = \{ [0, t] : t \geq 0 \}$$

y al λ-sistema

$$\mathscr{M} = \left\{ A \in \mathscr{B}_{[0,\infty)} : \Xi(A) \text{ es variable aleatoria} \right\}.$$

Puesto que si $0 \leq a < b$ entonces

$$(a, b) = \bigcup_{n=1}^{\infty} [0, b - 1/n] \setminus [0, a],$$

se sigue que $(a,b) \in \sigma(\mathscr{C})$ y por lo tanto $\mathscr{B}_{[0,\infty)} = \mathscr{C}$. Por otra parte, hemos verificado que $\mathscr{C} \subset \mathscr{M}$ y por lo tanto $\mathscr{B} = \mathscr{B}_{[0,\infty)}$. □

Definición. Sea $(S, \mathscr{S})$ un espacio medible. Una medida aleatoria es una aplicación $\Xi : \Omega \times \mathscr{S} \to [0,\infty]$ tal que

(1) para toda $\omega \in \Omega$ la aplicación $A \mapsto \Xi(\omega, A)$ es una medida en $\mathscr{S}$ y
(2) para toda $A \in \mathscr{S}$ la aplicación $\omega \mapsto \Xi(\omega, A)$ es una variable aleatoria.

Así, vemos que la medida Ξ construida poniendo una masa puntual en cada instante de ocurrencia de un proceso de Poisson es en efecto una medida aleatoria. Para esta medida, el lema de Dynkin implica el siguiente resultado.

Proposición 4.5. *Para todo boreliano A la variable $\Xi(A)$ tiene distribución Poisson de parámetro $\lambda Leb(A)$. Además, si $A_1, \ldots, A_n$ son borelianos ajenos por pares entonces $\Xi(A_1), \ldots, \xi(A_n)$ son independientes.*

Antes de pasar a la demostración, veamos la siguiente definición.

Definición. Sea $(S, \mathscr{S})$ un espacio medible y ν una medida. Una medida aleatoria Ξ es una **medida de Poisson aleatoria** de medida de intensidad ν si

(1) para todo $A \in \mathscr{S}$ la variable $\Xi(A)$ tiene distribución Poisson de parámetro $\nu(A)$ y
(2) si $A_1, \ldots, A_n \in \mathscr{S}$ son ajenos por pares entonces $\Xi(A_1), \ldots, \Xi(A_n)$ son independientes.

En otras palabras, la Proposición 4.5 nos dice que la medida asociada al proceso de Poisson de intensidad λ es una medida aleatoria de Poisson en $[0,\infty)$ cuya intensidad es λ veces la medida de Lebesgue.

Prueba de la proposición 4.5. Consideremos a la clase $\mathscr{C}$ que consta de las uniones ajenas de intervalos de la forma $(a,b] \subset [0,\infty)$. Esta clase es un álgebra y como su σ-álgebra generada contiene a los intervalos abiertos, entonces genera a $\mathscr{B}_{[0,\infty)}$. Si $A \in \mathscr{C}$, digamos que

$$A = \bigcup_{i=1}^{n} (a_i, b_i]$$

donde $a_i < b_i < a_{i+1}$ entonces puesto que el proceso de Poisson tiene incrementos independientes y que la suma de variables Poisson independientes tiene tambén distribución Poisson cuyo parámetro es la suma de los parámetros de los sumandos, vemos que

$$\Xi(A) = \sum_{i=1}^{n} N_{b_i} - N_{a_i} \sim \text{Poisson}\left(\lambda \sum_{i=1}^{n} (b_i - a_i)\right) = \text{Poisson}(\lambda \text{Leb}(A)).$$

Por otra parte, sea $\mathscr{M}$ la clase monótona siguiente:

$$\mathscr{M} = \{A \in \mathscr{B}[0,\infty) : \Xi(A) \text{ tiene distribución Poisson de parámetro } \lambda\text{Leb}(A)\}.$$

En efecto, si $A_1, A_2, \ldots \in \mathscr{M}$, $A_i \subset A_{i+1}$ y definimos a $A = \cup_i A_i$, entonces $\Xi(A_i) \uparrow \Xi(A)$ y como una sucesión convergence de enteros es eventualmente constante, vemos que

$$\begin{aligned}\mathbb{P}(\Xi(A) = n) &= \lim_{i\to\infty} \mathbb{P}(\Xi(A_i) = n)\\ &= \lim_{i\to\infty} e^{-\lambda \mathrm{Leb}(A_i)} \frac{(\lambda \mathrm{Leb}(A_i))^n}{n!}\\ &= e^{-\lambda \mathrm{Leb}(A)} \frac{(\lambda \mathrm{Leb}(A))^n}{n!},\end{aligned}$$

por lo que $A \in \mathscr{M}$. Puesto que $\mathscr{C} \subset \mathscr{M}$, se tiene que $\mathscr{M} = \mathscr{B}_{[0,\infty)}$.

Para la misma álgebra $\tilde{C}$ consideremos a la clase $\mathscr{M}_1$ de eventos $A_1 \in \mathscr{B}_{[0,\infty)}$ tales que si $A_1, \ldots, A_n \in \mathscr{C}$ y $A_1, A_2, \ldots, A_n$ son ajenos por pares entonces $\Xi(A_1)$ es independiente de $\Xi(A_2), \ldots, \Xi(A_n)$. Puesto que la independencia es estable cuando tomamos límites casi seguros, se sigue que $\mathscr{M}_1$ es una clase monótona. Por otra parte, al utilizar la independencia de los incrementos de N, se sigue que $\tilde{C} \subset \mathscr{M}_1$, lo cual implica que $\mathscr{M}_1 = \mathscr{B}_{[0,\infty)}$. Para proceder, consideramos a la clase $\mathscr{M}_2$ de eventos $A_2 \in \mathscr{B}_{[0,\infty)}$ tales que si $A_1 \in \mathscr{B}_{\mathbb{R}_+}$, $A_3, \ldots, A_n \in \mathscr{C}$ y si $A_1, A_2, \ldots, A_n$ son ajenos por pares entonces $\Xi(A_2)$ es independiente de $\Xi(A_1), \Xi(A_3), \ldots, \Xi(A_n)$. por independencia de los incrementos sabemos que si $A_2 \in \mathscr{C}$ entonces $\Xi(A_2)$ es independiente de $\Xi(A_3), \ldots, \Xi(A_n)$. Por otra parte, puesto que $\mathscr{M}_1 = \mathscr{B}_{\mathbb{R}_+}$ se sigue que A_1 es independiente de $A_2, \ldots, A_n$ y por lo tanto A_2 es independiente de $A_1, A_3, \ldots, A_n$. Esto prueba que $\mathscr{C} \subset \mathscr{M}_2$ y por lo tanto $\mathscr{M}_2 = \mathscr{B}_{\mathbb{R}_+}$. El resto de la demostración es análogo. □

Ahora mostraremos que en general, existen las medidas de Poisson aleatorias.

Teorema 4.3. *Sea ν una medida σ-finita en el espacio medible $(S, \mathscr{S})$. Entonces existe (un espacio de probabilidad en el que está definida) la medida de Poisson aleatoria con medida de intensidad ν.*

Demostración. Se hará en dos partes: la primera cuando ν es una medida finita y la segunda cuando ν es infinita.

Supongamos que ν es una medida finita. Sea $\nu^n(A) = \nu(A)/\nu(S)$ y consideremos a una sucesión de variables aleatorias independientes $U_1, U_2, \ldots$ con distribución ν^n independientes también de una variable aleatoria N con distribución Poisson de parámetro $\nu(S)$. Definamos a la medida

$$\Xi = \begin{cases} \sum_{1\leq i\leq N} \delta_{U_i} & N > 0\\ 0 & N = 0\end{cases}.$$

Sea $A \in \mathscr{S}$. Puesto que $\Xi(A) = m$ si y sólo si existe $k \geq m$ tal que $N = k$ y existe $I \subset \{1, \ldots, k\}$ de cardinalidad m tal que si $i \leq k$ entonces $U_i \in A$ si y sólo si $i \in I$, vemos que $\Xi(A)$ es una variable aleatoria; en otras palabras, $\Xi(A)$ es una variable aleatoria gracias a la igualdad de conjuntos

$$\{\Xi(A) = m\} = \bigcup_{n \geq m} \{N = n\} \cap \bigcup_{\substack{I \subset \{1,\ldots,n\} \\ \#I = m}} \cup_{i \in I} \{U_i \in A\} \cap \cup_{i \in I^c} \{U_i \in A^c\}.$$

Esto muestra que Ξ es una medida aleatoria. Para ver que es de Poisson consideremos una partición $A_1, \ldots, A_n$ de S consistente de elementos de $\mathscr{S}$ y $k_1, \ldots, k_n \geq 0$; denotemos por k a $k_1 + \cdots + k_n$. Entonces el evento $\Xi(A_1) = k_1, \ldots, \Xi(A_n) = k_n$ si y sólo si $N = k$ y existe una partición $I_1, \ldots, I_n$ de $\{1, \ldots, k\}$ tal que I_j tiene cardinalidad k_j y si $i \leq k$ entonces $U_i \in A_j$ si y sólo si $i \in I_j$. Entonces

$$\begin{aligned} &\mathbb{P}(\Xi(A_1) = k_1, \ldots, \Xi(A_n) = k_n) \\ &= \sum_{I_1,\ldots,I_n} \mathbb{P}(N = n) \prod_{j=1}^{n} \mathbb{P}(U_1 \in A_j)^{k_j} \\ &= \sum_{I_1,\ldots,I_n} e^{-\nu(S)} \frac{\nu(S)^k}{k!} \prod_{j=1}^{n} \mathbb{P}(U_1 \in A_j)^{k_j}. \end{aligned}$$

Notemos que los sumandos no dependen de la partición y puesto que hay

$$\frac{k!}{k_1! \cdots k_n!}$$

tales particiones obtenemos

$$\begin{aligned} &\mathbb{P}(\Xi(A_1) = k_1, \ldots, \Xi(A_n) = k_n) \\ &= \frac{k!}{k_1! \cdots k_n!} e^{-\nu(A_1)} \cdots e^{-\nu(A_n)} \frac{\nu(S)^k}{k!} \prod_{j=1}^{n} \left[\frac{\nu(A_j)}{\nu(S)}\right]^{k_j} \\ &= \prod_{j=1}^{n} e^{-\nu(A_j)} \frac{\nu(A_k)_j^k}{k_j!}, \end{aligned}$$

lo cual muestra que $\Xi(A_1), \ldots, \Xi(A_n)$ son independientes y con distribución Poisson de parámetros respectivos $\nu(A_1), \ldots, \nu(A_n)$.

Cuando ν es una medida infinita, consideremos una partición $A_1, A_2, \ldots$ de S que consta de elementos de $\mathscr{S}$ y tal que $\nu(A_i) < \infty$. Sea $\nu_i(A) = \nu(A \cap A_i)$. Sean $\Xi_1, \Xi_2, \ldots$ medidas aleatoris de Poisson independientes de intensidades $\nu_1, \nu_2, \ldots$, y definamos a

$$\Xi = \sum_{i=1}^{\infty} \Xi_i.$$

Claramente Ξ es una medida aleatoria; para ver que es de Poisson, consideremos a $B_1, \ldots, B_n \in \mathscr{S}$ ajenos por pares. Puesto que

$$\Xi_1(B_i \cap A_1), \Xi_2(B_i \cap A_2), \ldots$$

son independientes y de distribución Poisson de parámetros respectivos $\nu(B_i \cap A_1)$, $\nu(B_i \cap A_2)$, ..., se sigue que su suma es Poisson de parámetro $\nu(B_j)$. Por otra parte, puesto que Ξ_1, Ξ_2 son medidas de Poisson aleatorias independientes se sigue que

$$\{\Xi_i(B_1) : i \geq 1\}, \ldots, \{\Xi_i(B_n) : i \geq 1\}$$

son colecciones de variables aleatorias independientes. Por lo tanto

$$\Xi(B_1), \ldots, \Xi(B_n)$$

son independientes. □

Cabe preguntarse si cualquier medida de Poisson aleatoria Ξ en $\mathbb{R}_+$ cuya intensidad sea $\lambda \times \mathrm{Leb}$ corresponde a un proceso de Poisson. Esto es, podría ser que $\Xi(\{x\}) = 2$ para algún $x \in \mathbb{R}$. En efecto, puesto que $\mathrm{Leb}(\{x\}) = 0$ vemos que, para cada $x \geq 0$ fijo, $\Xi(\{x\}) = 0$ casi seguramente. Sin embargo, puesto que $\Xi(\mathbb{R}) = \infty$ casi seguramente, eso quiere decir que existe x (aleatorio) tal que $\Xi(\{x\}) \geq 1$. Así, no es absolutamente trivial que casi seguramente no exista algún x aleatorio tal que $\xi(\{x\}) \geq 2$. Atacaremos el problema primero si $x \in [0, 1)$. Primero, para $n \geq 1$ fijo, calculemos la distribución condicional de

$$\Xi([0, 1/2^n)), \ldots, \Xi([(2^n - 1)/n, 1)) \mid \Xi([0, 1)) = k.$$

Si $0 \leq k_i$ para $i = 1, \ldots, 2^n$ y $\sum k_i = k$

$$\begin{aligned}
&\mathbb{P}(\Xi([(i-1)/2^n, i/2^n)) = k_i \text{ para } 1 \leq i \leq 2^n \mid \Xi([0, 1)) = k) \\
&= \frac{\prod_{i=1}^{2^n} e^{-\lambda/2^n} \left(\lambda/2^n\right)^{k_i} / k_i!}{e^{-\lambda} \left(\lambda\right)^k / k!} \\
&= \frac{k!}{\prod_{i=1}^{2^n} k_i! n^{k_i}}.
\end{aligned}$$

Se deduce que la distribución condicional es multinomial. Otra manera de generar esta distribución es tirar k uniformes independientes y ver la cantidad de uniformes que caen en cada uno de los intervalos $[0, 1/2^n), \ldots, [(2^n - 1)/2^n, 1)$; caen dos uniformes en el mismo intervalo con probabilidad $1/2^{2n}$. Por lo tanto, la probabilidad de que caigan 2 de las k uniformes en el mismo intervalo está acotada superiormente por $k(k-1)/2^{2n}$. Se sigue que

$$\begin{aligned}
&\mathbb{P}(\text{Existe } x \in [0, 1) \text{ tal que } \Xi(\{x\}) \geq 2) \\
&= \lim_{n \to \infty} \mathbb{P}(\text{Existe } 1 \leq i \leq 2^n \text{ tal que } \Xi([(i-1)/2^n, i/2^n)) \geq 2) \\
&\leq \lim_{n \to \infty} \frac{k^2}{2^{2n}} = 0.
\end{aligned}$$

Así, Ξ es una suma de masas puntuales; digamos que

$$\Xi = \sum_{n=1}^{\infty} \delta_{T_n},$$

donde $T_1 < T_2 < \cdots$. Ahora demostraremos que si $T_0 = 0$ entonces $T_n - T_{n-1}, n \geq 1$ son variables independientes y exponenciales de parámetro λ. En efecto, si definimos a

$$T_i^n = \lfloor nT_i \rfloor / n = \min \{ j \geq 0 : \Xi([0, j/n)) \geq i \},$$

notamos que $n\left(T_i^n - T_i^n\right)$ es una sucesión de geométricas independientes de parámetro λ/n. En efecto, basta notar que

$$\mathbf{1}_{\Xi([(j-1)/n, j/n)) \geq 1}, j \geq 1$$

es una sucesión de variables independientes de distribución Bernoulli de parámetro λ/n. Por lo tanto, $(T_i - T_{i-1}, i \geq 1)$ es una sucesión de exponenciales independientes.

Ejercicio 4.3. Sea N un proceso de Lévy tal que N_t tiene distribución de parámetro λt.

(1) Pruebe que casi seguramente las trayectorias de N son no-decrecientes.
(2) Sea Ξ la única medida en $\mathscr{B}_{\mathbb{R}_+}$ tal que $\Xi([0, t]) = N_t$. Pruebe que Ξ es una medida de Poisson aleatoria de intensidad $\lambda \times \text{Leb}$.
(3) Concluya que N es un proceso de Poisson de intensidad λ.

A continuación presentaremos una especie de transformada de Laplace de las medidas aleatorias de Poisson y una forma de calcularla conocida como fórmula exponencial.

Proposición 4.6. *Sea ν una medida σ-finita en $(S, \mathscr{S})$ una medida de Poisson aleatoria con medida de intensidad ν.*

(1) *(Fórmula exponencial) Si $f : S \to \mathbb{R}_+$ es medible entonces la integral de f respecto de Ξ, denotada por Ξf, es una variable aleatoria y*
$$\mathbb{E}\left(e^{-\Xi f}\right) = e^{-\int \left(1 - e^{-f}\right) d\nu}.$$
(2) *La variable aleatoria Ξf es casi seguramente finita o casi seguramente infinita de acuerdo a si la integral $\int 1 \wedge f \, d\nu$ es finita o no.*

Demostración. La fórmula exponencial se prueba mediante el método estándar notando que si $f = \sum_i c_i \mathbf{1}_{A_i}$ es una función simple con los A_i ajenos por pares entonces

$$\mathbb{E}\left(e^{-\Xi f}\right) = \prod_i e^{-\left(1 - e^{-c_i}\right)\nu(A_i)} = e^{-\int \left(1 - e^{-}f\right) d\nu},$$

gracias al cálculo explícito de la transformada de Laplace de una variable Poisson.

La finitud de la variable Ξf se obtiene de notar que por el teorema de convergencia monótona se sigue que

$$\mathbb{P}(\Xi f = \infty) = \lim_{q \to 0+} \mathbb{E}\left(e^{-q \Xi f}\right) = \lim_{q \to 0+} e^{\int \left(1 - e^{-qf}\right) d\nu}.$$

Si $\int 1 \wedge f \, d\nu < \infty$ entonces, podemos utilizar la desigualdad

$$1 - e^{-qf} \leq 1 \wedge qf \leq 1 \wedge f$$

para $q \leq 1$ y por el teorema de convergencia dominada concluir que

$$\lim_{q \to 0+} \int \left(1 - e^{-qf}\right) d\nu = 0.$$

Por otra parte, la integral $\int 1 \wedge f \, d\nu$ puede ser infinita por dos razones: si

$$\int f \mathbf{1}_{f>1} \, d\nu = \infty$$

entonces $\Xi \{f > 1\} = \infty$ casi seguramente por lo que $\Xi f = \infty$ casi seguramente. Si por otra parte $\int f \mathbf{1}_{f \leq 1} \, d\nu = \infty$ entonces utilizamos la desigualdad

$$1 - e^{-qf} \geq \frac{q}{2} f$$

válida sobre $f \leq 1$ para $q \leq 1$ y concluimos que

$$\int \left(1 - e^{-qf}\right) d\nu = \infty$$

para cualquier $q \leq 1$, por lo cual

$$\mathbb{P}(\Xi f < \infty) = 0. \qquad \square$$

5. Construcción de procesos de Lévy mediante medidas de Poisson

La construcción del proceso de Poisson compuesto puede reinterpretarse de la siguiente manera en términos de medidas aleatorias de Poisson. Consideremos un parámetro de intensidad $\lambda > 0$ y una distribución de salto μ en $\mathbb{R}$. Sea N un proceso de Poisson de parámetro $\lambda > 0$ independientemente de la sucesión iid $X_1, X_2, \ldots$ donde X_i tiene distribución μ. Sabemos que el proceso de Poisson compuesto se construye como

$$X_t = \sum_{i \leq N_t} X_i.$$

Una interpretación en términos de medidas aleatorias es la siguiente: sean $T_1 < T_2 < \cdots$ los instantes de salto del proceso N y consideremos a la medida

$$\Xi = \sum_n \delta_{(T_n, X_n)}.$$

Puesto que $\sum_n \delta_{T_n}$ es una medida de Poisson aleatoria de intensidad $\lambda\mathrm{Leb}$ es fácil ver que Ξ es una medida de Poisson aleatoria de intensidad $\lambda\mathrm{Leb} \times \mu$. Además, si $f_t : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ está dada por $f_t(s, x) = \mathbf{1}_{s \leq t} x$ entonces

$$X_t = \Xi f_t.$$

Elaboraremos esta idea para construir a una clase importante de procesos de Levy.

Definición. Un **subordinador** es un proceso de Lévy X tal que casi seguramente las trayectorias $t \mapsto X_t$ son no-decrecientes.

Ejercicio 4.4. Pruebe que un proceso de Lévy X es un subordinador si y sólo si la distribución de X_t tiene soporte en $[0, \infty)$ para alguna $t > 0$. *Sugerencia:* comience el ejercicio en el caso en que el soporte esté contenido en $[0, \infty)$ para toda $t > 0$.

Ahora daremos una construcción de un subordinador mediante una medida aleatoria de Poisson. Sean μ una medida en $(0, \infty)$ tal que

$$\int_0^\infty 1 \wedge x \, \mu(dx) < \infty.$$

y $d \geq 0$. Sea Ξ una medida de Poisson aleatoria en $\mathbb{R}_+ \times \mathbb{R}_+$ con intensidad $\mathrm{Leb} \times \mu$. Para las funciones f_t que definimos anteriormente, definamos a

$$X_t = dt + \Xi f_t = dt + \int_0^t \int_0^\infty x \, \Xi(ds, dx) \, .$$

Proposición 4.7. *El proceso X es un subordinador. Si μ es una medida infinita entonces el conjunto de discontinuidades de X es denso y numerable. Además, las distribuciones unidimensionales de X están caracterizadas por*

$$\mathbb{E}\big(e^{-\lambda X_t}\big) = e^{-t\Phi(\lambda)} \quad \textit{donde} \quad \Phi(\lambda) = d + \int_0^\infty \left(1 - e^{-\lambda x}\right) \mu(dx) \, .$$

Demostración. Primero notemos que puesto que $s \leq t$ implica $f_s \leq f_t$ entonces X tiene trayectorias no-decrecientes. Además, nuestra hipótesis sobre μ implica que $X_n < \infty$ para toda $n \in \mathbb{N}$ casi seguramente y por lo tanto $X_t < \infty$ para toda $t \geq 0$ casi seguramente. Claramente $X_0 = 0$. Si $t \downarrow s$ entonces $f_t \downarrow f_s$ y por lo tanto X tiene trayectorias continuas por la derecha. Puesto que X es no decreciente, tiene límites por la izquierda. Si (t_n, x_n) es una enumeración de los átomos de Ξ se sigue que X tiene una discontinuidad en t_n y que $X_{t_n} - X_{t_n-} = x_n$. Si μ es una medida infinita entonces $\Xi([0,t] \times \mathbb{R}_+) = \infty$ casi seguramente por lo que X tiene una cantidad numerable de discontinuidades en cada intervalo compacto. Esto implica que las discontinuidades de X son densas.

Sean $0 = t_0 < t_1 < \cdots < t_n$ y $\lambda_1, \ldots, \lambda_n \geq 0$. Al aplicar la fórmula exponencial a la función

$$f(s,x) = \sum_i \mathbf{1}_{t_i \leq s \leq t} \lambda_i x$$

se tiene que

$$\Xi f = \sum_{i=1}^n \lambda_i \left(X_{t_i} - X_{t_{i-1}}\right)$$

y por lo tanto:

$$\mathbb{E}\left(e^{-\sum_{i=1}^n \lambda_i \left(X_{t_i} - X_{t_{i-1}}\right)}\right) = e^{-\sum_{i=1}^n (t_i - t_{i-1}) \int \left(1 - e^{-\lambda_i x}\right) \mu(dx)}$$

Al utilizar $\lambda_j = 0$ si $j \neq i$ nos damos cuenta de que

$$\mathbb{E}\left(e^{-\lambda_i\left(X_{t_i}-X_{t_{i-1}}\right)}\right) = e^{-(t_i-t_{i-1})\int\left(1-e^{-\lambda_i x}\right)\mu(dx)}$$

y que por lo tanto los incrementos de X son independientes y estacionarios. Claramente X comienza en cero, por lo que resta probar que X tiene trayectorias càdlàg. $\square$

Corolario 3. *Se tiene que* $\mathbb{E}(X_t) = t\int x\,\mu(dx)$. *Si dicha cantidad es finita,* $Var(X_t) = t\int_0^\infty x^2\,\mu(dx)$.

CAPÍTULO 5

Procesos de Markov constantes por pedazos

En este capítulo analizaremos una clase de procesos estocásticos que generaliza al proceso de Poisson y tiene similitudes con las cadenas de Markov: los procesos de Markov a tiempo continuo con valores en un conjunto a lo más numerable E. Comenzaremos con un estudio del proceso de Poisson, al expresar su propiedad de Markov de forma que se parezca a la de las cadenas de Markov. Luego, se introducirá un segundo ejemplo, el de los procesos de nacimiento puro. Finalmente, comenzaremos el estudio de procesos con trayectorias constantes por pedazos y daremos una descripción probabilística de estos al introducir parámetros que determinan a un proceso de Markov: la distribución inicial y la matriz infinitesimal. Luego, estudiaremos sus probabilidades de transición mediante unas ecuaciones diferenciales que satisfacen y que están ligadas con su matriz infinitesimal: las ecuaciones backward y forward de Kolmogorov. Finalmente, veremos cómo las ecuaciones backward nos permiten estudiar a las distribuciones invariantes de los procesos de Markov a tiempo continuo.

1. El proceso de Poisson como proceso de Markov

El proceso de Poisson toma valores en los naturales. Como lo hemos definido, siempre comienza en cero, a diferencia de las cadenas de Markov que podemos comenzar en cualquier parte de su espacio de estados. Recordemos que si N es un proceso de Poisson de parámetro λ, al tener incrementos independientes y estacionarios, podemos escribir, para $= t_0 < t_1 < t_2 < \cdots < t_n$

$$\begin{aligned}&\mathbb{P}(N_{t_1} = k_1, N_{t_2} = k_2, \ldots, N_{t_n} = k_n)\\&= \mathbb{P}(N_{t_1} = k_1)\,\mathbb{P}(N_{t_2 - t_2} = k_2 - k_1)\cdots\mathbb{P}\big(N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1}\big)\,.\end{aligned}$$

Por lo tanto, si definimos

$$P_t(i,j) = \mathbb{P}(N_t = j - i) = e^{-\lambda t}\frac{(\lambda t)^n}{n!},$$

vemos que

$$\mathbb{P}(N_{t_1} = k_1, N_{t_2} = k_2, \ldots, N_{t_n} = k_n) = P_{t_1}(0, k_1)\,P_{t_2 - t_1}(k_1, k_2)\cdot P_{t_n}(k_{n-1}, k_n)\,.$$

La expresión anterior ya es muy paralela a la que vemos en cadenas de Markov; la entrada i, j de la matriz (infinita) P_t se puede interpretar como la probabilidad de ir de i a j en t unidades de tiempo. Esto además nos dice cómo podríamos definir a un proceso estocástico que fuera como el proceso de Poisson pero que comenzara en k: simplemente substituimos $P_{t_1}(0,k_1)$ por $P_{t_1}(k,k_1)$. Sin embargo, no tenemos un proceso estocástico que satisfaga lo anterior; demos una construcción de él.

Sea N un proceso de Poisson de parámetro λ. Si $t>0$, hemos visto que el proceso N^t dado por $N^t_s = N_{t+s} - N_t$ es un proceso de Poisson independiente de N_t. Por lo tanto:

$$\begin{aligned}
&\mathbb{P}(N_{t+s_1} = k_1, \ldots, N_{t+s_n} = k_n \,|N_t = k) \\
&= \frac{\mathbb{P}(N_{t+s_1} = k_1, \ldots, N_{t+s_n} = k_n, N_t = k)}{\mathbb{P}(N_t = k)} \\
&= \frac{\mathbb{P}(N_{s_1} = k_1 - k, \ldots, N_{s_n} = k_n - k)\,\mathbb{P}(N_t = k)}{\mathbb{P}(N_t = k)} \\
&= \mathbb{P}(N_{s_1} + k = k_1, \ldots, N_{s_n} + k = k_n)\,.
\end{aligned}$$

Vemos entonces que, la distribución condicional del proceso de Poisson en tiempos posteriores a t condicionalmente a $N_t = k$ es la misma que la del proceso $k+N$. Así, es natural definir al proceso de Poisson que comienza en k como $k+N$, pues este proceso tiene trayectorias como las del proceso de Poisson (aumenta de uno en uno en ciertos tiempos aleatorios) pero comienza en k. Además:

$$\mathbb{P}(N_{t+s_1} = k_1, \ldots, N_{t+s_n} = k_n \,|N_t = k) = P_{t_1}(k,k_1)\,P_{t_2-t_1}(k_1,k_2)\cdot P_{t_n}(k_{n-1},k_n)$$

Recordemos que en el caso de cadenas de Markov, se satisfacen las ecuaciones de Chapman-Kolmogorov. En este contexto, las ecuaciones se pueden expresar como sigue:

$$P_{t+s}(i,k) = \sum_j P_s(i,j)\,P_t(j,k)\,.$$

Ejercicio 5.1. Al utilizar el teorema del biniomio, pruebe directamente que la ecuación anterior se satisface. Dé además un argumento probabilístico, basado en condicionar con lo que sucede al tiempo s, para probar dicha ecuación.

La diferencia con las cadenas de Markov y el proceso de Poisson es que, al ser el segundo un proceso de Markov a tiempo continuo, en lugar de tener una sola matriz cuyas potencias nos permiten describir al proceso, tenemos toda una colección de matrices $(P_t, t \geq 0)$. Uno de los objetivos de este capítulo es mostrar como, con una definición adecuada, podemos generar a todas las matrices $(P_t, t \geq 0)$ mediante una sola matriz, la llamada matriz infinitesimal o matriz de tasas de transición. La idea es la siguiente: si x es un número, podemos interpolar a la sucesión $x^n, n \in \mathbb{N}$ mediante la función exponencial: $x^n = e^{n \log x}$. El lado derecho ya tiene sentido si $n \geq 0$ y no sólo si $n \in \mathbb{R}$. En cierto sentido, si $P_t = e^{tQ}$, podríamos interpretar a Q como el logaritmo de P_1. Para ver cómo podríamos obtener a Q, notemos que para obtener a $\log x$ si conocemos a $f(t) = e^{t \log x}$, simplemente calculamos $\log x = f'(0)$ y además $f'(t) = f'(0) f(t)$. En el caso de nuestra matriz P_t, vemos que

$$\begin{aligned}\frac{d}{dt} P_t(i,j) &= \frac{d}{dt} e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} \\ &= \frac{1}{(j-i)!} \left[e^{-\lambda t} (j-i) t^{j-i-1} \lambda^{j-i} \mathbf{1}_{j \geq i+1} - \lambda e^{-\lambda t} (\lambda t)^{j-i} \right]\end{aligned}$$

por lo que al evaluar en cero se obtiene:

$$\left. \frac{d}{dt} \right|_{t=0} P_t(i,j) = \begin{cases} -\lambda & j = i \\ \lambda & j = i+1 \\ 0 & j \neq i, i+1 \end{cases} .$$

Ejercicio 5.2. Sea

$$Q(i,j) = \begin{cases} -\lambda & j = i \\ \lambda & j = i+1 \\ 0 & j \neq i, i+1 \end{cases} .$$

Pruebe directamente que

$$\frac{d}{dt} P_t(i,j) = Q P_t(i,j) = P_t Q(i,j) \,, \tag{11}$$

donde QP_t es el producto de las matrices Q y P_t.

A las ecuaciones (diferenciales) (11) se les conoce como **ecuaciones de Kolmogorov**. Esbocemos una interpretación y deducción probabilística de dichas ecuaciones. Para calcular $P_{t+h}(0,j)$, podemos descomponer respecto del valor de N_h para obtener

$$P_{t+h}(i,j) = \mathbb{P}(N_{t+h} = j) = \mathbb{P}(N_h = i) \, \mathbb{P}(N_t = j - i) \,.$$

Por otra parte, al utilizar explícitamente la distribución Poisson, vemos que

$$\lim_{h \to 0} \frac{1 - \mathbb{P}(N_h = 0)}{h} = \lambda \qquad \lim_{h \to 0} \frac{\mathbb{P}(N_h = 1)}{h} = \lambda \qquad \lim_{h \to 0} \frac{\mathbb{P}(N_h \geq 2)}{h} = 0.$$

La interpretación es que es muy probable que haya cero saltos en un intervalo de longitud h, hay probabilidad proporcional al tamaño del intervalo de que haya un sólo salto, y muy improbable que haya más de dos saltos. Se concluye que

$$\begin{aligned}\frac{\partial P_t(i,k)}{\partial t} &= \lim_{h \to 0} \frac{P_h(i,i) \, P_t(i,k) - P_t(i,k)}{h} \\ &\quad + \lim_{h \to 0} \frac{P_h(i,i+1) \, P_t(i+1,k)}{h} \\ &= P_t(i+1,k) \, \lambda - P_t(i,j) \, \lambda = P_t Q(i,k)\end{aligned}$$

A esta ecuación se le denomina **ecuación hacia atrás de Kolmogorov** (también llamada, aún en español, ecuación backward) y fué obtenida al descomponer a

N_{t+h} respecto del valor de N_h. Ésta implica automáticamente la **ecuación hacia adelante de Kolmogorov**

$$\frac{\partial P_t(i,k)}{\partial t} = QP_t(i,k)$$

al utilizar la relación $P_t(i+1,j) = P_t(i,j-1)$.

2. El proceso de nacimiento puro

El proceso de nacimiento puro es una generalización del proceso de Poisson que puede presentar un fenómeno interesante: la explosión. Daremos una definición de dicho proceso paralela a la del proceso de Poisson.

Definición. Sean $q_0, q_1, \ldots \in (0, \infty)$. Un **proceso de nacimiento puro** es el proceso de contéo cuyos tiempos interarribo conforman una sucesión de variables exponenciales independientes $S_1, S_2, \ldots$ de parámetros $q_0, q_1, \ldots$.

La interpretación del proceso de nacimiento puro es que λ_i representa la tasa a la que un nuevo individuo se agrega a una población cuando esta tiene i elementos. Así, el proceso de nacimiento puro se construye al definir a los tiempos de salto

$$0 = T_0 \quad T_n = S_1 + \cdots + S_n \quad T_\infty = \sum_{n=0}^{\infty} S_i$$

y al proceso de contéo asociado

$$N_t = \sum_{n \in \mathbb{N}} n \mathbf{1}_{T_n \leq t < T_{n+1}}.$$

Claramente, cuando $q_i = \lambda$ para toda i el proceso que hemos construido es el proceso de Poisson. Sin embargo, cuando tenemos q_i dependientes de i se presenta un nuevo fenómeno. En efecto, en el caso del proceso de Poisson, sabemos que $T_n/n \to 1/\lambda = \mathbb{E}(S_1) > 0$ por la ley fuerte de los grandes números y por lo tanto $T_\infty = \infty$ casi seguramente. Sin embargo, en el caso del proceso de nacimiento puro puede suceder que $T_\infty < \infty$, en cuyo caso:

$$\lim_{t \to T_\infty -} N_t = \infty.$$

Decimos entonces que ha ocurrido la explosión (en este caso demográfica) del proceso de nacimiento puro y es natural definir

$$N_t = \infty \mathbf{1}_{T_\infty \leq t} + \sum_{n \in \mathbb{N}} n \mathbf{1}_{T_n \leq t < T_{n+1}}$$

pues con la definición anterior el proceso se vuelve cero despues de que la población haya explotado.

Puesto que el proceso de nacimiento puro admite una construcción sencilla, se puede dar un criterio explícito para la explosión, el cual además prueba que la explosión se da o con probabilidad cero o con probabilidad uno.

Proposición 5.1. *$T_\infty = \infty$ casi seguramente si $\mathbb{E}(T_\infty) = \sum_i 1/q_i = \infty$. Si $\sum_i 1/q_i < \infty$ entonces $T_\infty < \infty$ casi seguramente.*

Demostración. Si $\sum_i 1/q_i < \infty$ entonces podemos aplicar el teorema de convergencia monótona a la sucesión T_n para deducir que

$$\mathbb{E}(T_\infty) = \lim_n \mathbb{E}(T_n) = \lim_n \sum_{m \leq n} 1/q_m = \sum_n 1/q_n < \infty,$$

por lo que $T_\infty < \infty$ casi seguramente.

Por otra parte, si $\sum_i 1/q_i = \infty$, entonces podemos aplicar el teorema de convergencia acotada para ver que

$$\mathbb{E}\big(e^{-T_\infty}\big) = \lim_n \mathbb{E}\big(e^{-T_n}\big) = \lim_n \prod_{i=1}^n \frac{q_i}{1+q_i}.$$

Ahora dividiremos en dos partes nuestro análisis: si existe $\mu > 0$ tal que $q_i \leq \mu$ para una cantidad infinita de índices i entonces

$$\limsup_{n\to\infty} \prod_{i=1}^n \frac{q_i}{1+q_i} \leq \limsup_{n\to\infty} \left(\frac{1}{1+1/\mu}\right)^n = 0$$

conforme $n \to \infty$. Si por otra parte $q_i \to \infty$ entonces utilizamos el hecho de que $q_i \log\left(1 + 1/q_i\right) \to 1$ conforme $i \to \infty$ y por lo tanto, para alguna constante $C > 1$

$$\prod_{i=1}^n \frac{q_i}{1+q_i} = e^{-\sum_{i=1}^n \log(1+1/q_i)} \leq e^{-C\sum_{i=1}^n 1/q_i} \to 0.$$

Vemos que en cualquier caso

$$\mathbb{E}\big(e^{-T_\infty}\big) = 0,$$

lo cual nos dice que $T_\infty = \infty$ casi seguramente. □

Para analizar la propiedad de Markov del proceso de nacimiento puro, necesitamos definirlo cuando comienza en $i \in \mathbb{N}$. Una definición posible es que se trata del proceso $(N_{T_i+t}, t \geq 0)$. Este proceso va tomando los valores sucesivos $i, i+1, i+2, \ldots$ y los tiempos que permanece en cada uno de los estados son $S_i, S_{i+1}, \ldots,$. Pues este proceso tiene la misma estructura probabilística que i más un proceso de contéo cuyos tiempos interarribo son exponenciales independientes de parámetros $q_i, q_{i+1}, \ldots$, o sea, que una definición equivalente es que un proceso de nacimiento puro con parámetros $q_0, q_1, \ldots$ que comienza en i es i más un proceso de nacimiento puro con parámetros $q_i, q_{i+1}, \ldots$. Con estos preliminares podemos enunciar la propiedad de Markov del proceso de nacimiento puro.

Proposición 5.2. *Sean N un proceso de nacimiento puro de parámetros q_0, $q_1, \ldots$ y $s \geq 0$. Condicionalmente a $N_s = i$, el proceso estocástico $(N_{t+s} - i, t \geq 0)$ es un proceso de nacimiento puro de parámetros $q_i, q_{i+1}, \ldots$ y es condicionalmente independiente de $\mathscr{F}_s = \sigma(N_r : r \leq t)$.*

En otras palabras, condicionalmente a $N_s = i$, el proceso $N_{t+s}, t \geq 0$ es un proceso de nacimiento puro que comienza en i.

Demostración. El proceso $(N_{t+s}, t \geq 0)$ es un proceso de contéo cuyos tiempos interarribo son $T_{N_s+1} - s, S_{N_s+2}, S_{N_s+3}, \ldots$. Denotémoslos como $S^s_0, S^s_1, \ldots$. Debemos ver que condicionalmente a $N_s = i$, estos tiempos interarribo son exponenciales, independientes y de parámetros $q_i, q_{i+1}, \ldots$. En efecto:

$$\begin{aligned}&\mathbb{P}(N_s = i, S^s_0 > s_0, \cdots, S^s_n > s_n)\\ &= \mathbb{P}(T_i \leq s < T_i + S_i, s_0 + s < T_i + S_i, s_1 < S_{i+1}, \ldots, s_n < S_{i+n}).\end{aligned}$$

Al condicionar por T_i, utilizar la independencia de las variables $T_i, S_i, \ldots, S_{i+n}$ y el carácter exponencial de S_i

$$\begin{aligned}&\mathbb{P}(T_i \leq s < T_i + S_i, s_0 + s < T_i + S_i, s_1 < S_{i+1}, \ldots, s_n < S_{i+n}).\\ &= \mathbb{E}\big(\mathbf{1}_{T_i \leq s} e^{-\lambda_i}\,(s - T_i)\big)\, e^{-q_i s_1} \cdots e^{-q_{i+n} s_n}.\end{aligned}$$

Al utilizar $s_1, \ldots, s_n = 0$, vemos que

$$\mathbb{P}(N_s = i) = \mathbb{E}\big(\mathbf{1}_{T_i \leq s} e^{-\lambda_i}\,(s - T_i)\big)$$

y por lo tanto, al condicional por $N_s = i$, vemos que N^s es un proceso de nacimiento puro que comienza en i.

Falta demostrar que N^s es condicionalmente independiente de $\mathscr{F}_s$ dado que $N_s = i$, esto es, que para todo $A \in \mathscr{F}_s$, se tiene que

$$\begin{aligned}&\mathbb{P}\big(N^s_{t_1} = k_1, \ldots, N^s_{t_n} = k_n, A \,|N_s = i\big)\\ &= \mathbb{P}\big(N^s_{t_1} = k_1, \ldots, N^s_{t_n} = k_n \,|N_s = i\big)\, \mathbb{P}(A \,|N_s = i).\end{aligned}$$

Por clases monótonas, basta verificarlo cuando $A = \{N_{s_1} = j_1, \ldots, N_{s_m} = j_m\}$ con $s_1 \leq \cdots \leq s_m \leq t$ y $j_1 \leq \cdots \leq j_m \leq i$. Para esto, seguimos el mismo razonamiento anterior al notar que

$$A \cap \{N_s = i\} = \{T_{j_1} \leq s_1 < T_{j_1+1}, \cdots, T_{j_m} \leq s < T_{j_m+1}\} \cap \{T_i \leq s < T_{i+1}\}. \quad \square$$

Ejercicio 5.3. Haga un programa en Octave que simule al proceso de nacimiento puro que comienza en 1 si $q_i = i\lambda$ para algún $\lambda > 0$. ¿Este proceso explota? Haga otro programa que simule el caso $q_i = \lambda i^2$ y diga si ocurre explosión o no.

La propiedad de Markov se puede interpretar de manera similar a la del proceso de Poisson. Dados los parámetros $\mathbf{q} = (q_0, q_1, \ldots)$, sea $P_t(i, j)$ la probabilidad de que un proceso de nacimiento puro de parámetro $\mathbf{q}$ que comienza en i se encuentre en j al tiempo t. Entonces:

$$\begin{aligned}&\mathbb{P}(N_{s+t_1} = k_1, \ldots, N_{s+t_n} = k_n \,|N_s = k)\\ &= P_{t_1}(k, k_1)\, P_{t_2-t_1}(k_1, k_2) \cdots P_{t_n-t_{n-1}}(k_{n-1}, k_n).\end{aligned}$$

Aquí es más difícil obtener una expresión explícita para $P_t(i, j)$. Ésto se puede lograr cuando por ejemplo $\lambda_i = i\lambda$ para alguna $\lambda > 0$ (que es el caso del **Proceso**

de Yule, básico en la teoría de procesos de ramificación a tiempo continuo). Sin embargo, podemos argumentar por qué son válidas las ecuaciones hacia atrás de Kolmogorov: Si h es pequeño y nuestro proceso de nacimiento puro comienza en i, será raro que haya más de dos saltos en un intervalo de tiempo pequeño. Lo que se afirma es que, exactamente como en el caso de procesos de Poisson,

$$\lim_{h\to 0}\frac{1-P_h(i,i)}{h}=\lambda_i \lim_{h\to 0}\frac{P_h(i,i+1)}{h}=\lambda_i \qquad \lim_{h\to 0}\frac{\sum_{j\geq 2}P_h(i,j)}{h}=0.$$

Por lo tanto, las ecuaciones hacia atrás de Kolmogorov serían

$$\frac{\partial}{\partial t}P_t(i,j)=\lambda_i P_t(i+1,j)-\lambda_i P_t(i,j)\,.$$

Vemos entonces que la matriz infinitesimal o de tasas de transición estaría dada por

$$Q(i,j)=\begin{cases}-\lambda_i & i=j\\ \lambda_i & j=i+1\\ 0 & \text{en otro caso}\end{cases}$$

y que las ecuaciones hacia atrás se pueden escribir como

$$\frac{\partial}{\partial t}P_t(i,j)=QP_t(i,j)\,.$$

3. Matrices infinitesimales y construcción de procesos de Markov

Comenzaremos por describir a un tercer ejemplo de proceso de Markov a tiempo continuo. Este ejemplo se basa en calcular el mínimo de variables exponenciales independientes.

Comenzaremos con el caso en el que el espacio de estados E es finito. Para cada $x,y\in E$ con $x\neq y$ sea $\lambda_{x,y}\geq 0$. Sean

$$\{E_{i,x,y}, i\geq 1 : x,y\in E, x\neq y\}$$

variables aleatorias exponenciales independientes, tales que $E_{i,x,y}$ tiene parámetro $\lambda_{x,y}$. A $\lambda_{x,y}$ lo interpretaremos como la tasa a la que dejo el estado x para acceder al estado y y supondremos que para cada $x\in E$ existe $y\in E$ distinta de x tal que $\lambda_{x,y}>0$. Si $x\in E$, definamos a

$$T_1=\min_{y\neq x}E_{1,x,y}$$

En un momento veremos que existe un único $X_1\in E$ tal que

$$T_1=E_{1,x,X_1}.$$

Recursivamente, definiremos a

$$T_{n+1}=\min_{y\neq X_n}E_{n+1,X_n,y} \quad \text{donde } T_{n+1}=E_{1,X_n,X_{n+1}}.$$

Definamos finalmente, al proceso a tiempo continuo de interés:

$$Z_t = X_n \quad \text{si} \quad t \in [T_n, T_{n+1}).$$

Hay una forma de dar la construcción del proceso sin utilizar tantas variables exponenciales. Nos basamos en el siguente resultado.

Proposición 5.3. *Sean $T_1, \ldots, T_n$ variables exponenciales independientes de parámetros respectivos $\lambda_1, \lambda_2, \ldots$. Sea $T = \inf_i T_i$. Si $\sum_i \lambda_i < \infty$ entonces, con probabilidad 1, existe un único índice K tal que $T = T_K$. T y K son independientes, T tiene distribución exponencial de parámetro $\lambda = \sum_i \lambda_i$ y $\mathbb{P}(K = k) = \lambda_k/\lambda$.*

Demostración. Calculemos:

$$\begin{aligned}
\mathbb{P}(T \geq t, T_j > T_k \text{ para toda } j \neq k) &= \mathbb{P}(T_k \geq t, T_j > T_k \text{ para toda } j \neq k) \\
&= \mathbb{E}\left(\mathbf{1}_{T_k \geq t} \prod_{j \neq k} e^{-\lambda_k T_k}\right) \\
&= \int_t^\infty \lambda_k e^{-\lambda_k s} \prod_{j \neq k} e^{-\lambda_j s}\, ds \\
&= \frac{\lambda_k}{\lambda} e^{-\lambda t}.
\end{aligned}$$

Al evaluar en $t = 0$ vemos que

$$\mathbb{P}(T_j > T_k \text{ para toda } j \neq k) = \frac{\lambda_k}{\lambda}$$

por lo que al utilizar que los eventos anteriores son ajenos conforme variamos a k y su unión es la probabilidad de que el ínfimo de $T_1, T_2, \ldots$ se alcance para un sólo índice, vemos que

$$\mathbb{P}(\text{Existe un único } k \in \{1, \ldots, n\} \text{ tal que } T = T_k) = \sum_k \frac{\lambda_k}{\lambda} = \frac{\lambda}{\lambda} = 1.$$

Al sumar sobre k, vemos que

$$\mathbb{P}(T \geq t) = e^{-\lambda t},$$

lo cual implica que T es exponencial de parámetro λ

Puesto que con probabilidad 1 hay un único índice (aleatorio) en el que se alcanza el mínimo, digamos K tal que $T_K = T$, vemos que

$$\mathbb{P}(K = k, T \geq t) = \frac{\lambda_k}{\lambda} e^{-\lambda t} = \mathbb{P}(K = k)\, \mathbb{P}(T \geq t)\,. \qquad \square$$

Continuemos con el análisis de la cadena general. Notamos que los parámetros se pueden reducir a una matriz de transición P con ceros en la diagonal y un vector l de tasas totales de salto. Explícitamente,

$$l(x) = \sum_y \lambda_{x,y} \quad \text{y} \quad P_{x,y} = \frac{\lambda_{x,y}}{l(x)}.$$

Además, es posible construir nuestra cadena mediante una sucesión de variables exponenciales estándar independientes $\xi_1, \xi_2, \ldots$ entre sí y de una sucesión iid $U_1, U_2, \ldots$ de variables uniformes. En efecto, podemos definir a

$$X_0 = x, \quad T_0 = x \quad \text{y a} \quad T_1 = S_1 = \xi_1/l(X_0)\,.$$

Luego, utilizamos a la variable uniforme U_1 para escoger a un elemento aleatorio X_1 de E tal que $\mathbb{P}(X_1 = y) = P_{X_0,y}$. Finalmente, continuamos este procedimiento de manera recursiva: si ya tenemos definidos a $X_0, \ldots, X_n$ y a $T_0, \ldots, T_n$ en términos de $S_1, \ldots, S_n$ y $U_1, \ldots, U_n$ entonces definimos

$$S_{n+1} = \xi_{n+1}/l(X_n)\,, \quad T_{n+1} = T_n + S_{n+1}$$

y utilizamos a las variables U_{n+1} y X_n para construir a X_{n+1} de tal manera que

$$\mathbb{P}(X_{n+1} = y \,|X_n = x) = P_{x,y}.$$

Finalmente, recordemos que

$$Z_t = X_n \quad \text{si} \quad T_n \leq t < T_{n+1}.$$

Puesto que el espacio de estados es finito, se sigue que existe $\varepsilon > 0$ tal que $l(x) > \varepsilon$ para toda $x \in E$. Por lo tanto:

$$T_n \geq \frac{\xi_1 + \cdots + \xi_n}{\varepsilon} \to \infty.$$

Así, en este caso no debemos preocuparnos por el fenómeno de explosión.

Resulta ser que Z es un proceso de Markov a tiempo continuo. Formalmente, sea $P_t(x, y)$ la probabilidad de que $Z_t = y$ cuando $Z_0 = x$ y probemos que

$$\begin{aligned}&\mathbb{P}_{x_0}(X_{t_1} = x_1, \ldots, X_{t_n} = y_n)\\&= P_{t_1-t_0}(x_0, x_1)\, P_{t_2-t_1}(x_1, x_2) \cdots P_{t_n-t_{n-1}}(x_{n-1}, x_n)\end{aligned}$$

donde $\mathbb{P}_x$ es la medida de probabilidad que rige a Z cuando comienza en x y $0 = t_0 < t_1 < \cdots < t_n$. El argumento es parecido a cuando probamos que el proceso de Poisson (o más generalmente el de nacimiento y muerte) es un proceso de Markov, se basa en notar que el proceso posterior a s, condicionalmente a $Z_s = x$ tiene la misma construcción probabilística que Z comenzando en x. Esto es, que sus tiempos y lugares de salto se pueden obtener a partir de una sucesión de variables uniformes y exponenciales estándar independientes. Lo haremos en la siguiente sección sin tener que asumir que el espacio de estados es finito.

Gracias a la propiedad de Markov concluimos que se satisfacen las ecuaciones de Chapman-Kolmogorov. En efecto, vemos que

$$\begin{aligned} P_{t+s}(x,z) &= \mathbb{P}_x(Z_{t+s}=z) \\ &= \sum_y \mathbb{P}_x(Z_s=y, Z_{t+s}=z) \\ &= \sum_y P_s(x,y)\, P_t(y,z)\,. \end{aligned}$$

Ahora utilizaremos que en espacio de estados finito, digamos de cardinalidad n, la ecuaciones backward

$$\frac{\partial P_t(x,t)}{\partial t} = QP_t(x,y)$$

admiten una solución en términos de la matriz infinitesimal Q y esto nos permitir introducir a las ecuaciones forward. Cuando I es finito, debemos resolver el sistema de ecuaciones

$$\frac{\partial}{\partial t}P_t = QP_t.$$

Este es un sistema lineal y si Q fuera de tamaño 1×1, tendría como única solución a la función exponencial. Lo mismo sucede en el caso $n\times n$, si definimos a la matriz

$$e^{tQ} = \sum_{n=0}^{\infty} \frac{t^nQ^n}{n!}$$

para cualquier $t\in\mathbb{R}$.La convergencia de la serie se sigue pues la sucesión

$$\sum_{n=0}^{N} \frac{t^nQ^n}{n!}.$$

es de Cauchy cuando se utiliza la norma

$$\|Q\| = \max\left\{\|Qx\| : x\in\mathbb{R}^n, \|x\|=1\right\}.$$

En efecto, puesto que esta norma es submultiplicativa, se sigue que:

$$\sup_{n\geq m} \|\sum_{k=0}^{n}\frac{t^kQ^k}{k!} - \sum_{k=0}^{m}\frac{t^kQ^k}{k!}\| \leq \sum_{k=m+1}^{\infty}\frac{|t|^{\,k}\|Q\|^k}{k!} \to_{m\to\infty} 0.$$

Ahora veremos que $e^{tQ}, t\geq 0$ es la única solución a las ecuaciónes backward y forward de Kolmogorov:

$$\frac{\partial}{\partial t}e^{tQ} = Qe^{tQ} \quad \text{y} \quad \frac{\partial}{\partial t}e^{tQ} = e^{tQ}Q.$$

Además, satisfacen las ecuaciones de Chapman-Kolmogorov

$$e^{(s+t)Q} = e^{sQ}e^{tQ}.$$

En efecto, Chapman-Kolmogorov se sigue de la definición de la exponencial de una matriz. Por otra parte, podemos derivar término a término la serie de potencias (cuyo radio de convergencia, con la norma matricial, es infinito) para obtener

$$\frac{\partial}{\partial t}e^{tQ}=\sum_{k=1}^{\infty}\frac{t^{k-1}Q^{k}}{k!}=\sum_{k=0}^{\infty}\frac{t^{k}Q^{k+1}}{k!}=\begin{cases}Qe^{tQ}\\ e^{tQ}Q\end{cases},$$

lo cual muestra que se satisfacen las ecuaciones backward y forward. Además $e^{0Q}=\mathrm{Id}$. Para la unicidad de la solución a estas últimas, supongamos que $P_t, t\geq 0$ es una colección de matrices en $\mathbb{R}^n$ que satisface las ecuaciones backward (el argumento para las forward es similar) y tal que $P_0=\mathrm{Id}$. Notemos que la inversa de e^{tQ} es e^{-tQ}. Entonces

$$\frac{\partial}{\partial t}e^{-tQ}P_t=-Qe^{-tQ}P_t+e^{-tQ}QP_t=-Qe^{-tQ}P_t+Qe^{-tQ}P_t=0.$$

Por lo tanto $e^{-tQ}P_t$ es constante y como la constante es $P_0=\mathrm{Id}$, vemos que $P_t=e^{tQ}$.

4. Cadenas de Markov constantes por pedazos

Ahora presentaremos un marco conceptual que incluye al proceso de Poisson, al proceso de nacimiento puro y a las cadenas construidas mediante tasas de transición. El objetivo será construir procesos de Markov con trayectorias continuas por la derecha y lo haremos mediante la introducción de un espacio canónico adecuado.

Sea E un conjunto a lo más numerable, $\Delta\notin E$ algún punto que denominaremos **cementerio** y a donde mandaremos las trayectorias de un proceso de Markov cuando explote, y sea Ω^{can} el conjunto de funciones $f:[0,\infty)\to E\cup\{\Delta\}$ que satisfacen:

(1) si $f(t)\in E$ entonces existe $\delta>0$ tal que $f(s)=f(t)$ para $s\in[t,t+\delta]$,
(2) si $f(t)=\Delta$ entonces $f(s)=\Delta$ para toda $s\geq t$ y
(3) si $t_1<t_2<\cdots$, $t_n\to t<\infty$ y $f(t_{n+1})\neq f(t_n)$ entonces $f(t)=\Delta$.

Definiremos también a $X_t:\Omega\to E\cup\{\Delta\}$ dado por $X_t(f)=f(t)$, a $\mathscr{F}=\sigma(X_t:t\geq 0)$ y a $\mathscr{F}_t=\sigma(X_s:s\leq t)$. Entonces $(\Omega,\mathscr{F})$ es el llamado **espacio canónico de trayectorias constantes por pedazos y minimales** sobre E (con cementerio Δ), X es el **proceso canónico** y $(\mathscr{F}_t,t\geq 0)$ es su **filtración canónica** asociada. Podemos además definir a los **operadores de translación** $\theta_t:\Omega\to\Omega$ mediante la asignación $\theta_t f(s)=f(s+t)$.

Definición. Una colección de matrices $(P_t,t\geq 0)$ indexada por E se llama **semigrupo de probabilidades de transición** si:

$$P_0(x,y)=\mathbf{1}_{x=y}\qquad\qquad P_t(x,y)\geq 0$$

$$\sum_{y\in E}P_t(x,y)\leq 1\qquad \text{y}\qquad P_{s+t}(x,z)=\sum_{y\in E}P_s(x,y)\,P_t(y,z)\,.$$

A la última condición se le puede escribir con la notación de producto de matrices como $P_{s+t} = P_s P_t$ y se conoce como **ecuación de Chapman-Kolmogorov**.

Sea $P = (P_t, t \geq 0)$ un semigrupo de probabilidades de transición. Una **cadena de Markov** con semigrupo P y distribución inicial ν es un proceso estocástico Z definido en algún espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$, con valores en E, cuyas trayectorias pertenecen a Ω^{can} y tal que si $x_1, \ldots, x_n \in E$ y $\leq t_1 \leq \ldots \leq t_n$ entonces

$$\mathbb{P}(Z_{t_0} =) = \sum_{x_0 \in E} \nu_{x_0} P_{t_1}(x_0, x_1)\, P_{t_2 - t_1}(x_1, x_2) \cdots P_{t_n - t_{n-1}}(x_{n-1}, x_n)\,.$$

Una cadena de Markov Z con distribución inicial ν y semigrupo de transición $P = (P_t, t \geq 0)$ induce una medida de probabilidad $\mathbb{P}_\nu$ en Ω^{can}. En términos del proceso canónico, la medida está determinada por

$$\mathbb{P}_x(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = \nu_x P_{t_1}(x, x_1)\, P_{t_2 - t_1}(x_1, x_2) \cdots P_{t_n - t_{n-1}}(x_{n-1}, x_n)$$

gracias a:

Ejercicio 5.4. Si $\mathbb{P}$ y $\mathbb{Q}$ son medidas de probabilidad en Ω^{can} y coinciden sobre $\mathscr{C} = \cup_{t \geq 0} \mathscr{F}_t$ entonces $\mathbb{P} = \mathbb{Q}$ en $\mathscr{F}$.

Si $\nu_x > 0$ para toda $x \in E$ entonces podemos definir a las medidas $\mathbb{P}_x$ en Ω^{can} donde $\mathbb{P}_x$ es igual a $\mathbb{P}_\nu$ condicionada por $X_0 = 0$. Notemos que una consecuencia de las ecuaciones de Chapman-Kolmogorov es que

$$\mathbb{P}_\nu(\, X_{t+s} = y \mid \mathscr{F}_s) = P_t(X_s, y)\,.$$

En efecto, el resultado se prueba por clases monótonas al notar que si $A = \{X_{s_1} = x_1, \ldots, X_{s_m} = x_m\}$ para $s_1 \leq \cdots \leq s_m \leq s$ entonces

$$\begin{aligned} \mathbb{E}_\nu\big(\mathbf{1}_A \mathbf{1}_{X_{t+s}=y}\big) &= \sum_x \mathbb{E}_\nu\big(\mathbf{1}_A \mathbf{1}_{X_s = x} \mathbf{1}_{X_{t+s}=y}\big) \\ &= \sum_x \mathbb{E}_\nu(\mathbf{1}_A \mathbf{1}_{X_s = x} P_t(x, y)) \\ &= \mathbb{E}_\nu(\mathbf{1}_A P_t(X_s, y))\,. \end{aligned}$$

Para cualquier $f : E \cup \{\Delta\} \to \mathbb{R}$ tal que $f(\Delta) = 0$, se define a la función $P_t f : E \cup \{\Delta\} \to \mathbb{R}$ mediante

$$P_t f(x) = \mathbb{E}_x(f(X_t))\,.$$

(Si a las funciones sobre $E \cup \{\Delta\}$ nulas en Δ se les interpreta vectores renglón, $P_t f$ corresponde a la multiplicación de P_t con f). Una primera muestra de la utilidad de la noción anterior es que nos permite expresar de manera compacta cálculos de esperanza condicional como el siguiente:

$$\mathbb{E}(\, f(X_{t+s}) \mid \mathscr{F}_s) = P_t f(X_s)\,.$$

Podemos generalizar lo anterior al utilizar operadores de traslación.

Proposición 5.4 (Propiedad de Markov). *Si $F : \Omega^{can} \to \mathbb{R}$ es medible, acotada y nula en la función idénticamente igual a Δ entonces*

$$\mathbb{E}_x(F \circ \theta_t \mid \mathscr{F}_t) = \mathbb{E}_{X_t}(F) .$$

Demostración. Primero se prueba por inducción que el resultado es válido si

$$F(f) = g_1 \circ X_{t_1} \cdots g_n \circ X_{t_n}$$

donde $g_1, \ldots, g_n : E \cup \{\Delta\} \to \mathbb{R}$ son nulas en Δ y $t_1 \leq \cdots \leq t_n$.

Luego, se prueba que las funciones anteriores generan a $\mathscr{F}$, lo cual prueba el resultado deseado. □

Definición. Una **familia Markoviana** en Ω es una colección de medidas de probabilidad $(\mathbb{P}_x, x \in E)$ definidas en $(\Omega, \mathscr{F})$ tales que:

(1) $\mathbb{P}_x(X_0 = x) = 1$,
(2) para toda función $F : \Omega \to \mathbb{R}$ medible y acotada que sea cero en $t \mapsto \Delta$:

$$\mathbb{E}_x(F \circ \theta_t \mid \mathscr{F}_t) = \mathbb{E}_{X_t}(F) .$$

Así, a cada semigrupo de probabilidades de transición es posible asignarle una familia markoviana. A continuación, veremos que cada familia Markoviana está caracterizada por dos colecciones numéricas, su distribución inicial y su matriz de tasas de transición. Inversamente, veremos que dado un parámetro, podemos construir a la familia Markoviana asociada. Veamos primero cómo definir este parámetro.

Sea $(\mathbb{P}_x)_{x \in E}$ una familia Markoviana. Consideremos a los tiempos aleatorios

$$T_0 = 0, \quad T_{n+1} = \inf \{t \geq T_n : X_t \neq X_{T_n}\} \quad \text{y} \quad \zeta = \lim_{n \to \infty} T_n$$

con la convención $\inf \emptyset = \infty$.

Proposición 5.5. *T_n y ζ son tiempos de paro respecto de la filtración canónica.*

Existen tres categorías para las trayectorias en términos de estos tiempos aleatorios:

Absorción: Cuando existe n tal que $T_n < \infty = T_{n+1}$, en cuyo caso $X_t = X_{T_n}$ para toda $t \geq T_n$,
Explosión: cuando $\zeta < \infty$ y
Movimiento perpetuo: cuando $T_n < \infty$ para toda n y $\zeta = \infty$.

Proposición 5.6. *Bajo $\mathbb{P}_x$, T_1 es exponencial de parámetro $c(x) \in [0, \infty)$. Si $c(x) > 0$ entonces las variables X_{T_1} y T_1 son independientes.*

Demostración. Al utilizar la propiedad de Markov, vemos que

$$\begin{aligned}\mathbb{P}_x(T_1 > t+s) &= \mathbb{P}_x(T_1 > s, T_1 \circ \theta_s > t)\\ &= \mathbb{E}_x(\mathbf{1}_{T_1>s}\mathbb{E}_{X_s}(T_1 > t))\\ &= \mathbb{E}_x(\mathbf{1}_{T_1>s}\mathbb{E}_x(T_1 > t))\\ &= \mathbb{P}_x(T_1 > s)\,\mathbb{P}_x(T_1 > t)\end{aligned}$$

y por lo tanto, bajo $\mathbb{P}_x$, T_1 es exponencial.

Sea ahora F medible y acotada en el espacio canónico. Entonces

$$\begin{aligned}&\mathbb{P}_x(\mathbf{1}_{T_1>t} F \circ \theta_{T_1})\\ &= \mathbb{P}_x(\mathbf{1}_{T_1>t} F \circ \theta_{T_1} \circ \theta_t)\\ &= \mathbb{P}_x(\mathbf{1}_{T_1>t})\,\mathbb{E}_x(F \circ \theta_{T_1}),\end{aligned}$$

por lo que T_1 es independiente de $X \circ \theta_{T_1}$. □

A $c(x)$ la interpretamos como la tasa a la que dejamos el estado x. Definamos ahora

$$P_{x,y} = \begin{cases} 0 & c(x) = 0 \\ \mathbb{P}_x(X_{T_1} = y) & c(x) \neq 0 \end{cases} \quad \text{y} \quad \alpha(x,y) = c(x)\,P_{x,y}.$$

A α se le conoce como la matriz de tasas de transición y la interpretación de $\alpha(x,y)$ es la tasa a la que dejamos el estado x para pasar al estado y. La matriz de tasas de transición es el parámetro que nos permitirá caracterizar a la familia Markoviana. Para verificar por qué, es necesario extender la propiedad de Markov.

Teorema 5.1 (Propiedad de Markov fuerte). *Sea T un tiempo de paro finito y*

$$\mathscr{F}_T = \{A \in F : F \cap \{T \leq t\} \in \mathscr{F}_t \text{ para toda } t \geq 0\}.$$

Entonces $\mathscr{F}_T$ es una σ-álgebra y X_T es $\mathscr{F}_T$-medible. Sea $\theta_T f(t) = f(T(f) + t)$. Entonces θ_T es medible y para toda variable aleatoria $F : \Omega \to \mathbb{R}$ acotada:

$$\mathbb{E}_x(F \circ \theta_T \mid \mathscr{F}_T) = \mathbb{E}_{X_T}(F).$$

Demostración. Es un buen ejercicio probar que $\mathscr{F}_T$ es una σ-álgebra. Al utilizar la constancia por pedazos de X y aproximar a T mediante la sucesión de tiempos de paro $T_n = \lceil T2^n \rceil / 2^n$, es otro buen ejercicio probar que X_T es $\mathscr{F}_T$-medible. De igual manera, al utilizar el hecho de que

$$\theta_T = \lim_{n\to\infty} \sum_k \mathbf{1}_{T_n = (k+1)/2^n} \theta_{(k+1)/2^n},$$

que se sigue de la constancia por pedazos de las trayectorias de X, vemos que θ_T es medible. Finalmente, al utilizar la propiedad de Markov al instante $(k+1)/2^n$

vemos que si $A \in \mathscr{F}_T$

$$\begin{aligned}\mathbb{E}_x\left(\mathbf{1}_{T_n=(k+1)/2^n}\mathbf{1}_A F\circ\theta_{(k+1)/2^n}\right) &= \mathbb{E}_x\left(\mathbf{1}_{k/2^n<k\leq(k+1)/2^n}\mathbf{1}_A F\circ\theta_{(k+1)/2^n}\right)\\ &= \mathbb{E}_x\left(\mathbf{1}_{k/2^n<T\leq(k+1)/2^n}\mathbf{1}_A \mathbb{E}_{X_{(k+1)/2^n}}(F)\right).\end{aligned}$$

Al sumar sobre k obtenemos

$$\mathbb{E}_x(\mathbf{1}_A F\circ\theta_{T_n}) = \mathbb{E}_x\left(\mathbf{1}_A\mathbb{E}_{X_{T_n}}(F)\right).$$

Ahora especialicemos a funciones F de la forma $f_1(X_{t_1})\cdots f_m(X_{t_m})$. Al utilizar la constancia por pedazos de X y tomar el límite conforme $n\to\infty$, vemos que

$$\mathbb{E}_x(\mathbf{1}_A F\circ\theta_T) = \mathbb{E}_x(\mathbf{1}_A\mathbb{E}_{X_T}(F))$$

al menos para estas funciones especiales. Sin embargo, el lema de clases de Dynkin nos permite extender la igualdad anterior primero a indicadoras de elementos de $\mathscr{F}$ y luego, por aproximación, a funciones F que sean $\mathscr{F}$ medibles y acotadas. □

El teorema anterior nos permitirá caracterizar a la familia Markoviana en términos de la matriz de tasas de transición α, o equivalentemente, de c y P. Sea Z el proceso estocástico a tiempo discreto definido por

$$Z_n = X_{T_n}$$

si $T_n < \infty$. En el caso absorbente, definimos $Z_{n+m} = Z_n$ para toda $m\geq 1$ si $T_n < \infty = T_{n+1}$.

Teorema 5.2. *El proceso Z es una cadena de Markov de matriz de transición P que comienza en x bajo $\mathbb{P}_x$. Si $c(x)>0$ para toda $x\in E$, condicionalmente a Z, las variables $S_1, S_2, \ldots$ con $S_i = T_i - T_{i-1}$ son independientes y exponenciales de parámetros $c(Z_0), c(Z_1), \ldots$.*

Demostración. Al utilizar el lema de clases de Dynkin, vemos que es suficiente verificar que

$$\begin{aligned}&\mathbb{P}_x(Z_1=x_1,\ldots,Z_n=x_n,S_1>t_1,\ldots,S_n>t_n)\\ &=P_{x,x_1}\cdots P_{x_{n-1},x_n}e^{-\lambda_1 t_1}\cdots e^{-\lambda_n t_n}.\end{aligned}\tag{12}$$

Esto se sigue por inducción al utilizar la propiedad de Markov fuerte. La base inductiva es la Proposición 5.6. Por otra parte, si suponemos válida la ecuación (12) vemos que al aplicar la propiedad de Markov fuerte al instante T_n y la Proposición 5.6 se sigue que

$$\begin{aligned}&\mathbb{P}_x(Z_1=x_1,\ldots,Z_{n+1}=x_{n+1},S_1>t_1,\ldots,S_{n+1}>t_{n+1})\\ &=\mathbb{P}_x(Z_1=x_1,\ldots,Z_n=x_n,S_1>t_1,\ldots,S_n>t_n)\,e^{-c(x_n)t_{n+1}}P_{x_n,x_{n+1}}\end{aligned}$$

puesto que $Z_{n+1}=Z_1\circ\theta_{T_n}$ y $S_{n+1}=S_1\circ\theta_{T_n}$. □

Dada una función $\alpha : E \times E \to [0, \infty)$ tal que $c(x) = \sum_y \alpha(x, y) < \infty$, podemos definir a $P_{x,y} = \alpha(x, y) / c(x)$ cuando $c(x) = 0$ y a $P_{x,y} = \delta_{x,y}$ cuando $c(x) = 0$ y preguntarnos cuándo existe una cadena de Markov a tiempo continuo cuya matriz de tasas de transición sea α. Nos abocaremos ahora a verificar que se puede construir una familia markoviana cuya matriz de tasas de transición sea α. En efecto, sean $\tilde{S}_1, \tilde{S}_2, \ldots$ variables aleatorias exponenciales de parámetro 1 y Z una cadena de Markov con matriz de transición P que comienza en X. Ahora definamos

$$T_0 = 0, \quad \text{y} \quad T_{n+1} = T_n + \tilde{S}_n / c(Z_n) .$$

Consideremos al proceso X definido mediante

$$\tilde{X}_t = \begin{cases} Z_n & \text{si } t \in [T_n, T_{n+1}) \\ \Delta & \text{si } T_\infty \leq t \end{cases} .$$

Las trayectorias del proceso así construido pertenecen a Ω y entonces podemos definir a $\mathbb{P}_x$ como la probabilidad imagen de $\tilde{X}$. Se afirma que $(\mathbb{P}_x, x \in E)$ es una familia Markoviana cuya matriz de tasas de transición es α. Por definición, bajo la medida de probabilidad $\mathbb{P}_x$ es válida la ecuación (12).

Teorema 5.3. *La colección $(\mathbb{P}_x)_{x \in E}$ es una familia Markoviana con matriz de tasas de transición α.*

Demostración. Puesto que $\mathscr{F} = \sigma(Z, S)$, basta mostrar que para toda $A \in \mathscr{F}_t$:

$$\begin{aligned} &\mathbb{P}_x(A, X_t = x_0, Z_1 \circ \theta_t = x_1, S_1 \circ \theta_t > t_1, \ldots, Z_n \circ \theta_t = x_n, S_n \circ \theta_t > t_n) \\ &= \mathbb{P}_x(A, X_t = x_0) \, \mathbb{P}_{x_0}(Z_1 = x_1, S_1 > t_1, \ldots, Z_n = x_n, S_n > t_n) , \end{aligned}$$

(donde por supuesto tenemos una expresión explícita para el segundo factor del lado derecho).

Sea N_t la cantidad de saltos que tiene X en $[0, t]$. Entonces

$$N_t = \sum_{n=1}^{\infty} \mathbf{1}_{T_n \leq t}$$

lo que implica la igualdad

$$S_1 \circ \theta_t = S_{N_t+1} - (t - T_{N_t}) , S_2 \circ \theta_t = S_{N_t+2}, \ldots, S_n \circ \theta_t = S_{N_t+n}.$$

Por otra parte, se tiene que

$$Z_1 \circ \theta_t = Z_{N_t+1}, \ldots, Z_n \circ \theta_t = Z_{N_t+n}.$$

Podemos entonces calcular

$$\begin{aligned}
&\mathbb{P}_x(A, X_t = x_0, N_t = m, Z_1 \circ \theta_t = x_1, S_1 \circ \theta_t > t_1, \ldots, Z_n \circ \theta_t = x_n, S_n \circ \theta_t > t_n) \\
&= \sum_m \mathbb{P}_x(A, Z_m = x_0, N_t = m, \\
&\qquad\qquad Z_{m+1} = x_1, S_{m+1} - (t - T_m) > t_1, \ldots, Z_{n+m} = x_n, S_{n+m} > t_n).
\end{aligned}$$

Es fácil ver que existe $B \in \sigma(S_1, \ldots, S_m, Z_1, \ldots, Z_m)$ tal que

$$A \cap \{Z_m = x_0, N_t = m\} = B \cap \{Z_m = x_0, T_m \leq t < T_m + S_{m+1}\}$$

Al condicionar por $S_1, \ldots, S_m$ y por $Z_0, \ldots, Z_m$ se obtiene

$$\begin{aligned}
&\mathbb{P}_x(A, X_t = x_0, N_t = m, Z_1 \circ \theta_t = x_1, S_1 \circ \theta_t > t_1, \ldots, Z_n \circ \theta_t = x_n, S_n \circ \theta_t > t_n) \\
&= \mathbb{E}_x\left(\mathbf{1}_{B, Z_m = x_0, T_m \leq t} e^{-c(x_0)(t_1 - (t - T_m))}\right) \\
&\qquad\qquad e^{-c(x_1)t_2} \cdots e^{-c(x_{n-1})t_n} P_{x_0, x_1} \cdots P_{x_{m-1}, x_m}.
\end{aligned}$$

Del caso particular $t_1 = 0$ se obtiene que de hecho

$$\begin{aligned}
&\mathbb{P}_x(A, X_t = x_0, N_t = m, Z_1 \circ \theta_t = x_1, S_1 \circ \theta_t > t_1, \ldots, Z_n \circ \theta_t = x_n, S_n \circ \theta_t > t_n) \\
&= \mathbb{P}_x(A, X_t = x_0, N_t = m)\, \mathbb{P}_{x_0}(S_1 > t_1, \ldots, S_n > t_n, Z_1 = x_1, \ldots, Z_n = x_n)\,.
\end{aligned}$$

□

Ahora utilizaremos el teorema anterior para dar una segunda demostración de la caracterización del proceso Poisson como proceso de Lévy.

TEOREMA 5.4. *Si N es un proceso de contéo con incrementos independientes y estacionarios entonces es un proceso de Poisson.*

DEMOSTRACIÓN. Sea $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ el espacio de probabilidad donde está definido N. Definiremos a $\mathbb{P}_i$ como la imagen de $\tilde{\mathbb{P}}$ bajo la función $i + N : \tilde{\Omega} \to \Omega$ donde $i + N(\omega) = (i + N_t(\omega))_{t \geq 0}$. Verificaremos que $(\mathbb{P}_i)_{i \in \mathbb{N}}$ es una familia Markoviana. Sea α la matriz de tasas de transición asociada. Puesto que N es un proceso de contéo, vemos que $\alpha(x, y) > 0$ si y sólo si $y = x + 1$. Por otra parte, puesto que $\mathbb{P}_i$ es la imagen de $\mathbb{P}_0$ bajo la aplicación $f \mapsto f + i$ entonces $\alpha(x, x + 1) = \alpha(0, 1)$, digamos que ambas cantidades son λ. Por lo tanto vemos que los tiempos entre los saltos de N son exponenciales independientes del mismo parámetro λ. Así, N es un proceso de Poisson de intensidad λ.

Para verificar que $(\mathbb{P}_i)_{i \in \mathbb{N}}$ es una familia markoviana, definamos a $f_t(x) = \tilde{\mathbb{P}}(N_t = x)$ y a $p_t(x, y) = f_t(y - x)$. Puesto que N tiene incrementos independientes y estacionarios, entonces

$$\mathbb{P}_i(X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = p_{t_1}(i, i_1)\, p_{t_2 - t_1}(i_1, i_2) \cdots p_{t_n - t_{n-1}}(i_{n-1}, i_n)\,.$$

Se sigue entonces que si $s_1 \leq \cdots \leq s_m \leq t \leq t_1 \leq t_n$ entonces

$$\begin{aligned}&\mathbb{P}_i(X_{s_1} = i_1, \ldots, X_{s_m} = i_m, X_t = j, X_{t_1} = j_1, \ldots, X_{t_n} = j_n)\\&= \mathbb{P}_i(X_{s_1} = i_1, \ldots, X_{s_m} = i_m, X_t = j)\, \mathbb{P}_j(X_{t_1} = j_1, \ldots, X_{t_n} = j_n)\,.\end{aligned}$$

Al utilizar el lema de clases de Dynkin se puede probar que si $A \in \mathscr{F}_t$ y $B \in \mathscr{F}$ entonces

$$\mathbb{P}_i\big(A, \theta_t^{-1}(B)\big) = \mathbb{E}_i(\mathbf{1}_A \mathbb{E}_{X_t}(B))\,,$$

de lo cual se sigue que $(\mathbb{P}_i, i \in \mathbb{N})$ es una familia markoviana. □

Esta prueba se puede abreviar al no utilizar directamente la caracterización de las familias markovianas sino sólo algunas de las ideas de la prueba. Primero se prueba que si T_1 es el momento del primer salto de N entonces $(N_{T_1+s} - 1, s \geq 0)$ tiene la misma distribución que N y es independiente de $\mathscr{F}^N_{T_1} = \sigma(T_1)$. Luego, se prueba que T_1 es exponencial al satisfacer la propiedad de pérdida de memoria mediante un argumento que apele a la propiedad de Markov: $T_1 > t + s$ si y sólo si $T_1 > s$ y el primer salto del proceso $(N_{r+s}, r \geq 0)$ es mayor a t. Finalmente, se utiliza inducción para probar que los tiempos que permanece en cada natural son exponenciales y que éstos tiempos son independientes.

Tal como la propiedad de Markov y de Markov fuerte nos llevan a relaciones de recurrencia para probabilidades que deseamos calcular, en tiempo continuo nos llevan a ecuaciones diferenciales. Una de ellas es la ecuación backward de Kolmogorov. Sea $(\mathbb{P}_x, x \in E)$ una familia markoviana con **probabilidades de transición** $P_t(x, y) = \mathbb{P}_x(X_t = y)$. Estas probabilidades de transición satisfacen las ecuaciones de Chapman-Kolmogorov

$$p_{t+s}(x, z) = \sum_y p_s(x, y)\, p_t(y, z)$$

puesto que al aplicar la propiedad de Markov

$$\mathbb{P}_x(X_{t+s} = z) = \sum_{y \in E} \mathbb{P}_x(X_s = y, X_{t+s} = z) = \sum_{y \in E} \mathbb{P}_x(X_s = y)\, \mathbb{P}_y(X_t = z)\,.$$

Definiremos al semigrupo de transición asociado $(P_t, t \geq 0)$ de la siguiente manera: para cada función $f : E \to \mathbb{R}$ acotada (que definimos en Δ como 0), $P_t f : E \to \mathbb{R}$ será la función dada por

$$P_t f(x) = \mathbb{E}_x(f(X_t))\,.$$

Vemos que entonces

$$p_t(x, y) = P_t \mathbf{1}_y(x)\,.$$

Así, las ecuaciones de Chapman-Kolmogorov y la linealidad de la esperanza implican la propiedad de semigrupo:

$$P_t(P_s f) = P_{t+s} f.$$

Teorema 5.5 (Ecuaciones backward de Kolmogorov). *Para toda $f : E \to \mathbb{R}$ acotada, la función $t \mapsto P_t f$ es derivable y*

$$\frac{\partial}{\partial t} P_t f(x) = \sum_{y \in E} \alpha(x, z) \left[P_t f(z) - P_t f(x) \right].$$

Dada la matriz de tasas de transición, definiremos a la matriz infinitesimal Q mediante:

$$Q_{x,y} = \begin{cases} \alpha(x, y) & x \neq y \\ -c(x) & x = y \end{cases}.$$

Entonces la ecuación backward de Kolmogorov se puede escribir en la forma

$$\frac{\partial}{\partial t} P_t f(x) = Q P_t f.$$

Esto explica la conexión con ecuaciones diferenciales: las probabilidades de transición de una familia markoviana satisfacen una ecuación diferencial. Hemos visto que en el caso de espacio de estados finito, la teoría clásica de ecuaciones diferenciales lineales nos permite verificar que existe una única solución para la ecuación backward de Kolmogorov y por lo tanto nos da una manera de obtener, a veces explícitamente pero inclusive también numéricamente, a las probabilidades de transición de la familia Markoviana.

Demostración. Heurísticamente, la prueba es una aplicación de la propiedad de Markov fuerte. Sin embargo, necesitamos una verisón que también depende del tiempo. Específicamente, notemos que si $s \leq t$

$$P_t f(x) = \mathbb{E}_x(f(X_t)) = \mathbb{E}_x(\mathbb{E}_{X_s}(f(X_{t-s}))) = \mathbb{E}_x(P_{t-s} f(X_s)).$$

Podemos por lo tanto pensar que por la propiedad de Markov fuerte aplicada al tiempo de paro $\sigma = t \wedge T_1$ se satisface

$$P_t f(x) = \mathbb{E}_x(P_{t-\sigma}(X_\sigma)) \tag{13}$$

para $t > 0$. Esto es en efecto cierto pero no se ha demostrado y se sigue del hecho de que, a partir de la representación de una familia markoviana, podemos aproximar al tiempo de paro σ por $\sigma^n = \lceil \sigma 2^n \rceil / 2^n$ y tomar el límite conforme $n \to \infty$ para verificar (13). Antes de esto, veamos cómo se aplica dicha ecuación. De (13) se deduce que:

$$P_t f(x) = \mathbb{E}_x(P_{t-\sigma} f(X_\sigma)) = f(x)\, e^{-c(x)t} + \int_0^t \sum_y e^{-c(x)s} \alpha(x, y)\, P_{t-s} f(y)\, ds.$$

Al multiplicar por $e^{c(x)t}$ de ambos lados se obtiene

$$e^{c(x)t} P_t f(x) = f(x) + \int_0^t e^{c(x)s} \sum_y \alpha(x, y)\, P_s f(y)\, ds.$$

Finalmente, la expresión del lado derecho muestra que el lado izquierdo es derivable y por la continuidad del integrando vemos que

$$\frac{\partial}{\partial t} P_t f(x) + c(x)\, P_t f(x) = \sum_y \alpha(x,y)\, P_t f(y)\,.$$

Para concluir la prueba falta justificar la ecuación (13). Notemos que $x \mapsto P_t f(x)$ es continua en t. Esto sucede pues

$$\mathbb{P}_x(t = T_n \text{ para alguna } n \geq 1) = 0,$$

puesto que cada T_n es (condicionalmente a $Z_1, Z_2, \ldots$) una suma de variables exponenciales. Entonces, vemos que

$$\begin{aligned} P_{t+1/2^n} f(x) &= \mathbb{E}_x\big(f\big(X_{t+1/2^n}\big)\big) \\ &= \mathbb{E}_x\left(\sum_{k:k/2^n \leq t} f\big(X_{t+1/2^n}\big)\, \mathbf{1}_{\sigma^n=(k+1)/2^n}\right) \\ &= \mathbb{E}_x\left(\sum_{k:k/2^n \leq t} \mathbf{1}_{\sigma^n=(k+1)/2^n} P_{t+1/2^n-\sigma^n} f(X_{\sigma^n})\right) \\ &= \mathbb{E}_x\big(P_{t+1/2^n-\sigma^n} f(X_{\sigma^n})\big) \\ &\to_{n\to\infty} \mathbb{E}_x(P_{t-\sigma} f(X_\sigma))\,, \end{aligned}$$

lo que prueba (13) y nos permite concluir. □

5. Distribuciones invariantes

Ahora pasaremos al estudio de las distribuciones invariantes para familias markovianas. La liga entre el tiempo continuo y discreto nos lo proporciona el siguiente resultado que se sigue de las ecuaciones backward de Kolmogorov.

Definición. Decimos que una distribución ν en E (identificada con la colección numérica $\nu_x = \nu(\{x\})$) es **invariante** para una familia Markoviana si

$$\sum_x \nu_x P_t(x,y) = \nu_y$$

para toda $t \geq 0$.

En otras palabras, la distribución ν es invariante si la distribución de X_t bajo $\mathbb{P}_\nu = \sum_x \nu_x \mathbb{P}_x$ es igual a la de X_0.

Teorema 5.6. *Una medida de probabilidad ν tal que $\sum_x \nu_x c(x) < \infty$ es invariante para X si y sólo si $c\nu = (c_x \nu_x, x \in E))$ es invariante para la cadena asociada.*

Demostración. Por la ecuación backward de Kolmogorov y el teorema de Fubini-Tonelli se sigue que
(14)
$$\sum_x \nu_x P_t(x,z) = \sum_x \nu_x P_0(x,z) + \int_0^t \sum_y \sum_x \nu_x \alpha(x,y)\, [P_s(y,z) - P_s(x,z)]\, ds.$$

Así, ν es invariante si y sólo si la integral del lado derecho es igual a 0 para cualquier t. Escribamos a $\alpha(x,y) = c(x)\, P(x,y)$, donde P es la matriz de transición de la cadena asociada. Puesto que $\sum_x c_x \nu_x < \infty$ y $t \mapsto P_t(x,y)$ es continua, podemos aplicar el teorema de convergencia dominada para concluir que el integrando en el lado derecho de (14) es continuo. Por lo tanto, ν es invariante si y sólo si

$$\begin{aligned} 0 &= \sum_y \sum_x \nu_x \alpha(x,y)\, [P_0(y,z) - P_0(x,z)] \\ &= \sum_y \sum_x c(x)\, \nu_x P_{x,y}\, [\mathbf{1}_{y=z} - \mathbf{1}_{x=z}] = \sum_x c(x)\, \nu_x P_{x,z} - c_z \nu_z \end{aligned}$$

En otras palabras, $c\nu$ es invariante para la cadena asociada. □

Recordemos que en el teorema fundamental de convergencia para cadenas de Markov (en tiempo discreto) la periodicidad juega un rol importante. Ahora veremos que en tiempo continuo, en cierto sentido el proceso ya es periódico.

Proposición 5.7. *$P_t(x,y) > 0$ para alguna $t > 0$ si y sólo si $P_t(x,y) > 0$ para toda $t > 0$. En particular $P_t(x,x) > 0$ para toda $t \geq 0$.*

Demostración. El caso particular es simple:

$$P_t(x,x) \geq \mathbb{P}_x(T_1 > t) > 0.$$

Por otra parte, si $y \neq x$ y para la cadena asociada se accede de x a y entonces existen $x_0, \ldots, x_n \in E$ tales que $x_0 = x$, $x_n = y$ y $x_{k+1} \neq x_k$ para los cuales

$$P_{x,x_1} \cdots P_{x_{n-1},y} > 0.$$

En particular, se tiene que $c(x_i) > 0$ para $i < n$.

Si $S_1, \ldots, S_{n+1}$ son exponenciales de parámetro 1 independientes entonces

$$P_t(x,y) \geq \mathbb{P}\left(\sum_{k \leq n} \frac{S_k}{c(x_{k-1})} \leq t < \sum_{k \leq n+1} \frac{S_k}{c(x_{k-1})} \right) P(x_0,x_1) \cdots P_{x_{n-1},y} > 0.$$

(Sólo se debe tener cuidado si $c(y) = 0$.)

Finalmente, si de x no se accede a y para la cadena asociada Z entonces

$$\mathbb{P}_x(X_t \neq y \text{ para toda } t \geq 0) = \mathbb{P}_x(Z_n \neq y \text{ para toda } n \geq 0) = 1.$$ □

Una familia markoviana es **irreducible** si $\mathbb{P}_x(X_t = y) > 0$ para toda $t > 0$ y todas $x, y \in E$.

Proposición 5.8. *Si la cadena asociada a una familia markoviana irreducible es recurrente entonces no hay explosión.*

Lo anterior nos dice que los conjuntos $\{t \geq 0 : X_t = y\}$ y $\{n \in \mathbb{N} : Z_n = y\}$ o son ambos acotados o ambos no acotados para familias markovianas irreducibles. En el primer caso hablamos de **transitoriedad** y en el segundo de **recurrencia**.

Demostración. Sólo hay que notar que si $\mathbb{P}_x(Z_n = x \text{ i.o. })$ entonces o Z_n se absorbe en x (que sucede si y sólo si $c(x) > 0$ y no es compatible con la irreducibilidad de la cadena) ó $c(x) > 0$ y

$$\sum_n \frac{1}{c(Z_n)} \geq \infty/c(x) = \infty$$

$\mathbb{P}_x$-casi seguramente, en cuyo caso, al condicionar con Z, vemos que no hay explosión. (Recordemos que si τ_i son exponenciales independientes de parámetro λ_i entonces $\sum \tau_i = \infty$ casi seguramente si y sólo si $\sum 1/\lambda_i = \infty$.) □

Teorema 5.7. *Si* $(\mathbb{P}_x)$ *es una familia markoviana irreducible entonces son equivalentes:*

(1) *Existe una única distribución invariante* ν *para la familia que satisface* $\nu_x > 0$ *para toda* $x \in E$ *y para cualquier distribución inicial* μ*:*

$$\lim_{t\to\infty} \sum_x |\mathbb{P}_\nu(X_t = y) - \nu_y| = 0.$$

(2) *Para alguna* $h > 0$*, la sucesión de variables aleatorias* $(X_{nh}, n \in \mathbb{N})$ *es una cadena de Markov positivo recurrente.*

En caso contrario, no existe ninguna distribución invariante y $\mathbb{P}_x(X_t = y) \to 0$ *conforme* $t \to \infty$*.*

Demostración. Sólo demostraremos la equivalencia. (La prueba completa se puede verificar en [**Kal02**].)

Sea $h > 0$. Notemos que $(X_{nh}, n \geq 0)$ es una cadena de Markov con matriz de transición $P_h(x,y)\,, x, y \in E$. En efecto, vemos que

$$\mathbb{P}_x(X_h = x_1, \ldots, X_{nh} = x_n) = P_h(x, x_1)\, P_h(x_1, x_2) \cdots P_h(x_{n-1}, x_n)\,.$$

Si para alguna h, dicha cadena de Markov es positivo recurrente, entonces al ser irreducible y aperiódica, existe una única distribución invariante ν^h. Por otra parte, la cadena de Markov $X_{nh/2}n \geq 0$ debe también ser positivo recurrente pues su tiempo de primer retorno está acotado por dos veces el tiempo de primer retorno de $X_{nh}, n \geq 0$, el cual es integrable. Así, existe una única distribución invariante para $X_{nh/2}$, digamos $\nu^{h/2}$ pero como ésta también es invariante para X_{nh}, vemos que $\nu_{h/2} = \nu^h$. Escribamos por lo tanto $\nu = \nu^h$. Generalizando, vemos que para cualquier racional no-negativo q, la distribución de X_{qh} bajo $\mathbb{P}_\nu$ es ν y, al aproximar a cualquier $t > 0$ por la derecha por reales de la forma qh, vemos que ν es invariante para la familia markoviana. Para mostrar la convergencia en variación, notemos que, de acuerdo al teorema fundamental de convergencia para cadenas de Markov, se tiene que

$$\sum_x |P_{nh}(x,y) - \nu_y| \to 0$$

conforme $n \to \infty$. Por lo tanto, al escribir a t (de manera única) en la forma $nh+r$ con $n \in \mathbb{N}$ y $0 \leq r < h$, las ecuaciones de Chapman-Kolmogorov y la invariancia de ν nos dicen que

$$\sum_x |P_t(x,y) - \nu_y| \leq \sum_x \sum_y |P_{nh}(x,z) - \nu_z| P_r(z,y) \to 0.$$

Por lo tanto, el teorema de convergencia dominada nos permite afirmar que

$$\lim_{t\to\infty} \sum_x |\mathbb{P}_\nu(X_t = y) - \nu_y| = 0.$$

Por otra parte, si existe una distribución invariante ν para la familia markoviana, entonces ν es una distribución invariante para X_{nh}, lo que implica que esta es positivo recurrente para cualquier $h > 0$. □

Finalmente, pasamos a la relación entre el comportamiento asintótico de la probabilidad de transición y los tiempos medios de recurrencia. Sea

$$T^y = \min\{t > T_1 : X_t = y\}.$$

Teorema 5.8. *Si y no es absorbente entonces*

$$\lim_{t\to\infty} P_t(x,y) = \frac{\mathbb{P}_x(T^y < \infty)}{c(y)\,\mathbb{E}_y(T^y)}.$$

Demostración. Sólo podremos probarlo en el caso transitorio y positivo recurrente. En el caso nulo recurrente, tendremos la convergencia en el sentido de Cesàro.

Primero nos concentraremos en el caso $x = y$. Si x es transitorio entonces $\mathbb{E}_y(T^y) = \infty$ y por lo tanto el enunciado es válido. Si por otra parte y es positivo recurrente y nos concentramos en su clase de comunicación, esta será irreducible y sabemos que $P_t(x,y)$ converge a ν_y donde ν es la distribución invariante única (en la clase de comunicación de y). Así, los tiempos medios de ocupación

$$L_t = \frac{1}{t}\int_0^t \mathbf{1}_{X_s=y}\,ds$$

satisfacen:

$$\mathbb{E}_x(L_t) = \frac{1}{t}\int_0^t \mathbb{P}_x(X_s = y)\,ds \to \nu_y.$$

Por otra parte, si $\tilde{T}^y_n = T^y + \tilde{T}^y_{n-1} \circ \theta_{T^y}$, la propiedad de Markov fuerte nos dice que $\tilde{T}^y_n$ es una caminata aleatoria. Como $T^y \circ \theta_{T_1}$ se puede acotar en términos del tiempo de visita a y por la cadena $X_{nh} \circ \theta_{T_1}$, que es finito por ser positivo recurrente, vemos que $\mathbb{E}_x\left(\tilde{T}^y\right) < \infty$, por lo que podemos aplicar la ley fuerte de los grandes números y deducir que bajo $\mathbb{P}_y$ se tiene que $\tilde{T}^y_n/n \to \mathbb{E}_y\left(\tilde{T}^y_n\right)$. Por esto, observamos que

$$\frac{L_{\tilde{T}^y_n}}{\tilde{T}^y_n} = \frac{\xi_1 + \cdots + \xi_n}{T^y_n} \to \frac{1}{c(y)\,\mathbb{E}_x(T_y)}$$

donde $\xi_i = T_1 \circ \theta_{\tilde{T}^y_n}$ son variables exponenciales de parámetro $c(y)$ (a la cuales también les aplicamos la ley fuerte de los grandes números). Finalmente, por convergencia dominada vemos que

$$\mathbb{E}_x(L_t) \to \frac{1}{c(y)\,\mathbb{E}_x(T_y)},$$

lo cual prueba el resultado en este caso □

CAPÍTULO 6

El movimiento browniano

Consideremos una caminata aleatoria simple y simétrica $S = (S_n, n \in \mathbb{N})$. El teorema límite central afirma que $S_n/\sqrt{n}$ converge débilmente a una variable normal estándar. Una manera de interpretar al movimiento browniano es como una extensión multidimensional (inclusive infinito-dimensional o funcional) del teorema límite central. En efecto, si S se extiende por interpolación lineal en cada intervalo $[n, n+1]$ y consideramos al proceso estocástico S^n dado por $S^n_t = S_{nt}/\sqrt{n}$, vemos que S^n_t converge débilmente a una normal de media 0 y varianza t. Por otra parte, como S tiene incrementos independientes y estacionarios (cuando nos restringimos a instantes de tiempo naturales) entonces si $0 = t_0 < t_1 < \cdots < t_m$ entonces para n suficientemente grande los incrementos $S^n_{t_i} - S^n_{t_{i-1}}$, con $1 \leq i \leq m$ son independientes. Por lo tanto, vemos que dichos incrementos convergen débilmente a un vector aleatorio con entradas gaussianas independientes de varianzas respectivas $t_i - t_{i-1}$ para $1 \leq i \leq m$. El movimiento browniano es justamente un proceso estocástico que recoge este comportamiento límite de las caminatas aleatorias.

Definición. Un **movimiento browniano en ley** es un proceso estocástico $B = (B_t, t \geq 0)$ tal que:

(1) $B_0 = 0$
(2) B tiene incrementos independientes: si $0 = t_0 < t_1 < \cdots < t_m$ entonces $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq m$ son independientes
(3) B tiene incrementos estacionarios: $B_{t+s} - B_t$ tiene la misma distribución que B_s y
(4) la distribución de B_t es normal de media 0 y varianza t.

Un **movimiento browniano** es un movimiento browniano en ley que tiene trayectorias continuas.

Lo primero que se debe hacer es verificar que existen procesos estocásticos que son movimientos brownianos o movimientos brownianos en ley. Para esto, recordaremos algunos aspectos de vectores gaussianos.

1. Vectores gaussianos

A continuación, trabajaremos con variables aleatorias con valores en $\mathbb{R}^n$; a los elementos de $\mathbb{R}^n$ los tomaremos como vectores columna. Si $x \in \mathbb{R}^n$ e $i \in \{1, \ldots, n\}$, denotaremos por x_i a su i-ésima coordenada y si x' es el vector transpuesto de x, escribiremos $x' = (x_1, \dots, x_n)$ ó $x = (x_1, \dots, x_n)'$.

Definición. Un **vector gaussiano** es una variable aleatoria X definida en un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ y con valores en $\mathbb{R}^n$ y tal que para cualquier $\lambda \in \mathbb{R}^n$, $\lambda \cdot X$ tiene distribución normal. Asociado a un vector gaussiano está el vector de medias $\mu = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))'$ y la matriz de varianzas-covarianzas Σ (de tamaño $n \times n$) tal que $\Sigma_{i,j} = \Sigma_{j,i} = Cov(X_i, X_j)$.

Primero recordaremos, o más bien formalizaremos, algunos cálculos para el caso unidimensional. Sea X una variable aleatoria normal de media μ y varianza σ^2. Entonces X tiene la misma distribución que $\sigma N + \mu$ donde N es una variable normal estándar.

(1) Calculemos la función generadora de momentos de X en términos de la de N:

$$\mathbb{E}\left(e^{uX}\right) = \mathbb{E}\left(e^{u\sigma N + u\mu}\right) = e^{u\mu}\mathbb{E}\left(e^{u\sigma N}\right).$$

(2) Calculemos ahora la función generadora de momentos de N:

$$\mathbb{E}\left(e^{uN}\right) = \int_{-\infty}^{\infty} e^{ux} e^{-x^2/2} \frac{1}{\sqrt{2\pi}}\, dx = \int_{-\infty}^{\infty} e^{u^2/2} e^{-(x-u)^2/2} \frac{1}{\sqrt{2\pi}}\, dx = e^{u^2/2}.$$

(3) Concluimos que

$$\mathbb{E}\left(e^{uX}\right) = e^{u\mu} e^{-u^2\sigma^2/2}.$$

(4) Probemos la desigualdad

$$\mathbb{P}(N > x) \leq e^{-x^2/2}$$

si $x > 0$. Ésta desigualdad se sigue del siguiente razonamiento: para $x, \lambda > 0$:

$$e^{\lambda x}\mathbb{P}(\xi > x) \leq \mathbb{E}\left(e^{\lambda \xi}\mathbf{1}_{\xi > x}\right) \leq \mathbb{E}\left(e^{\lambda\xi}\right) = e^{-\lambda^2/2},$$

por lo cual para cualquier $\lambda > 0$,

$$\mathbb{P}(\xi > x) \leq e^{-\lambda x + \lambda^2/2}.$$

Al minimizar el lado derecho de la expresión anterior sobre $\lambda > 0$ (el mínimo ocurre cuando $\lambda = x$), se obtiene la desigualdad deseada.

(5) Calculemos ahora los momentos de N; como su distribución es simétrica, los momentos de orden impar son cero. De hecho, la simetría implica que $\mathbb{E}\left(e^{u|N|}\right) < \infty$ para toda $u \in \mathbb{R}$. Por el teorema de convergencia monótona

$$\sum_n u^n \mathbb{E}\left(|N|^{\,2n}\right) \frac{1}{2n!} = \mathbb{E}\left(e^{u|N|}\right) < \infty$$

por lo cual todos los momentos de orden par son finitos y además, por convergencia dominada vemos que

$$\sum_n u^n \mathbb{E}\big(N^{2n}\big) \frac{1}{2n!} = \mathbb{E}\big(e^{uN}\big)$$

(Otra forma de verlo es puesto que $x^n e^{-x^2/2} = x^n e^{-x^2/4} e^{-x^2/4}$, y $x \mapsto x^n e^{-x^2/4}$ es acotada y $x \mapsto e^{-x^2/4}$ es integrable, se sigue que todos los momentos son finitos). Esto implica que la función generadora de momentos es infinitamente diferenciable y que su enésima derivada en cero es el momento de orden n de N. Sea ϕ_N la generadora de momentos de la gaussiana. Hemos visto que $\Phi_N(u) = e^{u^2/2}$, por lo que $\phi_N^{(2n)}(0) = 2n!/n!2^n$. Así:

$$\mathbb{E}\big(N^{2n}\big) = \frac{2n!}{n!2^n}.$$

Ahora, como la serie de momentos de N es absolutamente convergence, el teorema de convergencia dominada nos permite afirmar que

$$\mathbb{E}\big(e^{iuN}\big) = \sum_{n=0}^{\infty} \mathbb{E}(N^n) \frac{u^n}{n!} (i)^n = e^{-u^2/2}.$$

(6) Un caso particular muy útil es que $\mathbb{E}\big(N^4\big) = 3$.
(7) Ahora calculemos los momentos de $|N|$, ya tenemos a los momentos de orden par; los de orden impar se calculan de manera distinta:

$$\begin{aligned}
\mathbb{E}\big(|N|^{2n+1}\big) &= \int_0^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} x^{2n}\, 2x\, dx \\
&= \int_0^\infty \frac{e^{-y/2}}{\sqrt{2\pi}} y^n\, dy \\
&= \frac{2^{n+1}}{\sqrt{2\pi}} n! \\
&= 2^{n+1/2} n!/\sqrt{\pi}.
\end{aligned}$$

Ahora veremos que la distribución de un vector gaussiano está determinada por μ y A, tal como la distribución gaussiana está determinada por la media y la varianza. Para esto, sea $\lambda \in \mathbb{R}^n$ y calculemos la media y la varianza de $\lambda \cdot X$: la media es

$$\mathbb{E}(\lambda \cdot X) = \sum_{i=1}^{n} \lambda_i \mathbb{E}(X_i) = \sum_{i=1}^{n} \lambda_i \mu_i = \lambda \cdot \mu$$

y la varianza es

$$\begin{aligned}
\mathrm{Var}(\lambda \cdot X) &= \mathrm{Var}\left(\sum_{i=1}^{n} \lambda_i X_i\right) \\
&= \sum_{i,j=1}^{n} \mathbb{E}(\lambda_i \, (X_i - \mu_i) \, \lambda_j \, (X_j - \mu_j)) \\
&= \sum_{i,j=1}^{n} \lambda_i \lambda_j A_{i,j} \\
&= \lambda' A \lambda.
\end{aligned}$$

Recordemos que las variables aleatorias en $\mathbb{R}^n$ están determinadas por su función característica. Como $\lambda \cdot X$ es una variable aleatoria gaussiana, se sigue que

$$\mathbb{E}\big(e^{i\lambda \cdot X}\big) = e^{i\lambda \cdot \mu} e^{-\lambda' A \lambda/2}.$$

Se sigue por lo tanto que X tiene una distribución normal multivariada con media μ y matriz de varianzas-covarianzas A. Se deduce el siguiente corolario importante.

Corolario 4. *Las entradas de un vector gaussiano son independientes si y sólo si son no-correlacionadas.*

La prueba se basa en notar que si las entradas de un vector gaussiano son no-correlacionadas entonces la matriz de varianzas-covarianzas es diagonal lo cual implica que la función característica se factoriza y que por lo tanto las entradas son independientes.

Necesitaremos ver que la convergencia débil de variables aleatorias gaussianas implica la convergencia de momentos.

Proposición 6.1. *Si X_n es una sucesión de variables gaussianas que converge débilmente a una variable aleatoria X entonces X es gaussiana, $|X_n|^{\,p}$ es uniformemente integrable para toda $p > 0$ y $\mathbb{E}(X_n^p) \to \mathbb{E}(X^p)$.*

Demostración. Sean $\mu_n = \mathbb{E}(X_n)$ y $\sigma_n^2 = \mathrm{Var}(X_n)$. Por hipótesis

$$e^{iu\mu_n - \sigma_n^2 u^2/2} = \mathbb{E}\big(e^{iuX_n}\big) \to \mathbb{E}\big(e^{iuX}\big)$$

para toda $u \in \mathbb{R}$. Vemos entonces que

$$e^{-\sigma_n^2 u^2/2} = \left|\mathbb{E}\big(e^{iuX_n}\big)\right| \to \left|\mathbb{E}\big(e^{iuX}\big)\right|,$$

por lo que $e^{-\sigma_n^2}$ es convergente, digamos a σ^2. Esto nos muestra que μ_n es también una sucesión acotada. En efecto, si por ejemplo $\mu_n \to \infty$ entonces

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(X_n - \mu_n \leq x - \mu_n) \leq \mathbb{P}\Big((X_n - \mu_n)^2 \leq (x - \mu_n)^2\Big) \leq \frac{\sigma_n^2}{(x - \mu_n^2)} \to 0,$$

por lo que X_n no puede converger en distribución y contradice la hipótesis. Por medio de subsucesiones y de considerar a $-X_n$ podemos reducirnos a este caso y concluir que μ_n es una sucesión acotada.

Como $e^{iu\mu_n}$ es convergente para cualquier $u \in \mathbb{R}$, esto no muestra que cualesquiera dos límites subsucesionales μ^1 y μ^2 de μ_n satisfacen $\mu^1 - \mu^2 = 2k\pi/u$ para todo $u \in \mathbb{R}$ donde $k \in \mathbb{Z}$, lo cual forza la igualdad $k = 0$ y por lo tanto $\mu^1 = \mu^2$. Esto implica que μ_n converge, digamos a μ, y por lo tanto

$$\mathbb{E}\left(e^{iuX}\right) = e^{iu\mu - \sigma^2 u^2/2},$$

por lo cual X es normal de media μ y varianza σ^2. Vemos que además,

$$\mathbb{E}\left(e^{uX_n}\right) \to \mathbb{E}\left(e^{uX}\right) < \infty,$$

por lo que para toda $p > 1$ se tiene que $\sup_n \mathbb{E}(|X_n|^{\,p}) < \infty$ y por lo tanto $(|X_n|^{\,p})$ es uniformemente integrable para toda $p \geq 1$, lo cual a su vez implica que $\mathbb{E}(|X_n|^{\,p}) \to \mathbb{E}(|X|^{\,p})$ para todo $p \geq 1$. $\square$

Ahora haremos algunos cálculos con la distribución gaussiana. Primero, calculemos la distribución de N^2:

$$\mathbb{P}\left(N^2 \leq x\right) = 2\mathbb{P}\left(0 \leq N \leq \sqrt{x}\right),$$

por lo que

$$f_{N^2}(x) = \frac{e^{-x/2}}{\sqrt{2\pi}} \frac{1}{\sqrt{x}}.$$

Se concluye que N^2 tiene distribución Γ de parámetros $(1/2, 1/2)$, donde el primer parámetro es el de posición y el segundo el de escala. (Si $\gamma_{a,b}$ tiene distribución Γ de parámetros (a, b) entonces

$$\mathbb{P}(\gamma_{a,b} \in dx) = \frac{1}{\Gamma(a)} (bx)^{a-1}\, be^{-bx}\, dx,$$

por lo que $c\gamma_{a,b} \sim \Gamma(a, b/c)$ y esto último explica el nombre de parámetro de escala.) Ahora calcularemos la distribución de N_2/N_1, donde N_1 y N_2 son gaussianas independientes: primero calculamos la densidad de $(N_1, N_2/N_1)$ al utilizar la transformación $(x, z) \mapsto (x, zx)$, cuyo jacobiano es x, vemos que

$$f_{N_1, N_2/N_1}(x, z) = f_{N_1, N_2}(x, zx)\, x = e^{-x^2\left(1+z^2\right)/2} \frac{x}{2\pi}.$$

Al integrar z en la expresión anterior, utilizando el cambio de variable $y = x^2/2$, obtenemos:

$$f_{N_2/N_1}(z) = \int e^{-x^2\left(1+z^2\right)/2} \frac{x}{2\pi}\, dx = 2\int e^{-y\left(1+z^2\right)} \frac{1}{2\pi}\, dy = \frac{1}{\pi\left(1+z^2\right)}.$$

Se sigue que N_2/N_1 tiene ley Cauchy; la distribución asociada se puede explicitar en términos de la función arcoseno. Sea C una variable aleatoria Cauchy; ahora caracterizaremos a la distribución de $A = 1/(1+C^2)$. Como

$$\mathbb{P}(A \geq x) = \mathbb{P}\big(1/x \geq 1 + C^2\big) = \mathbb{P}\big(1/x - 1 \geq C^2\big) = 2\mathbb{P}\Big(0 \leq C \leq \sqrt{1/x - 1}\Big),$$

entonces

$$f_A(x) = -2 f_C\Big(\sqrt{(1-x)/x}\Big) \frac{1}{2\sqrt{(1-x)/x}} \frac{-1}{x^2} = \frac{1}{\pi\sqrt{x(1-x)}}.$$

Por lo tanto, A tiene distribución Beta de parámetros $1/2, 1/2$, que es la llamada distribución arcoseno. Así, vemos también que $N_1^2/\left(N_1^2 + N_2^2\right)$ tiene distribución arcoseno.

Cuando $X_1, \ldots, X_\delta$ son independientes y normales estándar, se puede calcular la distribución de $X_1/\|\mathbf{X}\|$ mediante el siguiente razonamiento: como para $x > 0$

$$2\mathbb{P}(0 \leq X_1/\|\mathbf{X}\| < x) = \mathbb{P}\big(X_1^2 < \left(X_2^2 + \cdots X_\delta^2\right) x^2/(1-x^2)\big),$$

entonces

$$\begin{aligned}
&2 f_{X_1/\|\mathbf{X}\|}(x) \\
&= \frac{\partial}{\partial x} \int_0^\infty dy\, \frac{1}{2^{\nu+1/2}\Gamma(\nu+1/2)} y^{\nu-1/2} e^{-y/2} \int_0^{yx^2/(1-x^2)} dz\, \frac{1}{\sqrt{2}\Gamma(1/2)} z^{-1/2} e^{-1/2z} \\
&= \int_0^\infty dy\, \frac{1}{2^{\nu+1/2}\Gamma(\nu+1/2)} y^{\nu-1/2} e^{-y/2} y \frac{2x}{(1-x^2)^2} \frac{1}{\sqrt{2}\Gamma(1/2)} \left(y\frac{x^2}{1-x^2}\right)^{-1/2} e^{-y\frac{x^2}{2(1-x^2)}} \\
&= 2\frac{1}{\Gamma(\nu+1/2)\,\Gamma(1/2)} \left(1-x^2\right)^{-3/2} \int_0^\infty dy\, \frac{1}{2^{\nu+1}} y^{\nu-1} e^{y\frac{1}{2x^2}} \\
&= 2\frac{\Gamma(\nu+1)}{\Gamma(\nu+1/2)\,\Gamma(1/2)} \left(1-x^2\right)^{\nu-1/2}.
\end{aligned}$$

Notemos $X_1/\|\mathbf{X}\|$ y $\|\mathbf{X}\|$ son independientes pues la distribución de $\mathbf{X}$ es invariante ante transformaciones ortogonales. La interpretación ahora es clara: como X_1^2 tiene distribución Γ de parámetros $1/2$ y $1/2$ se sigue que $\|\mathbf{X}\|$ tiene distribución Γ de parámetros $\delta/2$ y $1/2$, es independiente de $X_1/\|\mathbf{X}\|$ cuya distribución es la de S_δ, por lo que se ha verificado la factorización de la distribución normal cuando δ es un entero positivo. La interpretación de la ley arcoseno es ahora clara: se trata de la distribución de $|X_1|/\|\mathbf{X}\|$ cuando $\delta = 2$.

Una muestra de la utilidad de la definición de vector gaussiano es la siguiente caracterización del movimiento browniano en ley.

Definición. Un proceso estocástico $X = (X_t, t \geq 0)$ es un **proceso gaussiano** si para cualesquiera $t_1, \ldots, t_n \geq 0$ el vector aleatorio $(X_{t_1}, \ldots, X_{t_n})$ es un vector gaussiano.

Ejercicio 6.1. Un proceso estocástico $B = (B_t, t \geq 0)$ es un movimiento browniano en ley si y sólo si es un proceso gaussiano centrado y $\mathbb{E}(B_s B_t) = s \wedge t$.

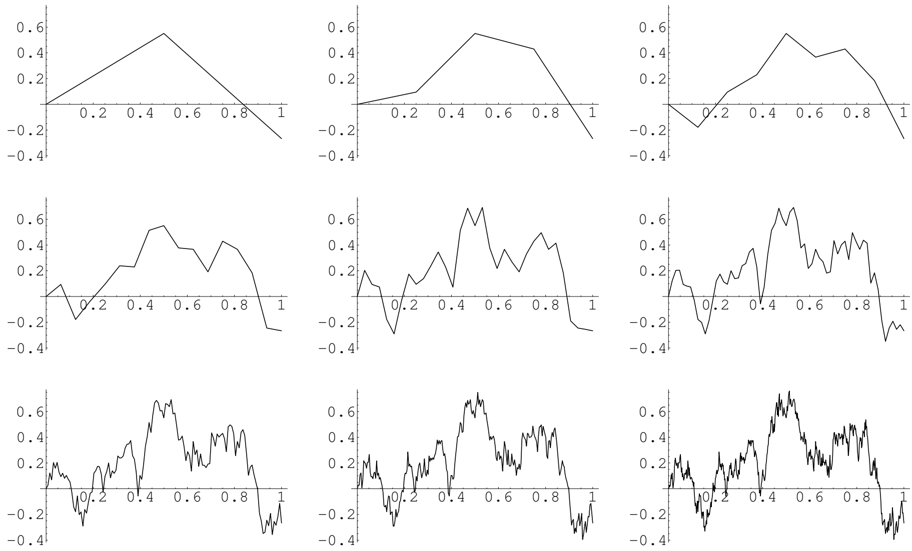

Figura 1. Aproximaciones al movimiento browniano

2. Existencia del movimiento browniano

2.1. El método de Lévy. Sea $(\Omega, \mathscr{F}, \mathbb{P})$ un espacio de probabilidad en el que están definidas una colección de variables aleatorias independientes e idénticamente distribuidas

$$(\xi_{i,n})_{0\leq i\leq 2^n, n\neq 0}$$

de distribución $N(0,1)$.

Definamos $X_0(0)=0$, $X_0(1)=\xi_{0,0}$ y extendamos linealmente la definición de X_0 al intervalo $[0,1]$. Definiremos una sucesión de procesos continuos con trayectorias continuas $(X_n)_{n\geq 0}$ postulando que X_n sea lineal sobre los intervalos de la forma $[k/2^n, (k+1)/2^n]$ y que

$$X_n\left(\frac{2j}{2^n}\right) = X_{n-1}\left(\frac{2j}{2^n}\right) \quad \text{y} \quad X_n\left(\frac{2j+1}{2^n}\right) = X_{n-1}\left(\frac{2j+1}{2^n}\right) + \frac{\xi_{2j+1,n}}{2^{(n+1)/2}}.$$

Se pueden apreciar los 9 primeros pasos de la aproximación en la Figure 1.

Notemos que para toda $n\geq 0$, $Y_n = (X_n(k/2^n))_{0\leq k\leq 2^n}$ es un vector aleatorio gaussiano, por lo que el proceso X_n es un proceso gaussiano ya que $X_n(t)$ es una combinación lineal de las entradas de Y^n. Para determinar a dicho proceso es suficiente explicitar su función de covarianza, que a su vez se obtiene por interpolación lineal en cada intervalo $[j/2^n, (j+1)/2^n]$ a partir de las cantidades:

$$\mathbb{E}\left(X_n\left(\frac{k}{2^n}\right)X_n\left(\frac{l}{2^n}\right)\right);$$

notemos que la función de media es cero.

Lema 5. *Se tiene la igualdad*

$$\mathbb{E}\left(X_n\left(\frac{k}{2^n}\right)X_n\left(\frac{l}{2^n}\right)\right)=\frac{k\wedge l}{2^n}$$

Demostración. La prueba se hará por inducción sobre n, siendo la base inductiva ($n=0$) inmediata. Si el lema es cierto para $n-1$ y k y l son pares, entonces también es válido para n. Por otra parte, si $k=2j+1$ y l es par entonces, al utilizar la independencia entre $\xi_{2j+1,n}$ y X_{n-1}, se obtiene

$$X_n\left(\frac{2j+1}{2^n}\right)=\frac{1}{2}X_{n-1}\left(\frac{j}{2^{n-1}}\right)+\frac{1}{2}X_{n-1}\left(\frac{j+1}{2^{n-1}}\right)+\frac{\xi_{2j+1,n}}{2^{(n+1)/2}},$$

por lo que

$$\begin{aligned}&\mathbb{E}\left(X_n\left(\frac{k}{2^n}\right)X_n\left(\frac{l}{2^n}\right)\right)\\&=\mathbb{E}\left(\frac{1}{2}X_{n-1}\left(\frac{j}{2^{n-1}}\right)X_{n-1}\left(\frac{l/2}{2^{n-1}}\right)\right)+\mathbb{E}\left(\frac{1}{2}X_{n-1}\left(\frac{j+1}{2^{n-1}}\right)X_{n-1}\left(\frac{l/2}{2^{n-1}}\right)\right)+0\\&=\frac{1}{2}\frac{j\wedge(l/2)}{2^{n-1}}+\frac{1}{2}\frac{(j+1)\wedge(l/2)}{2^{n-1}}.\end{aligned}$$

Al analizar los distintos casos que pueden darse, nos damos cuenta de que

$$\mathbb{E}\left(X_n\left(\frac{k}{2^n}\right)X_n\left(\frac{l}{2^n}\right)\right)=\frac{(2j+1)\wedge l}{2^n}.$$

por otra parte, si tanto l como k son impares pero distintos, el análisis es análogo. Finalmente, si $k=l=2j+1$, al escribir a $X_n(k/2^n)$ en términos de X_{n-1} y utilizar la hipótesis de inducción y la independencia entre $\xi_{k,n}$ y X_{n-1}, se observa que

$$\begin{aligned}\mathbb{E}\left(X_n\left(\frac{k}{2^n}\right)\right)&=\frac{1}{4}\frac{j}{2^{n-1}}+\frac{1}{4}\frac{j+1}{2^{n-1}}+2\frac{1}{4}\frac{j}{2^{n-1}}+\mathbb{E}\left(\left(\frac{\xi_{k,n}}{2^{(n+1)/2}}\right)^2\right)\\&=\frac{4j+1}{2^{n+1}}+\frac{1}{2^{n+1}}=\frac{2j+1}{2^n}.\end{aligned}$$

□

Verifiquemos ahora que la sucesión de procesos $(X_n)_{n\in\mathbb{N}}$ converge uniformemente. Para ésto, consideremos el evento

$$\begin{aligned}A_n &= \left\{\sup_{t\in[0,1]} |X_n(t) - X_{n-1}(t)| > 2^{-n/4}\right\}\\ &= \left\{\max_{0\leq j\leq 2^{n-1}-1} \left|X_n\left(\frac{2j+1}{2^n}\right) - X_{n-1}\left(\frac{2j+1}{2^n}\right)\right| > 2^{-n/4}\right\}\\ &= \left\{\max_{0\leq j\leq 2^{n-1}-1} |\xi_{2j+1,n}| > 2^{(n+2)/4}\right\}.\end{aligned}$$

Entonces, por la subaditividad de $\mathbb{P}$, el hecho de que las variables $\xi_{i,j}$ tengan ditribución $N(0,1)$ y la cota para la cola de la distribución normal estandar:

$$\begin{aligned}\mathbb{P}(A_n) &\leq \sum_{j=0}^{2^{n-1}-1} \mathbb{P}\Big(|\xi_{1,1}| > 2^{(n+2)/4}\Big)\\ &\leq 2^{n-1} \times 2 \times e^{-2^{(n+2)/2}/2}.\end{aligned}$$

De la cota anterior, se conluye la convergencia de la serie $\sum_i \mathbb{P}(A_i)$, por lo cual, el lema de Borel-Cantelli nos permite afirmar que existe $E \in \mathscr{F}$ tal que $\mathbb{P}(E) = 1$ tal que si $\omega \in E$, existe $n_0 = n_0(\omega)$ tal que para $n \geq n_0$ se tiene que

$$|X_n(t) - X_{n-1}(t)| \leq 2^{-n/4},$$

de lo cual se deduce la convergencia uniforme de la sucesión $(X_n)_{n\in\mathbb{N}}$ hacia un límite $X = (X_t)_{t\in[0,1]}$ que es entonces continuo. La prueba estará (basicamente) terminada cuando verifiquemos que X es un movimiento browniano en $[0,1]$; sin embargo, ésto se sigue del hecho de que una sucesión de variables aleatorias gaussianas que converge en probabilidad (lo cual está implicado por la convergencia casi segura) también converge en L^p para toda $p \geq 1$, se puede tomar el límite cuando $n \to \infty$ con $k = \lfloor 2^n s\rfloor/2^n$ y $l = \lfloor 2^n t\rfloor/2^n$ en el lema anterior para concluir que la función de media de X es cero y que

$$\mathbb{E}(X_t X_s) = t \wedge s.$$

Si $s_1 \leq s_2 \leq t_1 \leq t_2$, al igualdad anterior implica que

$$\mathbb{E}((X_{s_2} - X_{s_1})(X_{t_2} - X_{t_1})) = s_2 - s_1 - s_2 + s_1 = 0,$$

por lo que X tiene incrementos independientes (recordemos que se trata de un proceso gaussiano) y como tiene trayectorias continuas, empieza en cero y X_t tiene distribución $N(0,1)$, se sigue que X es un movimiento browniano.

Para concluir, falta construir un movimiento browniano en $[0,\infty)$ en vez de en $[0,1]$, pero ésto se puede lograr considerando una sucesión de movimientos brownianos independientes en $[0,1]$ y concatenando sus trayectorias.

Ejercicio 6.2. El objetivo de este ejercicio es construir, a partir de movimientos brownianos en $[0, 1]$, al movimiento browniano en $[0, \infty)$.

(1) Pruebe que existe un espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ en el que existe una sucesión $B^1, B^2, \ldots$ de movimientos brownianos en $[0, 1]$ independientes. (Sugerencia: utilice la construcción del movimiento browniano de Lévy para que la solución sea corta.)

(2) Defina a $B_t = B^1_1 + \cdots + B^{\lfloor t \rfloor}_1 + B^{\lceil t \rceil}_{t - \lfloor t \rfloor}$ para $t \geq 0$. Pruebe que B es un movimiento browniano.

2.2. El método de Kolmogorov. Puesto que hemos probado el teorema de consistencia de Kolmogorov en tiempo continuo, esbozaremos uno de los métodos más generales para construir procesos con trayectorias continuas, basados en el criterio de continuidad de Kolmogorov. Si $0 \leq t_1 < \cdots < t_n$, definamos $\mu_{t_1, \ldots, t_n}$ como la distribución normal multivariada de media cero y matriz de varianza-covarianza Σ dada por $\Sigma_{i,j} = t_j \wedge t_j$. Al quitar la coordenada i a un vector aleatorio con distribución $\mu_{t_1, \ldots, t_n}$ obtenemos un vector aleatorio cuya distribución es $\mu_{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n}$, por lo cual se puede aplicar el teorema de consistencia de Kolmogorov para concluir que existe un espacio de probabilidad en el que están definido un proceso estocástico $(B_t, t \geq 0)$ tal que la distribución de $(B_{t_1}, \ldots, B_{t_n})$ es $\mu_{t_1, \ldots, t_n}$. Es claro que entonces B es un movimiento browniano en ley.

Teorema 6.1 (Criterio de continuidad de Kolmogorov). *Sea X un proceso estocástico real tal que existen constantes $\alpha, \beta, K \geq 0$ tales que*

$$\mathbb{E}(|X_t - X_s|^{\alpha}) \leq K (t - s)^{1+\beta}.$$

Entonces existe un proceso estocástico $\tilde{X}$ tal que X es modificación de $\tilde{X}$, es decir que $\mathbb{P}\Big(X_t = \tilde{X}_t\Big) = 1$ para toda $t \geq 0$, y cuyas trayectorias son continuas.

Puesto que

$$\mathbb{E}\big(|B_t - B_t|^{4}\big) = 3 (t - s)^2$$

si $0 \leq s \leq t$, vemos que el criterio de Kolmogorov aplica para construir una modificación de B con trayectorias continuas.

Ejercicio 6.3. Pruebe que si $\tilde{X}$ es una modificación de X entonces ambos procesos tienen las mismas distribuciones finito-dimensionales. Concluya que si B es un movimiento browniano en ley y $\tilde{B}$ es una modificación de B con trayectorias continuas entonces $\tilde{B}$ es un movimiento browniano.

Demostración. De nuevo razonamos en $[0, 1]$. Sea D_n el conjunto de los racionales diádicos de orden n, es decir

$$D_n = \left\{ \frac{k}{2^n} : 0 \leq k \leq 2^n \right\}.$$

Definamos a

$$\overline{\xi}_n = \max_{0 \leq k \leq 2^n - 1} \left| X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}} \right|.$$

Al utilizar nuestra hipótesis, vemos que

$$\mathbb{E}\left(\sum_n \left(2^{nc}\overline{\xi}_n\right)^\alpha \right) \leq \sum_n 2^{cn\alpha} 2^n K 2^{-n(1+\beta)} = K \sum_n 2^{n(c\alpha-\beta)}.$$

Por lo tanto, si $c < \alpha/\beta$ entonces la esperanza del lado izquierdo es finita y por lo tanto casi seguramente existe una variable aleatoria K_2 tal que

$$\xi_n \leq K_2 2^{-nc} \text{ para toda } n \geq 1.$$

Veamos que entonces X tiene trayectorias c-Hölder, al menos en el conjunto

$$D = \bigcup_n D_n.$$

En efecto, si $\delta > 0$, consideremos m tal que $\delta \in [2^{-m-1}, 2^{-}m]$. Si $s, t \in D$ y $|t - s| \leq \delta$ entonces existe $n \geq m$ tal que $t, s \in D_n$, por lo cual

$$|X_t - X_s| \leq \overline{\xi}_n \leq \sum_{n' \geq m} \overline{\xi}_{n'} \leq K_2 \sum_{n' \geq m} 2^{-cn'} \leq K_2 2^{-cm} \leq \frac{K_2}{2^c} r^c.$$

Al ser X c-Hölder continua sobre D y ser D denso, existe una única extensión c-Hölder $\tilde{X}$ de X sobre D a $[0,1]$ que satisface

$$\tilde{X}_t = \liminf_{n\to\infty} X_{\lceil 2^n t \rceil / 2^n} \text{ casi seguramente.}$$

Veamos que $\tilde{X}$ es una modificación de X. Esto se sigue pues si $t \in [0,1]$ y $t_n = \lceil 2^n t \rceil / 2^n$ entonces

$$\mathbb{E}\left(\left| \tilde{X}_t - X_t \right|^{\alpha} \right) \leq \liminf_{n\to\infty} \mathbb{E}(|X_{t_n} - X_t|^{\alpha}) \leq \limsup_{n\to\infty} K\left(t_n - t\right)^{1+\beta} = 0. \qquad \square$$

3. La propiedad de Markov

La propiedad de homogeneidad temporal del movimiento browniano (que también comparte con el proceso de Poisson y otros procesos de Lévy) se puede interpretar también como una propiedad de Markov. Sean $b\mathscr{B}_{\mathbb{R}}$ el conjunto de funciones medibles de $\mathbb{R}$ en $\mathbb{R}$ que son medibles y acotadas y C_0 el subconjunto de $b\mathscr{B}_{\mathbb{R}}$ que consta de funciones continuas que se anulan al infinito, es decir, tales que $\lim_{x\to\pm\infty} f(x) = 0$.

Definiremos $P_t : b\mathscr{B}_{\mathbb{R}} \to b\mathscr{B}_{\mathbb{R}}$ dado por

$$P_t f(x) = \mathbb{E}(f(B_t + x)) = \int f(y) \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} \, dy.$$

Nuestra primera versión de la propiedad de Markov es entonces

Proposición 6.2. *Para toda $f \in b\mathscr{B}_{\mathbb{R}}$*

$$\mathbb{E}(\, f(B_t) \mid \mathscr{F}_s) = P_{t-s} f(B_s)\,.$$

Además $P_t(C_0) \subset C_0$ y P tiene la propiedad de Feller:

$$\lim_{t\to 0} \sup_{f\in C_0, \|f\|\leq 1} \|P_t f - f\| = 0,$$

donde $\|\cdot\|$ denota la norma uniforme.

Por supuesto, la propiedad de Markov es mucho más útil cuando la expresamos en el espacio canónico.

Sea Ω el espacio de funciones continuas de $[0,\infty)$ en $\mathbb{R}$ y definamos $X_t : \Omega \to \mathbb{R}$ mediante $X_t(f) = f(t)$, así como la σ-álgebra $\mathscr{F} = \sigma(X_t : t \geq 0)$. Si B es un movimiento browniano definido en cualquier espacio de probabilidad, podemos interpretarlo como una variable aleatoria con valores en Ω (que será medible respecto de la σ-álgebra $\mathscr{F}$ que hemos escogido) y denotamos por $\mathbb{W}_0$ a su ley. A $\mathbb{W}_0$ se le conoce como medida de Wiener. De la misma manera definimos a $\mathbb{W}_x$ como la distribución de $B + x$.

Sea $\theta_t : \Omega \to \Omega$ donde $\theta_t f(s) = f(t+s)$. Ahora daremos la versión de la propiedad de Markov fuerte del movimiento browniano en el espacio canónico.

Proposición 6.3. *Para toda $F \in b\mathscr{F}$, la función $x \mapsto \mathbb{W}_x(F)$ pertenece a $b\mathscr{B}_{\mathbb{R}}$ y*

$$\mathbb{W}_x(\, F \circ \theta_t \mid \mathscr{F}_t) = \mathbb{W}_{X_t}(F)\,.$$

De igual manera, la homogeneidad se puede extender a tiempos de paro e interpretar como una propiedad de Markov fuerte

Proposición 6.4 (Propiedad de Markov fuerte para el movimiento browniano). *Sea B un movimiento browniano y $\mathscr{F}_t, t \geq 0$ su filtración canónica. Si T es un tiempo de paro finito, el proceso B^T dado por $B^T_s = B_{T+s} - B_T$ es un movimiento browniano independiente de $\mathscr{F}_T$. En el espacio canónico, se tiene que para toda $F \in b\mathscr{F}$ y todo tiempo de paro T*

$$\mathbb{W}_x(\, F \circ \theta_T \mid \mathscr{F}_T) = \mathbb{W}_{X_T}(F)\,.$$

Hemos probado más generalmente la propiedad de Markov fuerte para procesos de Lévy. La versión en el espacio canónico admite una verificación similar.

Como una aplicación a la propiedad de Markov fuerte, daremos el principio de invarianza de Donsker, primero en su versión para la caminata aleatoria simple.

4. Martingalas y procesos asociados

Continuaremos con algunos procesos asociados al Browniano que resultan ser útiles para su análisis. Comenzaremos con algunas martingalas.

Proposición 6.5. *Sea B un movimiento browniano. Entonces los siguientes procesos son martingalas.*

(1) $B_t, t \geq 0$,
(2) $B_t^2 - t, t \geq 0$,
(3) $e^{\lambda B_t - \lambda^2 t/2}$ y
(4) $\cosh(\lambda B_t)\, e^{-\lambda^2 t/2}$.

Demostración. Se tiene que $B_t - B_s$ es independiente de $\mathscr{F}_s$ para $s \leq t$; se deduce lo anterior pues por una parte $B_t - B_s$ es independiente de $\left(B_{s_i} - B_{s_{i-1}}\right)_{i=0}^{n}$ para cualquier $n \geq 0$ y cualquier colección de reales

$$0 = s_0 \leq s_1 \leq \cdots \leq s_n \leq s.$$

y por otra, dichas variables aleatorias generan $\mathscr{F}_s$.

(1) Vemos que

$$0 = \mathbb{E}(\, B_t - B_s \mid \mathscr{F}_s) = \mathbb{E}(\, B_t \mid \mathscr{F}_s) - B_s,$$

pues B_s es $\mathscr{F}_s$ medible. Se conlcuye que B es una $(\mathscr{F}_t)_{t \geq 0}$-martingala.

(2) Al ser B una martingala y $B_t - B_s$ independiente de $\mathscr{F}_s$, se tiene que

$$\begin{aligned} t - s &= \mathbb{E}(B_t - B_s) \\ &= \mathbb{E}\Big(\, (B_t - B_s)^2 \,\Big|\, \mathscr{F}_t\Big) \\ &= \mathbb{E}\big(B_t^2 \,\big|\, \mathscr{F}_t\big) - 2\mathbb{E}(\, B_t B_s \mid \mathscr{F}_t) + B_s^2 \\ &= B_s^2 - \mathbb{E}\big(B_t^2 \,\big|\, \mathscr{F}_s\big), \end{aligned}$$

de acuerdo a las propiedades de la esperanza condicional.

(3) Basta recordar que el cálculo de la transformada de Laplace de una variable normal estandar y utilizar el que $B_t - B_s$ es independiente de $\mathscr{F}_s$ para $s \leq t$ y se distribuye $N(0, t-s,)$ pues entonces:

$$e^{\lambda^2 (t-s)/2} = \mathbb{E}\Big(e^{\lambda(B_t - B_s)}\Big) = \mathbb{E}\Big(\, e^{\lambda(B_t - B_s)} \,\Big|\, \mathscr{F}_t\Big) = \mathbb{E}\big(e^{\lambda B_t} \,\big|\, \mathscr{F}_t\big)\, e^{-\lambda B_s}. \qquad \square$$

Ejercicio 6.4. Sea

$$M_t^\lambda = e^{\lambda B_t - \lambda^2 t/2}.$$

(1) Explique y pruebe formalmente por qué, para toda $n \geq 1$, $\partial^n M_t^\lambda / \partial \lambda^n$ es una martingala.
(2) Sea $H_n(x) = (-1)^n\, e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$. A H_n se le conoce como enésimo polinomio de Hermite. Calcúlelo para $n \leq 5$. Pruebe que H_n es un polinomio para toda $n \in \mathbb{N}$ y que $\partial^n M_t^\lambda / \partial \lambda^n = t^{n/2} H_n\big(B_t/\sqrt{t}\big)\, M_t^\lambda$.
(3) Pruebe que $t^{n/2} H_n\big(B_t/\sqrt{t}\big)$ es una martingala para toda n y calcúlela para $n \leq 5$.
(4) Aplique muestreo opcional a las martingalas anteriores al tiempo aleatorio $T_{a,b} = \min\{t \geq 0 : B_t \in \{-a, b\}\}$ (para $a, b > 0$) con $n = 1, 2$ para calcular $\mathbb{P}\big(B_{T_{a,b}} = b\big)$ y $\mathbb{E}(T_{a,b})$, Qué concluye cuando $n = 3, 4$? ¿ Cree que $T_{a,b}$ tenga momentos finitos de cualquier orden? Justifique su respuesta.

(5) Aplique el teorema de muestreo opcional a la martingala M^λ al tiempo aleatorio $T_a = \inf\{t \geq 0 : B_t \geq a\}$ si $\lambda > 0$. Diga por qué es necesaria la última hipótesis y calcule la transformada de Laplace de T_a.
(6) Opcional (para subir calificación en esta u otra tarea):
   (a) Modifique el ejercicio para que aplique al proceso Poisson.
   (b) Resuélva el ejercicio modificado.

Contruyamos ahora una martingala a dos parámetros con el movimiento browniano: consideremos

$$M_{t,s} = B_t - B_s$$

para $0 \leq s < t$ y $\mathscr{F}_{s,t} = \sigma(B_u - B_s : u \in [s,t])$. Entonces, como $\mathscr{F}_{s,t}$ es independiente de $\mathscr{F}_s$ (por la propiedad de incrementos independientes de B) y está contenida en $\mathscr{F}_t$, si $0 \leq u \leq s < t \leq v$, se tiene que

$$\begin{aligned}\mathbb{E}(\, M_{u,v} \mid \mathscr{F}_{s,t}) &= \mathbb{E}(\, B_v - B_u \mid \mathscr{F}_{s,t}) \\ &= \mathbb{E}(\, B_v - B_t \mid \mathscr{F}_{s,t}) + \mathbb{E}(\, B_s - B_u \mid \mathscr{F}_{s,t}) + B_t - B_s \\ &= B_t - B_s = M_{t,s}.\end{aligned}$$

Ahora analizaremos cuatro procesos importantes que ilustran propiedades de invariancia de la distribución del movimiento browniano.

Proposición 6.6. *El movimiento browniano B tiene las siguientes propiedades de invariancia.*

Simetría: *$-B$ es un movimiento browniano*
Homogeneidad temporal: *Para toda $t \geq 0$ el proceso B^t dado por $B^t_s = B_{t+s} - B_t$ es un movimiento browniano independiente de $\sigma(B_s : s \leq t)$.*
Autosimilitud: *Para toda $c > 0$ el proceso $B_{ct}/\sqrt{c}, t \geq 0$ es un movimiento browniano.*
Inversión temporal: *El proceso*

$$X_t = \begin{cases} 0 & t = 0 \\ tB_{1/t} & t > 0 \end{cases},$$

para $t \geq 0$, es un movimiento browniano.

Demostración.

(1) Los incrementos de $-B$ son iguales a menos los incrementos de B. Por lo tanto, los primeros serán independientes y estacionarios. Las trayectorias de $-B$ son continuas y comienzan en cero. Finalmente, puesto que la distribución normal centrada es invariante ante la transformación $x \mapsto -x$, vemos que $-B_t$ y B_t tienen la misma distribución y por lo tanto $-B$ es un movimiento browniano.

(2) Notemos que las trayectorias de B^t son continuas y comienzan en cero. Si $0 = s_0 < s_1 < \cdots < s_n$, entonces

$$\left(B^t_{s_1} - B^t_{s_0}, \ldots, B^t_{s_n} - B^t_{s_{n-1}}\right) = \left(B_{t+s_1} - B_t, \ldots, B_{t+s_n} - B_{t+s_{n-1}}\right);$$

puesto que los incrementos de B son independientes y estacionarios, vemos que los de B^t también lo son. Además, ya que $B^t_s = B_{t+s} - B_t$, vemos B^t_s tiene distribución normal $(0, s)$. Finalmente, para verificar que B^t es independiente de $\mathscr{F}_t$, notemos que por la propiedad de incrementos independientes de B, $B^t_{t_1}, \ldots, B^t_{t_n}$ es independiente de $(B_{s_1}, \ldots, B_{s_n})$ si $s_1, \ldots, s_n \leq t$. Por clases monótonas, se verifica entonces que B^t es independiente de $\mathscr{F}_t$.

(3) Se omite. Buen ejercicio

(4) Puesto el proceso de interés es gaussiano, se verifica mediante un cálculo de varianzas-covarianzas, que $(X_{t_1}, \ldots, X_{t_n})$ y $(B_{t_1}, \ldots, B_{t_n})$ si $t_1, \ldots, t_n \geq 0$. Por lo tanto, X es un movimiento browniano en ley. Sin embargo, no es nada trivial es que el proceso de interés tiene trayectorias continuas, en particular en cero. Ofrecemos dos pruebas: la primera es notar que B satisface la ley fuerte de los grandes números: $B_t/t \to 0$ conforme $t \to \infty$ casi seguramente.

Ejercicio 6.5.

(a) Al aplicar la desigualdad maximal de Doob sobre los racionales de orden n y pasar al límite conforme $n \to \infty$, pruebe que $\sup_{t\leq} |B_t - B_1|$ es cuadrado integrable.

(b) Pruebe que la sucesión de variables aleatorias

$$\left(\sup_{t\in[0,1]} |B_{n+t} - B_n|, n \in \mathbb{N}\right)$$

son independientes, idénticamente distribuidas y de media finita. (Utilice la propiedad de Markov.)

(c) Al utilizar Borel-Cantelli, pruebe que, para cualquier $C > 0$ fija, $\limsup_{n\to\infty} \sup_{t\in[0,1]} |B_{n+t} - B_n| / n \leq C$ casi seguramente.

(d) Pruebe que $(B_n/n, n \geq 1)$ converge casi seguramente a 0 y deduzca que $\lim_{t\to\infty} B_t/t = 0$.

La segunda prueba comienza con notar que B y X tienen las mismas distribuciones finito-dimensionales y trayectorias continuas en $(0, \infty)$. Luego, si $s^k_1, s^k_2, \ldots$ es una enumeración de los racionales en $[0, 1/k]$ para $k \geq 1$,se escribe

$$\left\{\lim_{t\to 0} B_t = 0\right\} = \bigcap_{n\geq 1} \bigcup_{k\geq 1} \bigcap_i \left\{\left|B_{s^k_i}\right| < 1/n\right\},$$

y se tiene una expresión similar para $\{\lim_{t\to 0} X_t = 0\}$. Por continuidad de $\mathbb{P}$, se sigue entonces que

$$\begin{aligned}
1 &= \mathbb{P}\Big(\lim_{t\to 0} B_t = 0\Big) \\
&= \lim_{n\to\infty} \lim_{k\to\infty} \lim_{i\to\infty} \mathbb{P}\Big(\Big|B_{s_1^k}\Big| < 1/n, \ldots, \Big|B_{s_i^k}\Big| < 1/n\Big) \\
&= \lim_{n\to\infty} \lim_{k\to\infty} \lim_{i\to\infty} \mathbb{P}\Big(\Big|X_{s_1^k}\Big| < 1/n, \ldots, \Big|X_{s_i^k}\Big| < 1/n\Big) \\
&= \mathbb{P}\Big(\lim_{t\to 0} X_t = 0\Big).
\end{aligned}$$

(Esta prueba es mucho más general si se expresa en el espacio canónico.)

□

Finalmente, estudiaremos algunos otros procesos importantes que se definen a partir del movimiento browniano.

Proceso de calor: Es el proceso estocástico $(t, B_t)\,, t \geq 0$.

Movimiento browniano multidimensional: Si $d \geq 1$, sean $B^1, \ldots, B^d$ d movimientos brownianos independientes. Entonces $B = \left(B^1, \ldots, B^d\right)$ es el llamado movimiento browniano en dimensión d.

Procesos de Bessel de dimensión entera: Si B es un movimiento browniano d dimensional, el proceso R dado por $R_t = \|B_t\|$ es el llamado proceso de Bessel d-dimensional.

Máximo acumulativo: Si B es un movimiento browniano unidimensional, su máximo acumulativo es el proceso $\overline{B}$ dado por $\overline{B}_t = \max_{s\leq t} B_s$. Es un proceso adaptado respecto a la filtración canónica de B y tiene trayectorias continuas.

Proceso de tiempos de arribo: Se trata del inverso generalizado del máximo acumulativo; formalmente se trata del proceso T dado por

$$T_a = \inf\{t \geq 0 : B_t \geq a\}$$

para $a \geq 0$. Es un proceso con trayectorias no decrecientes y continuas por la derecha. De hecho, veremos que es un subordinador estable. Al utilizar la martingala exponencial del movimiento browniano es fácil calcular su transformada de Laplace. (Tarea.)

Proceso de tiempo de positividad: Sea $A_t = \int_0^t \mathbf{1}_{B_s > 0}\, ds$. Entonces A_t es una variable aleatoria, lo cual se prueba al notar que la función $(t, \omega) \mapsto B_t(\omega)$ es medible en el espacio producto, lo que es consecuencia de que las trayectorias de B sean continuas. Entonces se puede aplicar Tonelli para concluir que A_t es variable aleatoria. Con esto, se deduce que A es un proceso con trayectorias continuas. Podrémos calcular explícitamente la distribución de A_t.

Proceso de edad de las excursiones: Sea

$$g_t = \sup\{s \le t : B_s = 0\}.$$

Puesto que para cada $t \ge 0$, $B_t \ne 0$ casi seguramente, entonces casi seguramente $g_t < t$.

5. Principio de invariancia de Donsker

El espacio canónico de funciones continuas de $[0,1]$ en $\mathbb{R}$, denotado $C = C([0,1])$ se puede dotar de la métrica d inducida por la norma uniforme. Por ejemplo, si X es el proceso canónico entonces X_t es continua en C. Esta métrica nos ayuda a definir funciones continuas en C y probar resultados de convergencia débil para variables aleatorias con valores en C.

Sea (S_n) una caminata aleatoria centrada y con varianza finita y positiva, que supondremos, sin pérdida de generalidad, igual a 1. Sea

$$X_t^n = \frac{1}{\sqrt{n}}\left[S_{\lfloor nt\rfloor}\left(\lceil nt\rceil - nt\right) + S_{\lceil nt\rceil}\left(nt - \lfloor nt\rfloor\right)\right].$$

Teorema 6.2 (Principio de invariancia de Donsker). *Para toda $F : C \to \mathbb{R}$ continua y acotada, $\mathbb{E}(F(X^n)) \to \mathbb{E}(F(B))$.*

Así por ejemplo, al considerar $F(f) = \sup_{s\le t} f(s) \wedge x$ vemos que la distribución de $\max_{k\le n} S_k/\sqrt{n}$ converge débilmente a la distribución de $\max_{s\in[0,1]} B_s$. Pero también, con $X_t \wedge x$, vemos que $S_n/\sqrt{n}$ converge débilmente a B_1 que tiene distribución normal $[0,1]$. En este sentido, el teorema anterior es una extensión funcional del teorema límite central.

La prueba del teorema anterior se basará en un resultado que permite encajar a la caminata aleatoria dentro del movimiento browniano.

Teorema 6.3 (Teorema de encaje de Skorohod). *Sea (S_n) una caminata aleatoria centrada y con varianza 1. Existe un espacio de probabilidad filtrado $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$ en el que existe un $(\mathscr{F}_t)$-movimiento browniano B y una sucesión creciente de tiempos de paro $(T_n, n \in \mathbb{N})$ tal que $((T_n, B_{T_n}), n \in \mathbb{N})$ es una caminata aleatoria, donde las primeras coordenadas tienen esperanza finita igual a 1 y la segundas coordenadas tienen la misma distribución que S.*

Un ejemplo sencillo de este teorema, que probaremos más adelante, consiste en considerar $a, b > 0$ y considerar el caso en que S_1 tiene distribución concentrada en $\{a, b\}$ y centrada por lo que

$$\mathbb{P}(S_1 = b) = \frac{a}{a+b} \quad \text{y} \quad \mathbb{P}(S_1 = -a) = \frac{b}{a+b}.$$

Si B es un movimiento browniano, definamos recursivamente $T_0 = 0$ y

$$T_{n+1} = \inf\{t \ge 0 : B_{t+T_n} - B_{T_n} \in \{-a, b\}\}.$$

Por la propiedad de Markov fuerte vemos que las variables aleatorias

$$\left(\left(T_n - T_{n-1}, B_{T_n} - B_{T_{n-1}}\right), n \in \mathbb{N}\right)$$

son independientes e idénticamente distribuidas. Además, al utilizar el teorema de muestreo opcional de Doob aplicado a las martingalas B y $B^2 - \mathrm{Id}$ al tiempo de paro T_1, vemos que B_{T_1} tiene la misma distribución que S_1 y que T_1 tiene esperanza finita e igual a ab.

Principio de invariancia de Donsker. Se asumirá el teorema de encaje de Skorohod.

Consideraremos el caso en que F es uniformemente continua. (Esto es suficiente por argumentos generales de convergencia débil.) Sea M una cota para F y para $\varepsilon > 0$, sea $\delta = \delta(\varepsilon) > 0$ tal que $|F(f) - F(g)| < \varepsilon$ si $\|f - g\| < \delta$. Notemos que

$$\lim_{\gamma \to 0} \mathbb{P}\left(\sup_{s \leq 1, h \leq \gamma} |B_s - B_{s+h}| > \delta\right) = 0$$

puesto que B tiene trayectorias continuas.

Puesto que $\mathbb{E}(T_1)$ =1, la ley fuerte de los grandes números implica que $T_n/n \to 1$ casi seguramente y por lo tanto, el lema de Pólya para convergencia uniforme implica la convergencia $T_{\lfloor nt \rfloor}/n \to t$ uniformemente para $t \in [0, 1]$ (casi seguramente).

Por el teorema de encaje de Skorohod, a la caminata aleatoria reescalada e interpolada la podemos escribir como sigue:

$$X_t^n = \frac{B_{T_{\lfloor nt \rfloor}}}{\sqrt{n}} \left(\lceil nt \rceil - nt\right) + \frac{B_{T_{\lceil nt \rceil}}}{\sqrt{n}} \left(nt - \lfloor nt \rfloor\right)$$

Por escalamiento, vemos que

$$|\mathbb{E}(F(X^n) - F(B))| = |\mathbb{E}\big(F(X^n) - F\big(B_{nt}/\sqrt{n}, t \leq 1\big)\big)|.$$

Ahora dividimos la integral en dos partes, dependiendo de si

$$\sup_{t \leq 1} \left|T_{\lfloor nt \rfloor}/n - t\right| \begin{cases} < \gamma & \text{ó} \\ \geq \gamma \end{cases}$$

para obtener

$$\begin{aligned} &|\mathbb{E}\big(F(X^n) - F\big(B_{nt}/\sqrt{n}, t \leq 1\big)\big)| \\ &\leq |\mathbb{E}\big(F(X^n) - F\big(B_{nt}/\sqrt{n}, t \leq 1\big)\big) \mathbf{1}_{\sup_{t \leq 1} \left|T_{\lfloor nt \rfloor}/n - t\right| \leq \gamma}| \\ &+ 2M\mathbb{P}\left(\sup_{t \leq 1} \left|T_{\lfloor nt \rfloor}/n - t\right| \geq \gamma\right). \end{aligned}$$

El segundo sumando tiende a cero conforme $n \to \infty$. Por otra parte

$$\begin{aligned}&\left|\mathbb{E}\Big(\big[F(X^n) - F\big(B_{nt}/\sqrt{n}, t \leq 1\big)\big]\,\mathbf{1}_{\sup_{t\leq 1}\left|T_{\lfloor nt\rfloor}/n - t\right| \leq \gamma}\Big)\right| \\ &\leq \varepsilon + 2M\mathbb{P}\left(\sup_{\substack{h\leq 2\gamma n\\ t\leq n}} |B_t - B_{t+h}| > \delta\right).\end{aligned}$$

Por autosimilitud podemos escribir al segundo sumando como

$$2M\mathbb{P}\left(\sup_{\substack{h\leq 2\gamma\\ t\leq 1}} |B_t - B_{t+h}| > \delta\right).$$

Ahora es claro que podemos escoger $\gamma > 0$ tal que el segundo sea tan pequeño como queramos. $\square$

La prueba del teorema de encaje de Skorohod se basa a su vez en el hecho de que cualquier distribución en $\mathbb{R}$ centrada se puede ver como mezcla de las distribuciones

$$\nu_{a,b} = \frac{a}{a+b}\delta_b + \frac{b}{a+b}\delta_{-a}.$$

Lema 6. *Sea μ una distribución en $\mathbb{R}$ centrada y de varianza finita. Entonces existe una distribución $\tilde{\mu}$ en $\mathbb{R}_- \times \mathbb{R}_+$ tal que*

$$\mu(A) = \int \nu_{a,b}(A)\ \tilde{\mu}(da, db).$$

Demostración. Sean μ_+ y μ_- las restricciones de μ a $(-\infty, 0)$ y a $(0, \infty)$ respectivamente, definamos

$$c = \int y\,\mu_+(dy) = -\int x\,\mu_-(dx),$$

y sea

$$\tilde{\mu}(dx, dy) = \mu(0)\,\delta_{0,0}(dx, dy) + \frac{y-x}{c}\mu_-(dx)\,\mu_+(dy).$$

Entonces, para cualquier f medible y acotada (por no decir funciones indicadoras) se tiene que

$$\begin{aligned}c\int f(x)\ \mu(dx) &= cf(0)\,\mu(0) + c\int f(y)\ \mu_-(dy) + c\int f(x)\ \mu_+(dx) \\ &= cf(0)\,\mu(0) + \int\int yf(x) - xf(y)\ \mu_+(dy)\,\mu_-(dx).\end{aligned}$$

Como

$$\nu_{x,y}(f) = \frac{f(x)\,y - f(y)\,x}{y-x},$$

se sigue que

$$c\int f\,\mu(dx) = cf(0)\,\mu(0) + \int\int \nu_{x,y}(f)\,(y-x)\;\mu_-(dx)\,\mu_+(dy)$$

o equivalentemente,

$$\int f\,\mu(dx) = \int\int \nu_{x,y}(f)\;\tilde{\mu}(dx,dy)\,. \qquad \square$$

Prueba del Teorema de encaje de Skorohod. Veamos ahora cómo encajar una caminata aleatoria centrada (con varianza 1 por cada paso) en el movimiento browniano B: sea μ una medida de probabilidad centrada en $\mathbb{R}$ y (α_n, β_n), $n \geq 1$ una sucesión de vectores aleatorios intependientes en $\mathbb{R}^2$ con distribución $\tilde{\mu}$. Asumimos que dicha sucesión es independiente de B. Consideremos a la sucesión de tiempos de paro

$$\tau_0 = 0 \quad \text{y} \quad \tau_{n+1} = \min\left\{t \geq \tau_n : |B_t - B_{\tau_n}| \in \{\alpha_{n=1}, \beta_{n+1}\}\right\}.$$

Sabemos que $\tau_n < \infty$ con probabilidad 1. Además,

$$\mathbb{E}(f(B_{\tau_1})) = \mathbb{E}(\mathbb{E}(\,f(B_{\tau_1})\mid \alpha_1, \beta_1)) = \mathbb{E}(\nu_{\alpha_1,\beta_1}(f)) = \int \nu_{x,y}(f)\;\tilde{\mu}(x,y)\,.$$

Se sigue que B_{τ_1} tiene distribución ν. Además, puesto que

$$\mathbb{E}(\,\tau_1 \mid \alpha_1, \beta_1) = \alpha_1\beta_1 = \int z^2\,\nu_{\alpha_1,\beta_1}(dz)$$

como vimos en la sección anterior, se sigue que

$$\mathbb{E}(\tau_1) = \int x^2 \mu(dx) = 1.$$

Finalmente, por la propiedad de Markov fuerte, vemos que las variables aleatorias $(\tau_{n+1} - \tau_n)$ son independientes e idénticamente distribuidas, al igual que las variables $B_{\tau_{n+1}} - B_{\tau_n}$ y que estás últimas tienen distribución μ. $\square$

6. El puente browniano

Ahora se introducirá un ejemplo interesante de proceso gausianno relacionado con el movimiento browniano. Si B es un movimiento browniano, notemos que $b_t = B_t - tB_1$ es independiente de B_1. Para ver esto es suficiente probar que la covariación entre b_t y B_1 es cero:

$$\mathbb{E}(b_t B_1) = \mathbb{E}(B_t B_1) - t\mathbb{E}\big(B_1^2\big) = t - t = 0.$$

Por lo tanto, si F es $\sigma(X_s : s \in [0,1])$-medible, se tiene que

$$\mathbb{E}(F(B)\,f(B_1)) = \mathbb{E}(F(b_t + tB_1)\,f(B_1)) = \int \mathbb{E}(F(b_t + tx))\,\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx.$$

Vemos que si $\mathbb{B}_x$ es la distribución de $b_t + tx, t \in [0,1]$ entonces $(\mathbb{B}_x)$ es una distribución condicional regular para B en $[0,1]$ dado el valor de B_1. La regularidad se sigue pues

$$x \mapsto \mathbb{P}(b_{t_1} + t_1 x \leq x_1, \ldots, b_{t_n} + t_n x \leq x_n)$$

es medible y de utilizar el lema de clases de Dynkin.

7. Algunos cálculos distribucionales

En esta sección utilizaremos la propiedad de Markov y de Markov fuerte del movimiento browniano para calcular ciertos aspectos distribucionales de este proceso.

Sea B un movimiento browniano, $\overline{B}$ su proceso de máximo acumulativo y T su proceso de tiempos de arribo.

PROPOSICIÓN 6.7. *El proceso T es un subordinador autosimilar. Además,*

$$\mathbb{E}\left(e^{-\lambda T_a}\right) = e^{-a\sqrt{2\lambda}} = e^{-a\int_0^\infty \left(1-e^{-\lambda x}\right)\nu(dx)}$$

donde

$$\nu(dx) = \frac{1}{\sqrt{2\pi x^3}}.$$

Finalmente, para cada $a > 0$, T_a tiene la misma distribución que $a/\overline{B}_1^2$.

DEMOSTRACIÓN. Comenzamos con el cálculo de la transformada de Laplace. Al aplicar muestreo opcional a la martingala $M_t = e^{\lambda B_t - \lambda^2 t}$ al tiempo de paro T_a (hasta el cual M permanece acotada) se ve que

$$\mathbb{E}\left(e^{\lambda T_a}\right) = e^{-a\sqrt{2\lambda}}.$$

Por una parte

$$\int_0^\infty \left(1 - e^{-\lambda x}\right) \nu(dx) = \lambda \int_0^\infty e^{-\lambda y} \overline{\nu}(y)\, dy,$$

donde $\overline{\nu}(y) = \int_y^\infty \nu(dx)$, y, utilizando la definición de la función Γ, vemos que

$$\sqrt{\frac{2}{\lambda}} = \int_0^\infty \frac{\sqrt{2} e^{-\lambda x}}{\sqrt{\pi x}}\, dx,$$

de lo cual se deduce que si

$$\int_0^\infty \left(1 - e^{-\lambda x}\right) \nu(dx) = \sqrt{2\lambda}$$

entonces

$$\overline{\nu}(y) = \sqrt{\frac{2}{\pi x}}$$

y por lo tanto

$$\nu(dx) = \frac{1}{\sqrt{2\pi x^3}} \mathbf{1}_{x>0}\, dx.$$

Por construcción, T es el inverso continuo por la derecha de S, por lo que es no-decreciente y continuo por la derecha y por lo tanto càdlàg. Para ver que $T_0 = 0$ casi seguramente, notemos que

$$\mathbb{E}\big(e^{-\lambda T_0}\big) = 1,$$

y como $e^{-\lambda T_0} \leq 1$ entonces $e^{-\lambda T_0} = 1$ casi seguramente.

Veamos ahora que T tiene incrementos independientes y estacionarios. Puesto que $T_{a+b} - T_a$ es el tiempo que transcurre para que $B_{\cdot + T_a} - a$ sobrepase b, por lo que la propiedad de Markov fuerte nos dice que $T_{a+b} - T_a$ tiene la misma distribución que T_b y además es independiente de $\mathscr{F}^B_{T_a}$. Por otra parte, $\mathscr{F}^T_a \subset \mathscr{F}^B_{T_a}$, por lo que T tiene incrementos independientes. Se concluye que T es un subordinador.

Finalmente, debemos ver que es un subordinador autosimilar. Veremos específicamente que T_{ca} y $c^2 T_a$ tienen la misma distribución. Esto se deduce de que ambas variables tienen transformada de Laplace $\lambda \mapsto e^{-ac\sqrt{2\lambda}}$. Una prueba basada en la autosimilitud del movimiento browniano además nos dice que los procesos $T_{ca}, a \geq 0$ y $c^2 T_a, a \geq 0$ tienen la misma distribución. En efecto, recordemos que puesto que $B_{ct}/\sqrt{c}, t \geq 0$ y $\sqrt{c} B_t, t \geq 0$ tienen la misma distribución. T es el proceso de tiempos de arribo del segundo mientras que $cT_{\sqrt{c}a}, a \geq 0$ es el proceso de tiempos de arribo del primero.

Finalmente, notamos que por autosimilitud y la relación entre $\overline{B}$ y T se sigue que

$$\mathbb{P}(T_a \leq t) = \mathbb{P}\big(a \leq \overline{B}_t\big) = \mathbb{P}\Big(a/\sqrt{t} \leq \overline{B}_1\Big) = \mathbb{P}\Big(a^2/\overline{B}_1^2 \leq t\Big).$$

□

Veamos ahora que la distribución de $\overline{B}_1$ se conoce explícitamente.

Proposición 6.8 (Principio de reflexión). *El proceso estocástico B^b dado por*

$$B^b_t = \begin{cases} B_t & t < T_b \\ 2b - B_t & t \geq T_b \end{cases}$$

es un movimiento browniano. En consequencia, la variable $\overline{B}_1$ tiene la misma distribución que $|B_1|$,

$$f_{B_1, \overline{B}_1}(a, b) = \mathbf{1}_{a<b} \frac{2b - a}{\sqrt{2\pi t^3}} e^{-(2b-a)^2/2t}$$

y

$$f_{T_a}(t) = \mathbf{1}_{t>0} \frac{1}{\sqrt{2\pi t^3}} e^{-a^2/2t}.$$

Demostración. Si $\tilde{B}_t = B_{t+T_b} - b$, entonces $\tilde{B}$ es un movimiento browniano independiente de $D_t = B_{t \wedge T_b}, t \geq 0$. Notemos que B se puede reconstruir a partir de D y $\tilde{B}$ a partir de la igualdad

$$B_t = \begin{cases} D_t & t < T_b \\ b + \tilde{B}_{t-T_b} & t > T_b \end{cases}.$$

Puesto que $-\tilde{B}$ también es un movimiento browniano, vemos que entonces el proceso

$$B_t^b = \begin{cases} D_t & t < T_b \\ b - \tilde{B}_{t-T_b} & t > T_b \end{cases} = \begin{cases} D_t & t < T_b \\ 2b - B_t & t \geq T_b \end{cases}$$

es un movimiento browniano.

A través de la igualdades de conjuntos

$$\{T_b \leq t, B_t \leq b\} = \left\{T_b \leq t, B_t^b \geq b\right\} = \left\{B_t^b \geq b\right\},$$

vemos que

$$\begin{aligned} \mathbb{P}(T_b \leq t) &= \mathbb{P}(B_t \geq b) + \mathbb{P}(B_t \leq b, T_b \leq t) \\ &= \mathbb{P}(B_t \geq b) + \mathbb{P}\big(B_t^b \geq b\big) \\ &= 2\mathbb{P}(B_t \geq b) \\ &= \mathbb{P}(|B_t| \geq b). \end{aligned}$$

Si $a \leq b$, apliquemos un argumento similar con los conjuntos

$$\left\{\overline{B}_t \geq b, B_t \leq a\right\} = \{T_b \leq t, B_t \leq a\} = \left\{B_t^b \geq 2b - a\right\}$$

para obtener

$$f_{B_t, \overline{B}_t}(a, b) = \mathbf{1}_{a<b} \frac{\partial}{\partial a} \frac{2e^{-(2b-a)^2/2t}}{\sqrt{2\pi t}} = \mathbf{1}_{a<b} \frac{2\,(2b-a)\,e^{-(2b-a)^2/2t}}{\sqrt{2\pi t^3}}.$$

□

Pasemos ahora al estudio del conjunto de ceros del movimiento browniano.

PROPOSICIÓN 6.9. *Sea $\mathscr{Z} = \{t \geq 0 : B_t = 0\}$. Entonces casi seguramente $\mathscr{Z}$ es un conjunto perfecto de medida de Lebesgue cero.*

DEMOSTRACIÓN. Puesto que $\mathbb{P}(B_t = 0) = 0$ para toda $t > 0$, el teorema de Fubini nos dice que $\lambda(\mathscr{Z} \cap [0, t]) = 0$ casi seguramente para toda $t \geq 0$.

Que $\mathscr{Z}$ es cerrado se sigue de que B tiene trayectorias continuas. Para ver que $\mathscr{Z}$ es perfecto, veamos primero que $0 \in \mathscr{Z}'$. Esto se sigue de la propiedad de Markov del movimiento browniano: sea d_t el primer punto a la derecha de t que pertenece a $\mathscr{Z}$. Entonces $d_t = t + T_0 \circ \theta_t$, por lo cual

$$\mathbb{E}\big(e^{-\lambda d_t}\big) = e^{-\lambda t}\mathbb{E}\Big(e^{-|B_t|\sqrt{2\lambda}}\Big).$$

La cantidad anterior converge a 1 conforme $t \to 0$, por lo cual $d_0 = \lim_{t\to 0} d_t = 0$. Sin embargo, $d_t \in \mathscr{Z}$, lo cual prueba que $0 \in \mathscr{Z}'$.

Por otra parte, d_t es tiempo de paro para cada $t > 0$, por lo cual $B_{d_t+\cdot}$ es un movimiento browniano para toda $t > 0$. Por lo tanto, $d_q \in \mathscr{Z}'$ para toda $q \in \mathbb{Q} \cap [0, \infty)$ casi seguramente. Si $h \in \mathscr{Z}$, sean q_n racionales que crecen a h. Entonces ó $d_{q_n} = h$ para n suficientemente grande ó $q_n \leq d_{q_n} < h$ para toda n, lo que implica que $h \in \mathscr{Z}'$. □

Enunciaremos dos resultados sobre el conjunto de ceros del movimiento browniano. El primero nos da la idea intuitiva de que el conjunto de ceros es una versión estocástica del conjunto triádico de Cantor.

Sea Ξ un proceso de Poisson puntual en $[0, \infty) \times [0, \infty)$ de intensidad $\mu(dx) = dx/2x^2$. Si los átomos de Ξ son (t_n, x_n), formemos al conjunto

$$\mathscr{U} = [0, \infty) \setminus \bigcup_n (s_n, s_n + x_n).$$

Entonces $\mathscr{U}$ tiene la misma distribución que $\mathscr{Z}$ en el sentido siguiente:

Proposición 6.10. *Para todo compacto $K \subset [0, \infty)$:*

$$\mathbb{P}(\mathscr{U} \cap K \neq \emptyset) = \mathbb{P}(\mathscr{Z} \cap K \neq \emptyset)\,.$$

El conjunto $\mathscr{U}$ tiene la particularidad de que podemos estudiar la intersección de conjuntos construidos con dos ó más procesos de Poisson independientes y probar que $\mathscr{U}_1 \cap \cdots \cap \mathscr{U}_n = \{0\}$ casi seguramente. En términos del browniano, estaríamos estudiando la intersección de los conjuntos de ceros de n brownianos independientes, que es el conjunto de ceros de un browniano n-dimensional. El que éste conjunto sea $\{0\}$ casi seguramente para $n \geq 2$ se conoce como polaridad del movimiento brownaino multidimensional.

Otro resultado interesante apunta en una dirección distinta. Se trata de probar que:

Proposición 6.11. *Los conjuntos cerrados aleatorios $\mathscr{Z}$ y $\left\{t \geq 0 : B_t = \overline{B}_t\right\}$ tienen la misma distribución.*

Terminar Hablando sobre tiempo local.

Recordemos que una variable Cauchy estándar tiene densidad

$$\frac{1}{\pi\,(1 + x^2)}.$$

En este caso la función de distribución es

$$\frac{1}{2} + \frac{\arctan(x)}{\pi}.$$

También se sabe que la distribución Cauchy es la misma que la del cociente de dos gaussianas estándar.

Proposición 6.12. *Sean B^1 y B^2 dos movimientos brownianos independientes y T^1 el proceso de tiempos de arribo de B^1. Entonces $B^2_{T^1_a}$ tiene la misma distribución que aC.*

Veamos ahora un resultado clásico conocido como primera ley arcoseno de Paul Lévy. Sean

$$d_t = \inf\{s \geq t : B_s = 0\} \quad \text{y} \quad g_t = \sup\{s \leq t : B_s < 0\}.$$

Proposición 6.13. *Sea C una variable Cauchy; entonces d_t tiene la misma distribución que $t\left(1 + C^2\right)$ y g_1 tiene la misma distribución que $1/(1 + C^2)$.*

Demostración. Notemos que $d_t = t + T_0 \circ \theta_t$. Por lo tanto, si N es una gaussiana estándar independiente de B, se tiene que

$$\begin{aligned}
\mathbb{P}(d_t > r) &= \mathbb{P}(t + T_0 \circ \theta_t > r) \\
&= \mathbb{P}\big(t + B_t^2/N^2 > r\big) \\
&= \mathbb{P}\big(t\left(1 + B_1^2/N^2\right) > r\big) \\
&= \mathbb{P}\big(t\left(1 + C^2\right) > r\big).
\end{aligned}$$

Por otra parte, vemos que

$$\mathbb{P}(g_1 < t) = \mathbb{P}(d_t > 1) = \mathbb{P}\big(t\left(1 + C^2\right) > 1\big) = \mathbb{P}\left(\frac{1}{1 + C^2} < t\right).$$

Al derivar, obtenemos la densidad arcoseno para g_1:

$$f_{g_1}(t) = 2\frac{1}{\pi\,(1/t)}\frac{1}{2}\sqrt{\frac{t}{t-1}}\frac{1}{t^2} = \frac{1}{\pi\sqrt{t}\sqrt{1-t}}.$$

$\square$

Bibliografía

[BGT87] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, Cambridge, 1987. MR 898871 (88i:26004)

[Kal02] Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)

[KS63] H. Kesten and F. Spitzer, *Ratio theorems for random walks. I*, J. Analyse Math. **11** (1963), 285–322. MR 0162279 (28 #5478)

[Pit75] J. W. Pitman, *One-dimensional Brownian motion and the three-dimensional Bessel process*, Advances in Appl. Probability **7** (1975), no. 3, 511–526. MR 0375485 (51 #11677)