Apunte Procesos Estocásticos I

Versión 2025-1

18 de noviembre de 2024

Índice

1. Introducción

Un **experimento aleatorio** es una receta o algoritmo para realizar una acción $\Rightarrow$, cuyo resultado no está completamente determinado. Al conjunto de posibles resultados asociado a un cierto experimento lo llamamos su espacio de resultados o su **espacio muestral**,denotado por S. Asociado a un experimento dado, podemos pensar también en el espacio de *todas sus posibles realizaciones*, que llamaremos Ω. Por cada subconjunto de posibles resultados $A \subseteq S$, tenemos un subconjunto $\tilde{A} \subseteq \Omega$ dado por

$$\tilde{A} = \{\text{todas las posibles realizaciones de mi experimento cuyo resultado cayó en A}\}.$$

Teniendo presente esta relación, usaremos el espacio de resultados S y el resultado de realizaciones Ω de manera paralela.[1]

Ahora, si repetimos un experimento de forma independiente, digamos N veces, podemos obtener una *frecuencia de aparición de eventos,* dada por el cociente

$$f^N(A) := \frac{\text{\#veces que obtengo un resultado incluído en A}}{\text{\#veces que realizo el experimento}} = \frac{\sum_{i=1}^{N} \mathbb{1}_{\{r_i \in A\}}}{N}.$$

Heurísticamente, se puede ver que existe $f(A) := \lim_{N\to\infty} f^N(A)$. Abstrayendo este concepto frecuentista (y sus propiedades), tenemos la siguiente definición axiomática.

Definición 1. *Dado un experimento con un espacio muestral Ω, y dada una σ-álgebra $\mathscr{F} \subseteq \mathscr{P}(\Omega)$ - cuyos elementos llamamos eventos-, definimos una **probabilidad** como una función*
$P : \mathscr{F} \to [0,1]$ *que satisface*

[A1] $P(A) \geq 0 \quad \forall A \in \mathscr{F}$,

[A2] $P(\Omega) = 1$,

[A3] *Si $A_1, A_2, \cdots \in \mathscr{F}$ son mutuamente excluyentes, entonces* $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

De esta definición se deducen varias propiedades y teoremas, relacionados también con el concepto de eventos independientes.

Recordemos también que partiendo de un *espacio de probabilidad* $(\Omega, \mathscr{F}, P)$ se pueden definir funciones que permitan describir los eventos del espacio original como eventos en $\mathbb{R}$, para poder

[1] En este sentido, abusaremos de la notación escribiendo A como subconjunto de Ω ;)

usar toda la teoría que ya conocemos. Llamamos entonces **variable aleatoria** a una función de la forma $X:\Omega \to \mathbb{R}$ tal que $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathscr{F} \quad \forall B \in \mathscr{B}$, dónde $\mathscr{B}$ corresponde a la σ-álgebra de los borelianos en $\mathbb{R}$. Así, $\forall B \in \mathscr{B}$, podemos definir $P_X(B) = P(X^{-1}(B))$, que denotaremos simplemente como $P(X \in B)$.

En Probabilidad I se trabajó con el concepto de variable aleatoria y las distribuciones de probabilidad más conocidas, así como también algunas propiedades y resultado relacionados.

Por otro lado, si queremos codificar más de una característica asociada a un experimento aleatorio dado, podemos describir los eventos de nuestro espacio original como eventos multidimensionales en $\mathbb{R}^d$, generalizando la noción de variable aleatoria. En este sentido, un **vector aleatorio** será entonces una función de la forma $\mathbf{X}:\Omega \to \mathbb{R}^d$ tal que $\mathbf{X^{-1}(B)} \in \mathscr{F} \quad \forall \mathbf{B} \in \mathscr{B}^d$, dónde $\mathscr{B}^d$ corresponde a la σ-álgebra de los borelianos en $\mathbb{R}^d$, lo que es equivalente a que cada coordenada $X_i:\Omega \to \mathbb{R}$, $i \in \{1,\dots,d\}$ es una variable aleatoria. En Probabilidad II aprendimos a definir la distribución conjunta asociada a estos vectores, y a trabajar con la dependencia entre distintas variables aleatorias.

Ahora bien, en muchos casos un experimento aleatorio tiene además una evolución temporal, y en ese caso querríamos codificar los resultados asociados al proceso como función del tiempo. Así, la continuación natural de la definición es pensar en *vectores aleatorios de tamaño infinito*, y con esa idea intuitiva definiremos los procesos estocásticos.

2. Procesos Estocásticos

Definición 2. *Dado $(\Omega,\mathscr{F},P)$ un espacio de probabilidad, un* ***proceso estocástico*** *es una colección de variables aleatorias $\mathbf{X} = (X_t,\ t \in T)$ en ese espacio, donde T se llama ESPACIO PARAMETRAL y todas las variables toman valores en un mismo conjunto S que llamaremos ESPACIO DE RESULTADOS.*

- Si T es un conjunto discreto (podemos considerar en general $T = \mathbb{N}$), decimos que el proceso es **a tiempo discreto**; y si T es un conjunto continuo (en general $T = \mathbb{R}_+ = [0,\infty)$), se dice que el proceso evoluciona **a tiempo continuo**.
- Si S es un conjunto discreto (en general $S \subseteq \mathbb{Z}^N$ para algun $N \in \mathbb{N}$), decimos que el proceso es **a valores discretos**; y si S es un conjunto continuo (en general $S \subseteq \mathbb{R}^N$ para algun $N \in \mathbb{N}$), decimos que el proceso es **a valores continuos**.

Nota. *Podemos pensar el proceso como una función* $\mathbf{X} : \Omega \to S^{\otimes T}$*, donde para cada realización* $\omega_0 \in \Omega$*, tenemos que* $\mathbf{X}(\omega_0) : T \to S$ *es una función* *determinista**, que asigna a cada tiempo* t *un valor* $X_t(\omega_0)$*, y que llamaremos una* ***realización*** *o* ***trayectoria*** *asociada al proceso.*

Por otro lado, dado $\mathbf{X}(\omega) = (X_t(\omega),\ t \in T)$ *para cada* $\omega \in \Omega$*, si fijamos un tiempo* $t_0 \in T$*, tenemos que* $X_{t_0} : \Omega \to S$ *es una variable aleatoria.*

¿Qué nos interesa estudiar? Nos interesa calcular probabilidades asociadas a eventos, que en el caso más simple tomarán la forma

$$\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\},$$

con $A_1, \ldots A_n$ eventos en la σ-álgebra correspondiente asociada al conjunto S. Luego veremos como definir eventos más generales, y como usarlos para estudiar el comportamiento del proceso a largo plazo.

Algunos tipos de procesos estocásticos son los siguientes:

★ ***Proceso de ensayos independientes.*** En este caso, $(X_n,\ n \in \mathbb{N}_0)$ es un conjunto de variables aleatorias independientes. En particular, si además son identicamente distribuídas, podemos pensar que este proceso corresponde a repeticiones independientes de un mismo experimento aleatorio en cada instante de tiempo, como por ejemplo lanzar un volado ($X_n \sim Be(0{,}5)$), hacer una medición para control de calidad ($X_n \sim \mathcal{N}(a, \varepsilon)$), etc.

¿Qué pasa si queremos modelar un proceso de evolución con dependencia?¿Cuál es el modelo más básico que se nos puede ocurrir?

★ ***Proceso de Markov.*** Decimos que un proceso $(Y_t,\ t \in T)$ satisface la **propiedad de Markov** si se cumple que para todo conjunto de tiempos $0 \le t_0 \le t_1 \le \cdots \le t_n \le t_{n+1} \in T$,

$$\mathbb{P}\left(Y_{t_{n+1}} = j | Y_{t_n} = i_n, Y_{t_{n-1}} = i_{n-1}, \ldots, Y_{t_0} = i_0\right) = \mathbb{P}\left(Y_{t_{n+1}} = j | Y_{t_n} = i_n\right)$$

para todo $\{i_0, \ldots, i_n, j\} \in S$. Esta propiedad puede resumirse de la siguiente manera: EL PASO SIGUIENTE EN EL FUTURO SÓLO DEPENDE DEL LUGAR EN EL QUE ESTOY EN EL PRESENTE, Y NO DE COMO LLEGUÉ HASTA AQUÍ.

Nota. *Es importante recordar que para que una igualdad se cumpla, ambos lados tienen que estar bien definidos. En este caso eso quiere decir que para poder calcular las respectivas probabilidades condicionales necesitamos que*

$$\mathbb{P}\left(Y_{t_n} = i_n, Y_{t_{n-1}} = i_{n-1}, \ldots, Y_{t_0} = i_0\right) > 0 \qquad y \qquad \mathbb{P}\left(Y_{t_n} = i_n\right) > 0,$$

lo que puede interpretarse como que el pasado al que estoy condicionando es un "camino" permitido para mi proceso.

¿Qué situaciones podemos modelar que satisfagan esta hipótesis?

- ***Borrachito***. Supongamos que un individuo se encuentra parado en una esquina, pero no recuerda en que dirección debe ir. Entonces elige *al azar* alguna de las esquinas adyacentes a las que se encuentra, y se mueve hacia allí. Si tomamos $(X_n,\ n \in \mathbb{N}_0)$ el proceso definido por

$$X_n = \textit{ubicación del individuo luego de n pasos},$$

  vemos que satisface la propiedad de Markov, ya que cada paso que doy (y por lo tanto la ubicación siguiente) solo depende de la ubicación actual, no del camino que recorrí en el pasado.

- ***Apostador.*** Supongamos que decido jugar *Blackjack*, apostando \$1 en cada partida, hasta que se me acabe el dinero o acumule una ganancia de $\$N$. Si empiezo con un capital de $Y_0 = \$c$, antes de empezar la k-ésima partida decido si seguir jugando o no según el dinero Y_{k-1} que tengo acumulado hasta ese momento (no necesito saber explícitamente cuando dinero tenía en todos los pasos anteriores).

- ***Modelo Web***. Supongamos que actualmente me encuentro en una cierta página web, y a partir de allí elijo la próxima página a visitar al azar usando sólo las ligas que aparecen en la página actual. Es fácil convencerse que el proceso así definido es Markoviano, pero si lo hacemos un poco más realista teniendo en cuenta que puedo "volver para atrás" desde la página en la que me encuentro (aunque no haya un link disponible en la página misma), ya pierdo esta propiedad.

Durante este curso analizaremos en profundidad estos ejemplos y varias de las propiedades de este tipo de procesos.

★ ***Proceso estacionario (homogéneo en el tiempo).*** Se dice que un proceso estocástico $(X_t,\ t \in T)$ es **estacionario** si $\forall\ 0 \leq t_1 < t_2 < \ldots < t_k$ se cumple que

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \stackrel{\mathscr{D}}{=} (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})^2 \quad \forall h : t_1 + h, \ldots, t_k + h \in T,$$

En particular $X_t \stackrel{\mathscr{D}}{=} X_{t+h}$ para todo $t, t + h \in T$.

[2]la notación $\stackrel{\mathscr{D}}{=}$ su usa para denotar la igualdad en distribución (o en ley).

★ ***Proceso a incrementos estacionarios.*** Se dice que un proceso estocástico $(X_t,\ t \in T)$ tiene **incrementos estacionarios** si para todo $s \leq t : s, t, (s+k), (t+k) \in T$ se cumple que

$$\Delta X_{s+k}^{t+k} := X_{t+k} - X_{s+k} \stackrel{\mathcal{D}}{=} X_t - X_s =: \Delta X_s^t,$$

es decir que la distribución de un incremento ΔX_s^t sólo depende del tamaño del intervalo de tiempo en el que estoy mirando el incremento, no del instante en el que lo estoy mirando. Si queremos modelar la cantidad de gente que entró a una tienda durante cada día, podemos simplificar nuestro modelo suponiendo que tiene incrementos estacionarios, es decir que la cantidad de personas sólo depende del tiempo de observación, y no si es jueves o martes.

Ojo! Un proceso estacionario no es lo mismo que un proceso a incrementos estacionarios!! Es un buen ejercicio pensar porqué ;)

★ ***Proceso a incrementos independientes.*** Se dice que un proceso estocástico $(X_t,\ t \in T)$ tiene **incrementos independientes** si $\forall\ t_0 \leq t_1 < t_2 < \ldots < t_k$

$$\Delta X_{t_0}^{t_1}, \Delta X_{t_1}^{t_2}, \ldots, \Delta X_{t_{k-1}}^{t_k} \text{ son variables independientes entre sí,}$$

es decir los eventos que ocurren en intervalos de tiempos disjuntos son independientes entre sí. Si pensamos nuevamente en el proceso que modela la gente que entra a una tienda, es natural pensar que tiene incrementos independientes, ya que la cantidad de gente que visita la tienda una cierta semana es independiente de la que la visitó la semana anterior.

★ ***Proceso de Lévy.*** Se conoce con este nombre a los procesos a incrementos estacionarios e independientes. Revisaremos luego en profundidad dos ejemplos muy importantes de este tipo de procesos: el *Movimiento Browniano* y el *proceso de Poisson*.

★ ***Martingalas.*** se dice que un proceso estocástico $\{X_n\}_{n \in \mathbb{N}}$ es una **martingala** si se cumple la siguiente condición

$$\mathbb{E}\left[X_{n+1} | X_0 = x_0, X_1 = x_1 \ldots, X_n = x_n\right] = x_n \quad \forall n \in \mathbb{N},$$

es decir que EL VALOR MEDIO DEL PROCESO A TIEMPO FUTURO ES IGUAL AL ÚLTIMO VALOR OBSERVADO. Su nombre deviene de la palabra francesa *martingale*, que usaban los apostadores en el siglo XVIII para denotar la estrategia de juego que consistía en apostar a que va a suceder lo mismo que acaba de pasar. Una forma de pensar en la propiedad que caracteriza a las martingalas es pensar que tienen una ley de evolución *equilibrada* o *simétrica*. Durante este curso también estudiaremos un poco más sobre este tipo de procesos.

Nota. *Esta clasificación no es exhaustiva ni mucho menos mutuamente excluyente. Por ejemplo, existen procesos markovianos que no son estacionarios y procesos estacionarios que no son markovianos; martingalas que no son markovianas y procesos de Markov que no son martingalas. La única inclusión que resaltamos es que, si notamos que siempre podemos escribir la variable* X_t *como* $X_t = (X_t - X_s) + X_s$, *para todo* $s \leq t$, *podemos deducir fácilmente que un proceso a incrementos independientes satisface la propiedad de Markov.*

3. Cadenas de Markov

Definición 3. *Diremos que un proceso estocástico a tiempo discreto* $(X_n,\ n \in \mathbb{N})$ *con espacio de estados* S *es una* ***cadena de Markov*** *si satisface la* ***propiedad de Markov****, es decir, si para todo* $0 \le k_0 \le k_1 \le \cdots \le k_n \le k_{n+1} \in \mathbb{N}$ *y* $i_0, \ldots, i_n, j \in S$ *se cumple que*

$$\mathbb{P}\left[X_{k_{n+1}} = j | X_{k_0} = i_0, X_{k_1} = i_1, \ldots, X_{k_n} = i_n\right] = \mathbb{P}\left[X_{k_{n+1}} = j | X_{k_n} = i_n\right]. \tag{1}$$

Para entender mejor esta definición, volvamos a nuestro ejemplo del apostador.

Ejemplo 4. [***Ruina del Jugador.***] Un jugador juega cierto juego de apuestas, en el que gana 1$ con probabilidad p y pierde 1$ con probabilidad $1-p$. Podemos pensar entonces las partidas como una sucesión $\{\xi_i\}_{i\in\mathbb{N}}$ de variables aleatorias i.i.d., con distribución

$$\xi_i = \begin{cases} 1 & \text{con probabilidad } p \\ -1 & \text{con probabilidad } 1-p \end{cases}.$$

El juego se termina cuando el jugador alcanza una cierta suma (digamos N$) o se queda sin dinero y lo echan del casino. Si queremos conocer el proceso $(X_n,\ n \in \mathbb{N})$ dado por

$$X_n = \textit{Cantidad de dinero luego de n partidas},$$

iniciando con un capital inicial $X_0 = c$, tenemos que

$$X_{n+1} = \begin{cases} X_n + \xi_{n+1} & \text{si } X_n \in (0, N) \\ X_n & \text{si } X_n = 0 \text{ ó } X_n = N \end{cases}.$$

Vemos entonces que $(X_n,\ n \in \mathbb{N})$ es un proceso estocástico con espacio de estados finito
$S = \{0, 1, 2, \ldots, N\}$, y que satisface la propiedad de Markov, ya que

$$\mathbb{P}\left[X_{k_{n+1}} = j | X_{k_0} = i_0, X_{k_1} = i_1, \ldots, X_{k_n} = i_n\right] = \mathbb{P}\left[X_{k_{n+1}} = j | X_{k_n} = i_n\right],$$

pues en cada paso $n+1$ me pregunto si sigo jugando o no dependiendo si ya llegué a 0 ó N; y si decido seguir jugando la próxima partida es independiente de todas las partidas anteriores.

Es fácil ver que

$$p(i,j) := \mathbb{P}\left[X_{n+1} = j | X_n = i\right] = \begin{cases} p & \text{si } i \in \{1, 2, \ldots N-1\},\ j = i+1 \\ 1-p & \text{si } i \in \{1, 2, \ldots N-1\},\ j = i-1 \\ 1 & \text{si } i = j = 0 \\ 1 & \text{si } i = j = N \\ 0 & \text{cualquier otro caso.} \end{cases}, \tag{2}$$

y, en este caso, estas probabilidades de transición no dependen del paso n en el que las estoy mirando (son **homogéneas en el tiempo**.) Todas las probabilidades asociadas a la evolución de esta cadena pueden calcularse usando las probabilidades anteriores, que llamaremos **probabilidades de transición**. Supongamos que el capital inicial no es un número fijo c, sino que $X_0 \sim U\{0,1,2,\ldots,N\}$. Si queremos conocer la distribución de X_1, que corresponde al dinero que tengo luego de la primera partida, tenemos que

$$P\left[X_1 = j\right] = \sum_{i=0}^{N} P\left[X_1 = j | X_0 = i\right] P\left[X_0 = i\right] \qquad \forall i, j \in S.$$

usando el Teorema de Probabilidad Total. Por ejemplo, si $j = 0$, notamos que si no tengo dinero luego de la primera partida puede ser por dos razones: o empecé sin dinero, por lo que me echaron del casino y continuaré asi eternamente; o empece con sólo 1$ y lo perdí en la primera partida. Es decir, usando (2) tenemos que

$$\begin{aligned} P\left[X_1 = 0\right] &= P\left[X_1 = 0 | X_0 = 0\right] P\left[X_0 = 0\right] + P\left[X_1 = 0 | X_0 = 1\right] P\left[X_0 = 1\right] \\ &= 1 \times \frac{1}{N+1} + (1-p) \times \frac{1}{N+1} = \frac{2-p}{N+1}. \end{aligned}$$

Haciendo un análisis similar, tenemos que

$$\begin{aligned} P\left[X_1 = N\right] &= P\left[X_1 = N | X_0 = N\right] P\left[X_0 = N\right] + P\left[X_1 = N | X_0 = N-1\right] P\left[X_0 = N-1\right] \\ &= 1 \times \frac{1}{N+1} + p \times \frac{1}{N+1} = \frac{1+p}{N+1}; \end{aligned}$$

y para todo $2 \le k \le N-2$

$$\begin{aligned} P\left[X_1 = k\right] &= P\left[X_1 = k | X_0 = k-1\right] P\left[X_0 = k-1\right] + P\left[X_1 = k | X_0 = k+1\right] P\left[X_0 = k+1\right] \\ &= p \times \frac{1}{N+1} + (1-p) \times \frac{1}{N+1} = \frac{1}{N+1}. \end{aligned}$$

Notemos que cuando $k = 1$ sucede algo diferente. Si me encuentro en 1 luego de la primera partida, quiere decir que solo pude haber empezado con 2$ y haber perdido, pues si hubiese empezado sin dinero no podría haber seguido jugando. Tenemos entonces que

$$P\left[X_1 = 1\right] = P\left[X_1 = 1 | X_0 = 2\right] P\left[X_0 = 2\right] = \frac{1-p}{N+1},$$

y análogamente

$$P\left[X_1 = N-1\right] = P\left[X_1 = N-1 | X_0 = N-2\right] P\left[X_0 = N-2\right] = \frac{p}{N+1}.$$

En resumen, dada una **distribución inicial** (asociada a X_0) y las **probabilidades de transición**, calculamos la distribución de X_1, que resulta ser

$$P\left[X_1 = j\right] = \begin{cases} \frac{2-p}{N+1} & \text{si } j = 0 \\ \frac{1-p}{N+1} & \text{si } j = 1 \\ \frac{1}{N+1} & \text{si } 2 \leq j \leq N-2 \\ \frac{p}{N+1} & \text{si } j = N-1 \\ \frac{1+p}{N+1} & \text{si } j = N \end{cases},$$

y podemos de forma similar calcular cualquier probabilidad de eventos asociados a este proceso.

Revisemos ahora de manera formal los conceptos que aparecieron en el ejemplo anterior.

Definición 5. *Decimos que una cadena de Markov es **homogénea** si* $\forall i, j \in S$,

$$p(i,j) := P\left[X_1 = j | X_0 = i\right] = P\left[X_{n+1} = j | X_n = i\right] \qquad \forall n \in \mathbb{N},$$

y llamaremos a $p(i,j)$ *la **probabilidad de transición del estado** i **al estado** j **en un paso**.*

Ojo! Cuando decimos que una cadena de Markov es homogénea, nos referimos a que tiene transiciones estacionarias en el tiempo, pero no es necesariamente un proceso estacionario en el sentido de la definición de la página 6.

Así, podemos acomodar la información de las probabilidades de transición en una **matriz de transición** P de tamaño $|S| \times |S|$. En nuestro ejemplo del apostador tenemos

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1-p & 0 & p & \cdots & 0 & 0 & 0 \\ 0 & 1-p & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1-p & 0 & p \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Vemos que $p(0,0) = 1$ y $p(N,N) = 1$, por lo que decimos que 0 y N son estados **absorbentes**.

Proposición 6. *Una matriz de transición P es una **matriz estocástica**, es decir*

(i) P tiene entradas no negativas $p(i,j) \geq 0$.

(ii) P tiene filas que suman 1, es decir $\sum_{j \in S} p(i,j) = 1 \; \forall i \in S$.

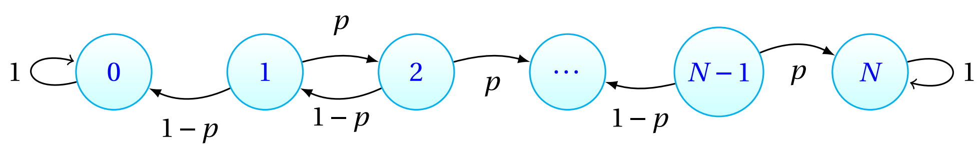
Dem.

(i) Por definición cada entrada de la matriz corresponde a una probabilidad, por lo que $p(i,j) \in [0,1]$.

(ii) $\sum_{j \in S} p(i,j) = \sum_{j \in S} \mathbb{P}\left[X_1 = j | X_0 = i\right] = \frac{\sum_{j \in S} \mathbb{P}[X_1 = j, X_0 = i]}{\mathbb{P}[X_0 = i]} = \frac{\mathbb{P}[X_0 = i]}{\mathbb{P}[X_0 = i]} = 1$

□

Otra forma de representar mis probabilidades de transición es a través de una gráfica. En este caso, los nodos representan los estados, y las aristas(direccionadas) entre dos nodos existirán cuando exista una probabidad positiva de pasar de un estado a otro, y tendrán asociado un peso (correspondiente a la probabilidad de transición respectiva). En nuestro ejemplo



Entonces, ¿qué necesito para estudiar la evolución de una cierta cadena? Además las transiciones, necesito saber "como empecé", es decir la distribución asociada al paso inicial.

Definición 7. *Denotamos por $\mu_0 = (\mu_0(i), i \in S)$ una* ***distribución inicial*** *de la cadena, que corresponde a la distribución de la variable aleatoria X_0, es decir*

$$\mu_0(i) = \mathbb{P}[X_0 = i] \qquad \forall i \in S.$$

En el ejemplo del apostador, trabajamos con la distribución uniforme sobre el conjunto $S = \{0, 1, \ldots, N\}$, es decir $\mu_0(i) = \frac{1}{N+1}$ para todo $i \in S$. Claramente, podríamos haber empezado de otra manera, por ejemplo desde un capital inicial fijo \$c , con $0 < c < N$. En ese caso, la distribución inicial estaria dada por $\tilde{\mu}_0(i) = \mathbb{1}_{\{i=c\}}$.

Tenemos entonces que todas las probabilidades asociadas a mi cadena pueden calcularse conociendo la distribución inicial μ_0 y la matriz de transición P. Cómo en el ejemplo del apostador, vimos que para todo $j \in S$,

$$\mathbb{P}\left[X_1 = j\right] = \sum_{i \in S} \mathbb{P}\left[X_1 = j | X_0 = i\right] \mathbb{P}[X_0 = i] = \sum_{i \in S} \mu_0(i) p(i,j) = \left[\mu_0 P\right]_j,$$

donde el último término corresponde a la entrada j-esima del vector obtenido al multiplicar el vector μ_0 por la matriz P. De la misma manera

$$\begin{aligned}\mathbb{P}\left[X_2 = j\right] &= \sum_{k\in S} \mathbb{P}\left[X_2 = j | X_1 = k\right] \mathbb{P}\left[X_1 = k\right] = \sum_{k\in S} p(k,j) \left[\sum_{i\in S} \mu_0(i) p(i,k)\right] \\ &= \sum_{i\in S} \mu_0(i) \left[\sum_{k\in S} p(i,k)p(k,j)\right] = \left[\mu_0 P^2\right]_j;\end{aligned}$$

y así sucesivamente. Formalmente, si para $i, j \in S$, $n \in \mathbb{N}$ definimos

$$p_n(i,j) := P\left[X_n = j | X_0 = i\right]$$

la **probabilidad de transición de i a j en n pasos**, tenemos el siguiente resultado.

Proposición 8 (**Ecuaciones de Chapman-Kolmogorov**). *Para todo par de naturales $m \le n$ y para todo par de estados $i, j \in S$, ocurre que*

$$p_n(i,j) = \sum_{k\in S} p_m(i,k) p_{n-m}(k,j) \tag{3}$$

Dem.

$$\begin{aligned}p_n(i,j) &= \frac{P\left[X_n = j, X_0 = i\right]}{P\left[X_0 = i\right]} = \sum_{k\in S} \frac{P\left[X_n = j, X_m = k, X_0 = i\right]}{P\left[X_0 = i\right]} \frac{P\left[X_m = k, X_0 = i\right]}{P\left[X_m = k, X_0 = i\right]} \\ &= \sum_{k\in S} P\left[X_n = j | X_m = k, X_0 = i\right] P\left[X_m = k | X_0 = i\right] \\ &= \sum_{k\in S} P\left[X_n = j | X_m = k\right] P\left[X_m = k | X_0 = i\right] = \sum_{k\in S} p_{n-m}(k,j) p_m(i,k).\end{aligned}$$

□

Como un corolario del resultado anterior, podemos usar inducción matemática completa para probar la siguiente afirmación.

Corolario 9. *Dada la matriz de transición P asociada a una cadena de Markov, se cumple que para todo $n \in \mathbb{N}$, las transiciones en n pasos satisfacen*

$$p_n(i,j) = [P^n]_{ij}.$$

Nota. *Los resultados anteriores nos dicen que la familia de matrices $\{P_n\}_{n\in\mathbb{N}}$, con P_n la matriz de transiciones en n pasos, forman lo que se conoce como un **semigrupo**; es decir satisfacen que*

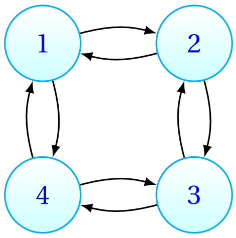
$$P^0 = I \qquad y \qquad P^{n+m} = P^n P^m \quad \forall n, m \in \mathbb{N}.$$

Corolario 10. *Sea $(X_n,\, n \in \mathbb{N})$ una cadena de Markov con matriz de transición P. Dada una distribución inicial μ_0, tenemos que*

$$\mathbb{P}\left[X_n = j\right] = \sum_{i \in S} \mu_0(i) p_n(i,j) = \sum_{i \in S} \mu_0(i) [P^n]_{ij} = [\mu_0 P^n]_j \qquad \forall j \in S.$$

Nota. *Este último corolario nos brinda un método para calcular las probabilidades de mi cadena en cualquier instante de tiempo. En la práctica calcular potencias de matrices no es siempre fácil, pero si la matriz P es diagonalizable, podemos usar el **Teorema de Descomposición Espectral**, que nos dice que existe una matriz diagonal D (formada por los autovalores asociados) y otra matriz Q (cuyas columnas son los autovectores asociados) tal que $P = QDQ^{-1}$. Entonces, tenemos que $P^n = QD^nQ^{-1}$, dónde D^n no es más que la matriz diagonal formada por las potencias n-ésimas de los autovalores asociados.*

Retomando el ejemplo del borrachito, podemos escribir la matriz de transición asociada y calcular todas sus potencias. En particular, podemos asociarle la gráfica



dónde cada arista tiene asociada una probabilidad de $p = 1/2$. Es fácil ver entonces que

$$P^{2k-1} = \begin{pmatrix} 0 & 0{,}5 & 0 & 0{,}5 \\ 0{,}5 & 0 & 0{,}5 & 0 \\ 0 & 0{,}5 & 0 & 0{,}5 \\ 0{,}5 & 0 & 0{,}5 & 0 \end{pmatrix} \qquad P^{2k} = \begin{pmatrix} 0{,}5 & 0 & 0{,}5 & 0 \\ 0 & 0{,}5 & 0 & 0{,}5 \\ 0{,}5 & 0 & 0{,}5 & 0 \\ 0 & 0{,}5 & 0 & 0{,}5 \end{pmatrix}$$

para todo $k \in \mathbb{N}$. Entonces, si la distribución inicial esta dada por

$$\mu_0(i) = 1/4 \qquad \forall i \in \{1,2,3,4\},$$

entonces tenemos que

$$\mathbb{P}\left[X_n = j\right] = \left[\mu_0 P^n\right]_j = 1/4 \qquad \forall j \in \{1,2,3,4\},$$

y las distribuciones de la cadena al tiempo n se mantienen uniformes para todo $n \in \mathbb{N}$.

En cambio, si tomamos como distribución inicial

$$\tilde{\mu}_0(i) = \mathbb{1}_{\{i=2\}} \qquad \forall i \in \{1,2,3,4\},$$

entonces

$$\mathbb{P}\left[\widetilde{X}_{2k} = j\right] = 0{,}5 \quad \text{si } j \in \{2,4\} \qquad \text{y} \qquad \mathbb{P}\left[\widetilde{X}_{2k-1} = j\right] = 0{,}5 \quad \text{si } j \in \{1,3\}.$$

El comportamiento que presenta esta cadena es un ejemplo de como a pesar de tener probabilidades de transición fijas, la distribución inicial afecta la distribución de la cadena en cada tiempo dado.

3.1. Ejemplos de Procesos de Markov

Ejemplo 11 (**Urna de Ehrenfest.**). Supongamos que tenemos N bolitas numeradas, $\{1,2,\ldots N\}$, distribuídas en dos urnas A y B. En cada paso elijo como una bolita al azar y la cambio de urna. Definimos el proceso $(X_n, n \in \mathbb{N})$ por

$$X_n = \text{ cantidad de bolitas en la urna A luego de } n \text{ extracciones.}$$

Es fácil deducir que este proceso es una cadena de Markov, ya que cada elección es independiente de las demás, y el movimiento de la bolita sólo depende de su ubicación en ese momento. Además es homógenea, ya que las transiciones no dependen de la cantidad del número de extracciones que haya hecho antes. Es importante notar que el proceso no toma en cuenta <u>cuales</u> bolitas están en cada urna, solo la <u>cantidad</u> de bolitas, por lo que $S = \{0,1,\ldots,N\}$.

Tenemos entonces que para todo $0 < i < N$

$$\begin{aligned}
p(i,i+1) = P[X_1 = i+1|X_0 = i] &= P[\{\text{elijo una bolita de la urna B}\}|\{\text{tengo } i \text{ bolitas en la urna A }\}] \\
&= P[\{\text{elijo una bolita de la urna B}\}|\{\text{tengo } N-i \text{ bolitas en la urna B }\}] \\
&= \frac{N-i}{N};
\end{aligned}$$

y por otro lado

$$\begin{aligned}
p(i,i-1) = P[X_1 = i-1|X_0 = i] &= P[\{\text{elijo una bolita de la urna A}\}|\{\text{tengo } i \text{ bolitas en la urna A }\}] \\
&= \frac{i}{N}.
\end{aligned}$$

En el caso $i = 0$ corresponde a que no hay bolitas en la urna A, por lo que necesariamente deberé elegir una bolita de la urna B y moverla, por lo que $p(1,0) = 1$; y haciendo un análisis análogo tenemos que $p(N-1,N) = 1$. En este caso, decimos que 0 y N son estados **reflectantes**.

La matriz de transición asociada será entonces de la forma

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & \cdots & 0 & 0 & 0 \\ 0 & \frac{2}{N} & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Ahora bien, si X_0 tiene distribución uniforme sobre $\{0,1,\ldots,N\}$ (es decir $\pi_0(i) = \frac{1}{N+1}$), ¿Qué distribución tendrá X_1?

Ejemplo 12 (**Cadena de Variables Aleatorias I.I.D.)**. Sea $(Z_n, n \in \mathbb{N})$ el proceso en el que las vs. as. Z_n son independientes e igualmente distribuidas, con distribución

$$\mathbb{P}[Z_n = i] = a_i \quad \text{para todo } i \in \mathbb{N}.^3$$

Es fácil ver que esta cadena satisface la propiedad de Markov, ya que

$$\mathbb{P}\left[Z_{k_{n+1}} = j | Z_{k_0} = i_0, \ldots, Z_{k_n} = i_n\right] = \mathbb{P}\left[Z_{k_{n+1}} = j\right] = \mathbb{P}\left[Z_{k_{n+1}} = j | Z_{k_n} = i_n\right]$$

por independencia, y ademas es homógenea, dado que

$$P[Z_{n+1} = j | Z_n = i] = P[Z_{n+1} = j] = a_j = p(i,j) \qquad \text{para todo } i, j \in \mathbb{N},\ n \in \mathbb{N}.$$

Entonces la matriz de transición toma la forma

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_n & \cdots \\ a_0 & a_1 & a_2 & \cdots & a_n & \cdots \\ a_0 & a_1 & a_2 & \cdots & a_n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_0 & a_1 & a_2 & \cdots & a_n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

De la matriz se nota claramente que no me importa de que estado parto, sino hacia que estado voy. ¿Que pasa con la matriz de transición en n pasos?

[3] obviamente, $\{a_i\}_{i\in\mathbb{N}}$ debe satisfacer $\sum_{i\in\mathbb{N}} a_i = 1$.

Ejemplo 13 (**Cadena de Máximos de Variables Aleatorias I.I.D.**). Sea $(M_n,\, n \in \mathbb{N})$ la cadena dada por $M_n = \text{máx}\{Z_0, Z_1, \ldots, Z_n\}$ para el proceso $(Z_n,\, n \in \mathbb{N})$ definido como en el ejemplo anterior. Vemos que podemos escribir

$$M_{n+1} = \text{máx}\{Z_0, Z_1, \ldots, Z_n, Z_{n+1}\} = \text{máx}\{\text{máx}\{Z_0, Z_1, \ldots, Z_n\}, Z_{n+1}\} = \text{máx}\{M_n, Z_{n+1}\},$$

de dónde podemos deducir la propiedad de Markov, notando además que $M_n \perp\!\!\!\perp Z_{n+1}$. Tenemos entonces que

$$\begin{aligned}
\mathbb{P}\left[M_{n+1} = j | M_n = i\right] &= \mathbb{P}\left[\text{máx}\{M_n, Z_{n+1}\} = j | M_n = i\right] = \mathbb{P}\left[\text{máx}\{i, Z_{n+1}\} = j | M_n = i\right] \\
&= \frac{\mathbb{P}\left[\text{máx}\{i, Z_{n+1}\} = j, M_n = i\right]}{\mathbb{P}\left[M_n = i\right]} = \frac{\mathbb{P}\left[\text{máx}\{i, Z_{n+1}\} = j\right]\mathbb{P}\left[M_n = i\right]}{\mathbb{P}\left[M_n = i\right]} \\
&= \mathbb{P}\left[\text{máx}\{i, Z_{n+1}\} = j\right],
\end{aligned}$$

y como las variables Z_k son i.i.d, tenemos que la cadena es homogénea, con

$$p(i,j) = \begin{cases} 0 & \text{si } i > j \\ a_j & \text{si } i < j \\ \sum\limits_{k \geq i} a_k & \text{si } i = j \end{cases},$$

y de aquí

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_n & \cdots \\ 0 & a_0 + a_1 & a_2 & \cdots & a_n & \cdots \\ 0 & 0 & a_0 + a_1 + a_2 & \cdots & a_n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sum\limits_{k=0}^{n} a_k & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

Notamos entonces que la matriz de transición es triangular, ¿cómo son las potencias de matrices triangulares?

Ejemplo 14 (**Castillo de Naipes.**). Queremos armar un castillo de naipes. Para eso contamos con infinitas cartas, pero sucede que dada una cierta configuración actual, sabemos que con cierta probabilidad p agregaremos dos cartas, y con probabilidad $1-p$ derribaremos todo el castillo en el intento. Por lo tanto, si definimos la cadena $(X_n,\, n \in \mathbb{N})$ dada por

$$X_n = \text{cantidad de naipes en el castillo luego de } n \text{ pasos},$$

con espacio de estados $S = \{0,2,4,\ldots\}$, es fácil deducir que satisface la propiedad de Markov y es homógenea, con probabilidad de transición

$$p(i,j) = \mathbb{P}\left[X_{n+1} = j | X_n = i\right] = \begin{cases} p & \text{si } j = i+2, i \in S \\ 1-p & \text{si } j = 0, i \in S \\ 0 & \text{si no.} \end{cases}$$

Si empezamos de $X_0 \equiv 0$, ¿cuál es la probabilidad de estar en 0 luego de n pasos?

3.2. Características de una Cadena de Markov

3.2.1. Clases de Comunicación

Queremos estudiar más en profundidad las probabilidades de transición de un estado a otro. Para eso, tenemos la siguiente definición

Definición 15. *Decimos que el estado j es* ***accesible*** *desde el estado i si existe $n \in \mathbb{N}_0$ tal que $p_n(i,j) > 0$. Lo denotamos como $i \to j$.*

En el ejemplo del castillo de naipes, es fácil ver que $i \to 0$ para todo $i \in \{0,2,4,6,\ldots\}$, ya que $p(i,0) = 1-p > 0$; y también se cumple que $i \to i+2k$ para todo $k \in \mathbb{N}$ ya que $p_k(i,i+2k) = p^k > 0$.

Definición 16. *Si se cumple que $i \to j$ y $j \to i$ decimos que los estados i y j* ***están comunicados****, y denotamos $i \longleftrightarrow j$.*

Proposición 17. *La relación $\longleftrightarrow$ es una relación de equivalencia, y por lo tanto define una partición (en clases de equivalencia) en el espacio de estados S.*

Dem. Para ver que es una relación de equivalencia tenemos que mostrar que se cumplen las siguientes tres condiciones

(i) Reflexividad: Notemos que siempre se cumple $i \longleftrightarrow i$, ya que $p_0(i,i) = 1$.

(ii) Simetría: Por definición $i \longleftrightarrow j$ implica $i \to j$ y $j \to i$, por lo que $j \longleftrightarrow i$.

(iii) Transitividad: Supongamos que se cumple $i \longleftrightarrow j$ y $j \longleftrightarrow k$, y por lo tanto existen $n_1, n_2, m_1, m_2 \in \mathbb{N}_0$ tales que $p_{n_1}(i,j) > 0$, $p_{n_2}(j,i) > 0$, $p_{m_1}(j,k) > 0$ y $p_{m_2}(k,j) > 0$. Usando Ch-K 3, tenemos que

$$p_{n_1+m_1}(i,k) \geq p_{n_1}(i,j)p_{m_1}(j,k) > 0 \quad \text{y} \quad p_{m_2+n_2}(k,i) \geq p_{m_2}(k,j)p_{n_2}(j,i) > 0,$$

por lo que $i \longleftrightarrow k$.

$\square$

Llamaremos entonces ***clases de comunicación*** a las clases de equivalencia definidas por la relación de comunicación, es decir para cada estado $i \in S$, tenemos

$$\mathscr{C}_i := \{j \in S : i \longleftrightarrow j\}.$$

Definición 18. *Decimos que una cadena de Markov es **irreducible** si existe una única clase de comunicación, es decir todos los estados se comunican entre sí.*

En los ejemplos anteriores, ¿hay alguno que NO sea irreducible?

Ejemplo 19 (**Proceso de ramificación (Proceso de Galton-Watson).**[4]). Sean $\{\xi^{i,n}\}_{i,n\in\mathbb{N}}$ una familia de variables aleatorias independientes e idénticamente distribuídas a valores en $\mathbb{N}$, con función de probabilidad asociada $\{p_k\}_{k\in\mathbb{N}_0}$, es decir $P\left(\xi^{i,n} = k\right) = p_k$. Podemos pensar que

$$\xi^{i,n} = \text{ cantidad de hijos que tiene el } i\text{-ésimo individuo en la } n\text{-ésima generación,}$$

con la intención de estudiar la evolución de una población en la que cada individuo muere luego de una generación , dejando k descendientes con probabilidad p_k. Entonces el proceso $(Z_n, n \in \mathbb{N}_0)$, dado por

$$Z_n = \text{ cantidad de individuos en la generación } n,$$

es una cadena de Markov homogénea con espacio de estados $S = \mathbb{N}$, donde dada una población inicial Z_0, se cumple que para todo $n \geq 0$, $i, j > 0$

$$p(i,j) = \mathbb{P}\left[Z_{n+1} = j | Z_n = i\right] = \mathbb{P}\left[\sum_{r=1}^{Z_n} \xi^{r,n} = j | Z_n = i\right] = \mathbb{P}\left[\sum_{r=1}^{i} \xi^{r,n} = j | Z_n = i\right] = \mathbb{P}\left[\sum_{r=1}^{i} \xi^{r,n} = j\right],$$

dónde esta última es la distribución de la suma de i vs.as. i.i.d., independientes de las anteriores. Notemos que $p(0,0) = 1$, pues si no tengo ningún individuo no puede haber reproducción, y el estado cero es un estado absorbente. En este sentido es claro que $p_n(0, j) = 0$ para todo $j \in \{1,2,\ldots\}$, $n \geq 0$, de donde $\mathscr{C}_0 = \{0\}$, y por lo tanto la cadena no será irreducible.

3.2.2. Período de una cadena

Definición 20. *El **período** de un estado $i \in S$ está dado por*

$$d(i) = \text{mcd}\{n \geq 1 : p_n(i,i) > 0\},$$

[4] Si quieren saber más...

es decir el máximo común divisor del conjunto de número de pasos en los que que puedo volver al estado i. Decimos que un estado es ***aperiódico si*** $d(i) = 1$.

Por convención diremos que $d(i) = 0$ *si* $p_n(i,i) = 0$ *para todo* $n > 0$.

Definición 21. *Decimos que una cadena de Markov es* ***aperiódica*** *si* $d(i) = 1$ *para todo* $i \in S$.

En el ejemplo del borrachito que camina en una cuadra, vemos que $d(1) = \text{mcd}\{2,4,6,\dots\} = 2$, y lo mismo vale para los demás estados posibles de la cadena.

En el caso del apostador, vemos que

$$d(0) = \text{mcd}\{1,2,\dots\} = 1 = d(N), \qquad \text{y} \qquad d(i) = \text{mcd}\{2,4,6,\dots\} = 2 \quad \forall i \in \{1,2,\dots,N-1\}.$$

¿Qué pasa en los demás ejemplos? ¿Hay algún caso en que la cadena sea aperiódica?

OJO! Notar que, por ejemplo, $\text{mcd}\{2,3\} = 1$, por lo que el período de un estado podría ser 1 aún en el caso cuando no hay probabilidad positiva de volver en un paso.

Proposición 22. *El período es una propiedad de las clases de comunicación. Es decir,*

$$i \longleftrightarrow j \qquad \Longrightarrow \qquad d(i) = d(j).$$

Dem. Dado $i \neq j$, si $i \longleftrightarrow j$ entonces $\exists\ n,m : p_n(i,j) > 0$ y $p_m(j,i) > 0$. Entonces tenemos que $p_{n+m}(i,i) > 0$ por Ch-K, y por lo tanto $n+m \in \{l \geq 1 : p_l(i,i) > 0\}$, por lo que este conjunto es no vacío. Ahora, dado $s_0 \in \{l \geq 1 : p_l(i,i) > 0\}$, se cumple que para todo $k \in \mathbb{N}$

$$p_{m+ks_0+n}(j,j) \geq p_m(j,i)p_{s_0}(i,i)\cdots p_{s_0}(i,i)p_n(i,j) > 0,$$

por lo que $(m+ks_0+n) \in \{v \geq 1 : p_v(j,j) > 0\}$. Notemos que por definición de período se cumple que $d(i)$ divide a s_0, y por lo anterior $d(j)$ divide a $(m+ks_0+n)$ para todo k, por lo que en particular $d(j)$ divide a s_0. Esto último se debe a que podemos escribir

$$s_0 = (m+2s_0+n) - (m+ks_0+n) = d(j)r_2 - d(j)r_1 = d(j)(r_2 - r_1).$$

Entonces, como $d(j)$ es divisor del conjunto $\{l \geq 1 : p_l(i,i) > 0\}$, debe ser que $d(j) \leq d(i)$ por ser este último el **máximo comun divisor**.

De forma análoga (solo intercambiando nombres) podemos deducir que $d(i) \leq d(j)$, por lo que $d(i) = d(j)$ como afirmamos. $\square$

Notemos del resultado anterior que, si una cadena es irreducible, basta comprobar que un estado cualquiera es aperiódico para afirmar que toda la cadena lo es.

3.2.3. Tiempos de primera visita (*hitting times*)

Definición 23. *Sea $A \subseteq S$ un subconjunto de estados de una cadena de Markov $(X_n,\ n \in \mathbb{N}_0)$. El* ***tiempo de primera visita*** *a A es la variable aleatoria*

$$\tau_A := \text{mín}\{n \geq 1 : X_n \in A\} = \textit{ la primera vez que el proceso visita } A.$$

Por convención, $\text{mín}\,\emptyset = \infty$*, es decir, si $X_n \notin A\ \forall n \in \mathbb{N}$, $\tau_A = \infty$ (por lo que τ_A es una variable aleatoria que tomará valores en $\overline{\mathbb{R}}_0^+$).*

Si $A = \{j\}$, escribimos $\tau_{\{j\}} = \tau_j$, y si tenemos como distribución inicial $X_0 \equiv i$, escribimos

$$\tau_j = \tau_{ij} = \textit{ la primera vez que visitamos } j \textit{ dado que partimos de } i\,.$$

Para conocer la distribución de esta nueva variable aleatoria, llamaremos $f_n(i,j)$ a la probabilidad de que la cadena visite el estado j por primera vez en n pasos dado que partimos del estado i, es decir

$$f_n(i,j) = \mathbb{P}\left[X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j | X_0 = i\right] = \mathbb{P}\left[\tau_{ij} = n\right].$$

Por definición, tenemos que $f_1(i,j) = p(i,j)$.

Volvamos a nuestro castillo de naipes para calcular algunas probabilidades asociadas a tiempos de primera visita. En este caso, tenemos que

$$f_n(0,2) = \mathbb{P}\left[X_n = 2, X_{n-1} \neq 2, \ldots, X_1 \neq 2 | X_0 = 0\right] = (1-p)^{(n-1)}p$$

$$f_n(0,0) = \mathbb{P}\left[X_n = 0, X_{n-1} \neq 0, \ldots, X_1 \neq 0 | X_0 = 0\right] = p^{(n-1)}(1-p)$$

$$f_1(2k, 2k+2) = p, \quad f_{k+2}(2k,2k+2) = (1-p)p^{k+1}, \quad f_r(2k,2k+2) = 0 \quad \forall r \in \{2,4,\ldots,k,k+1\}.$$

¿que pasa con $f_n(2k,2k+2)$ para $n > k+2$?

Notemos que en general $f_n(i,j) \neq p_n(i,j)$, ya que la probabilidad de transición en n pasos tiene en cuenta <u>todos</u> los caminos que me llevan del estado i al estado j en n pasos (incluídos los que pasan por j en algún tiempo intermedio). En particular, tenemos el siguiente resultado.

Proposición 24. *Dado $n \geq 1$, $i, j \in S$ se cumple que*

$$p_n(i,j) = \sum_{k=1}^{n} f_k(i,j)\, p_{(n-k)}(j,j).$$

Dem. Aplicando el Teorema de Probabilidad Total, podemos particionar las trayectorias que me llevan de i a j en n pasos según el momento que visitamos j por primera vez, y tenemos entonces que

$$\begin{aligned}
p_n(i,j) &= \frac{\mathbb{P}\left[X_n = j, X_0 = i\right]}{\mathbb{P}\left[X_0 = i\right]} \\
&= \sum_{k=1}^{n} \frac{\mathbb{P}\left[X_n = j, X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i\right]}{\mathbb{P}\left[X_0 = i\right]} \frac{\mathbb{P}\left[X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i\right]}{\mathbb{P}\left[X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i\right]} \\
&= \sum_{k=1}^{n} \mathbb{P}\left[X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j | X_0 = i\right] \mathbb{P}\left[X_n = j | X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j, X_0 = i\right] \\
&= \sum_{k=1}^{n} \mathbb{P}\left[X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j | X_0 = i\right] \mathbb{P}\left[X_n = j | X_k = j\right] = \sum_{k=1}^{n} f_k(i,j) p_{n-k}(j,j)
\end{aligned}$$

□

3.2.4. Recurrencia y Transitoriedad

Estamos especialmente interesados en los tiempos de paro de la forma

$$\tau_i := \tau_{ii} = \textit{ tiempo de primer retorno al estado } i, \quad i \in S.$$

Usando estas variables, podemos clasificar los estados de una cadena dependiendo de si sabemos con certeza si regresaremos o no.

Definición 25. *Un estado $i \in S$ se dice **recurrente** si*

$$f_{ii} = \mathbb{P}\left[\tau_i < \infty\right] = 1,$$

*es decir si la probabilidad de regresar a i (partiendo de i) es segura. Un estado no recurrente, es decir cuando $f_{ii} < 1$, se dice **transitorio**.*

Nota. *Notemos que, como*

$$\{X_n = i, X_{n-1} \neq i, \dots, X_1 \neq i, X_0 = i\} \cap \{X_k = i, X_{k-1} \neq i, \dots, X_1 \neq i, X_0 = i\} = \emptyset \quad \forall k \neq n,$$

entonces

$$f_{ii} = \mathbb{P}\left[\tau_i < \infty\right] = \sum_{n=1}^{\infty} \mathbb{P}\left[\tau_i = n\right] = \sum_{n=1}^{\infty} f_n(i,i).$$

El siguiente resultado nos dá un criterio para esta clasificación

Proposición 26. *El estado $i \in S$ es*

a) recurrente si $\sum_{n=1}^{\infty} p_n(i,i) = \infty$;

b) transitorio si $\sum\limits_{n=1}^{\infty} p_n(i,i) < \infty$.

Dem. Dada la variable

$$N^i\Big|_{X_0=i} = \text{ cantidad de veces que regresamos al estado i (partiendo de i) } = \left(\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=i\}}\right)\Big|_{X_0=i},$$

se cumple que

$$\begin{aligned}
\mathbb{P}\left[N^i \geq k | X_0 = i\right] &= \sum_{m=1}^{\infty} \frac{\mathbb{P}\left[N^i \geq k, X_m = i, X_{m-1} \neq i, \ldots, X_1 \neq i, X_0 = i\right]}{\mathbb{P}\left[X_0 = i\right]} \\
&= \sum_{m=1}^{\infty} \mathbb{P}\left[N^i \geq k | X_m = i, X_{m-1} \neq i, \ldots, X_1 \neq i, X_0 = i\right] f_m(i,i) \\
&= \sum_{m=1}^{\infty} \mathbb{P}\left[N^i \geq k-1 | X_0 = i\right] f_m(i,i) = \mathbb{P}\left[N^i \geq k-1 | X_0 = i\right] \sum_{m=1}^{\infty} f_m(i,i) \\
&= \mathbb{P}\left[N^i \geq k-1 | X_0 = i\right] f_{ii}.
\end{aligned}$$

dónde en el penúltimo renglón aplicamos la propiedad de Markov para deducir que si sabemos que hasta m sólo hubo un retorno, desde ahí debe haber por lo menos $k-1$ retornos. Si hacemos esto de forma recursiva, obtenemos que

$$\mathbb{P}\left[N^i \geq k | X_0 = i\right] = \mathbb{P}\left[N^i \geq k-1 | X_0 = i\right] f_{ii} = \cdots = \mathbb{P}\left[N^i \geq 1 | X_0 = i\right] (f_{ii})^{k-1} = (f_{ii})^k.$$

Podemos deducir entonces que $N^i|_{X_0=i} \sim Geom(f_{ii})$, y por lo tanto podemos calcular su esperanza como

$$E[N^i | X_0 = i] = \sum_{k=1}^{\infty} \mathbb{P}\left[N_i \geq k | X_0 = i\right] = \begin{cases} \frac{1}{1-f_{ii}} & \text{si } f_{ii} < 1 \\ \infty & \text{si } f_{ii} = 1. \end{cases}$$

Por otro lado, tenemos que por definición

$$E[N^i | X_0 = i] = E[\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=i\}} | X_0 = i] = \sum_{n=1}^{\infty} E[\mathbb{1}_{\{X_n=i\}} | X_0 = i] = \sum_{n=1}^{\infty} \mathbb{P}[X_n = i | X_0 = i] = \sum_{n=1}^{\infty} p_n(i,i),$$

por lo que igualando ambos resultados tenemos que

$$\sum_{n=1}^{\infty} p_n(i,i) = \begin{cases} \frac{1}{1-f_{ii}} & \text{si } f_{ii} < 1 & \Leftrightarrow & i \text{ es transitorio} \\ \infty & \text{si } f_{ii} = 1 & \Leftrightarrow & i \text{ es recurrente}. \end{cases}$$

□

En el ejemplo del castillo de naipes, tenemos que

$$f_{00} = \sum_{n=1}^{\infty} f_n(0,0) = (1-p) \sum_{n=1}^{\infty} p^{n-1} = (1-p)\frac{1}{1-p} = 1,$$

por lo que el estado 0 es recurrente. ¿Qué pasa con los demás estados? Podemos contestar a esta pregunta fácilmente usando el siguiente resultado.

Proposición 27. *La recurrencia es una propiedad de las clases de comunicación, es decir*

i) si i es recurrente e $i \longleftrightarrow j$, entonces j es recurrente.

ii) si i es transitorio e $i \longleftrightarrow j$, entonces j es transitorio.

Dem.

i) Por definición, si $i \longleftrightarrow j$, $\exists n, m : p_n(i,j) > 0$ y $p_m(j,i) > 0$, por lo que

$$p_{m+r+n}(j,j) \geq p_m(j,i)p_r(i,i)p_n(i,j) \quad \forall r \in \mathbb{N},$$

y entonces

$$\sum_{r=1}^{\infty} p_{m+r+n}(j,j) \geq p_m(j,i) \left(\sum_{r=1}^{\infty} p_r(i,i) \right) p_n(i,j).$$

Por lo tanto, si i es recurrente, por la Proposición 26 tenemos que $\sum_{r=1}^{\infty} p_r(i,i) = \infty$, por lo que

$$\sum_{l=1}^{\infty} p_l(j,j) \geq \sum_{r=1}^{\infty} p_{m+r+n}(j,j) \geq p_m(j,i) \left(\sum_{r=1}^{\infty} p_r(i,i) \right) p_n(i,j) = \infty,$$

de donde j es recurrente.

ii) Supongamos que la implicancia no es cierta, es decir i es transitorio con $i \longleftrightarrow j$ pero j es recurrente. Podemos aplicar el ítem anterior para j recurrente, deduciendo así que i es recurrente, lo que contradice nuestra hipótesis.

□

Estudiemos ahora algunas nociones más relacionadas con la recurrencia y transitoriedad, que serán de gran utilidad para estudiar luego que pasa con el comportamiento de la cadena a largo plazo.

Definición 28. *El **tiempo medio de recurrencia de un estado** j **partiendo de un estado** i está definido como*

$$\mu_{ij} := E[\tau_{ij}] = \sum_{n=1}^{\infty} n f_n(i,j).$$

En particular $\mu_i := E[\tau_i] = \sum_{n=1}^{\infty} n f_n(i,i)$.

Nota. *Claramente, la definición anterior solo es interesante cuando i es recurrente, pues si i es transiente (es decir, hay una probabilidad positiva de no visitar nunca más el estado i) sabemos que* $\mathbb{P}[\tau_i < \infty] = \sum_{n=1}^{\infty} f_n(i,i) < 1$, *por lo que* $\mathbb{P}[\tau_i = \infty] > 0$, *y por lo tanto se cumple que*

$$E[\tau_i] = E\left[\tau_i \mathbb{1}_{\{\tau_i < \infty\}}\right] + E\left[\tau_i \mathbb{1}_{\{\tau_i = \infty\}}\right] = \infty,$$

y por lo tanto no estará bien definida.

Ojo! podría darse el caso en que i es recurrente, es decir $\sum_{n=1}^{\infty} f_n(i,i) = 1$, pero aun así $\mu_i = \sum_{n=1}^{\infty} n f_n(i,i)$ no converge.

Definición 29. *Un estado i recurrente se dice* ***recurrente positivo*** *si $\mu_i < \infty$, y se dice* ***recurrente nulo*** *si $\mu_i = \infty$.*

Ejemplo 30. En el caso del castillo de naipes, tenemos que

$$\mu_0 = \sum_{n=1}^{\infty} n f_n(0,0) = \sum_{n=1}^{\infty} n(1-p)p^{(n-1)} = \frac{(1-p)}{(1-p)^2} = \frac{1}{1-p}$$

Ejemplo 31. ¿Que pasa con la caminata aleatoria simétrica en los enteros? Es fácil ver que es irreducible, por lo que basta estudiar la recurrencia para un cierto estado, digamos el origen. ¿El origen es un estado recurrente? Si es así, ¿cuál es el tiempo medio de recurrencia?

Si un estado es aperiódico y recurrente positivo decimos que es **ergódico**. Los resultados que veremos a continuación nos permitirán conocer en profundidad el comportamiento a largo plazo de las cadenas ergódicas, es decir en las que todos sus estados son ergódicos.

El siguiente teorema relaciona la cantidad de visitas promedio a un cierto estado i con su tiempo medio de recurrencia.

Teorema 32 (TEOREMA ERGÓDICO PARA CADENAS DE MARKOV)**.** *Dada una cadena irreducible, si definimos, para cualesquiera $i, j \in S$, la sucesión de variables $\{N_n^{ij}\}_{n\in\mathbb{N}}$ dada por*

$$N_n^{ij} = \textit{cantidad de visitas al estado } j \textit{ (partiendo de } i\textit{) en los } n \textit{ primeros pasos},$$

se cumple que

$$\lim_{n\to\infty} \frac{N_n^{ij}}{n} = \frac{1}{\mu_j} \qquad \textit{casi seguramente},$$

dónde $\lim_{n\to\infty} \frac{N_{i,j}(n)}{n} = 0$ si $\mu_j = \infty$.

Dem. Queremos estudiar el orden de convergencia de la sucesión de variables $\{N_n^{ij}\}_{n\in\mathbb{N}}$. Primero que nada, de la definición notamos que $N_n^{ij} = \left(\sum_{k=1}^{n} \mathbb{1}_{\{X_k=j\}}\right)\Big|_{X_0=i}$, por lo que podemos definir

$$N^j\Big|_{X_0=i} = \lim_{n\to\infty} N_n^{ij} = \left(\sum_{k=1}^{\infty} \mathbb{1}_{\{X_k=j\}}\right)\Big|_{X_0=i},$$

en el mismo espíritu de la variable $N^i|_{X_0=i}$ usada en la demostración de la Proposición 26. Es fácil ver entonces, de forma análoga a la prueba para N^i, que $\mathbb{P}\left[N^j \geq k|X_0=i\right] = f_{ij}\left(f_{jj}\right)^{k-1}$. De aqui podemos deducir algunas cosas, por ejemplo

$$\mathbb{P}\left[N^j=0|X_0=i\right]=1-\mathbb{P}\left[N^j\geq 1|X_0=i\right]=1-f_{ij};$$

$$\mathbb{P}\left[N^j=\infty|X_0=i\right]=\lim_{k\to\infty}\mathbb{P}\left[N^j>k|X_0=i\right]=\begin{cases}0 & \text{si } f_{jj}<1\\ f_{ij} & \text{si } f_{jj}=1\end{cases};$$

y además

$$\mathbb{E}\left[N^j|X_0=i\right]=\sum_{k\geq 0}\mathbb{P}\left[N^j>k|X_0=i\right]=\sum_{k\geq 0}f_{ij}\left(f_{jj}\right)^k=\begin{cases}\frac{f_{ij}}{1-f_{jj}} & \text{si } f_{jj}<1\\ \text{diverge!} & \text{si } f_{jj}=1\end{cases},$$

Entonces, tenemos dos casos bien diferenciados

- Si la cadena es transiente, $f_{jj}<1$ (y $\mu_j=\infty$), por lo que $\mathbb{P}\left[N^j=\infty|X_0=i\right]=0$, de donde existe un conjunto $A\subseteq\Omega$, con $\mathbb{P}(A)=1$ tal que $\forall\omega\in A$, $\{N_n^{ij}(\omega)\}_{n\in\mathbb{N}}$ es creciente y acotada (por $N^j(\omega)\Big|_{X_0=i}$), y por lo tanto

$$\lim_{n\to\infty}\frac{N_n^{ij}}{n}=0=\frac{1}{\mu_j}\qquad c.s.$$

- Si la cadena es recurrente, $f_{jj}=1$, por lo que

$$\mathbb{P}\left[N^j=\infty|X_0=i\right]=f_{ij}\qquad \text{y}\qquad \mathbb{P}\left[N^j=0|X_0=i\right]=1-f_{ij}.$$

  Probaremos ahora que la cadena sea irreducible y recurrente implica que $f_{ij}=1$ y por lo tanto $\mathbb{P}\left[N^j=\infty|X_0=i\right]=1$, y podemos probar que

$$\lim_{n\to\infty}\frac{N_n^{ij}}{n}=\lim_{n\to\infty}\frac{N_n^{jj}}{n}\qquad c.s. \tag{4}$$

  Para esto, notemos primero que, como $i\rightleftarrows j$, $\exists m\in\mathbb{N}: p_m(j,i)>0$. Definimos entonces la sucesión de eventos $\{A_n^m\}_{n\in\mathbb{N}}$ dados por

$$A_n^m=\left\{X_{(n-1)m+\sum_{k=1}^{n-1}k}=i,\quad X_{(n-1)m+\sum_{k=1}^{n}k}=j,\quad \text{dado que } X_{nm+\sum_{k=1}^{m}k}=i\right\},$$

  que tienen probabilidad

$$\mathbb{P}\left[A_n^m\right]=\mathbb{P}\left[X_{n+m}=i,X_n=j|X_0=i\right]=p_n(i,j)p_m(j,i),$$

dónde en la primera igualdad usamos la homogeneidad del proceso. Además es fácil ver que para todo $n \neq k$, $A_n^m \perp\!\!\!\perp A_k^m$ por propiedad de Markov, ya que

$$\mathbb{P}\left[A_n^m \cap A_k^m\right] = p_n(i,j)p_m(j,i)p_k(i,j)p_m(j,i) = \mathbb{P}\left[A_n^m\right]\mathbb{P}\left[A_k^m\right].$$

Usando la primera parte del Lema de Borel-Cantelli, como

$$\sum_{n\geq 1}\mathbb{P}(A_n^m) = \sum_{n\geq 1} p_n(i,j)p_m(j,i) = p_m(j,i)E\left[\sum_{n\geq 1}\mathbb{1}_{\{X_n=j\}}|X_0=i\right] = p_m(j,i)E\left[N^j|X_0=i\right] = \infty$$

y los eventos son independientes tenemos que $\mathbb{P}\left[\limsup A_n^m\right] = 1$, es decir

$$\mathbb{P}\left[\bigcap_{n\geq 1}\bigcup_{k\geq n} A_n^m\right] = 1 \Rightarrow \mathbb{P}\left[\bigcup_{n\geq 1}\bigcap_{k\geq n}(A_n^m)^C\right] = 0 \Rightarrow \mathbb{P}\left[\bigcap_{k\geq n}(A_n^m)^C\right] = 0\ \forall n\geq 1,$$

y por lo tanto

$$\mathbb{P}\left[N^{ij}=0\right] = \mathbb{P}\left[\bigcap_{k\geq n}(A_n^m)^C\right] = 0 \Rightarrow f_{ij} = \mathbb{P}\left[\tau_{ij}<\infty\right] = 1.$$

De aquí se deduce que existe un conjunto $\tilde{A}\subseteq\Omega$, con $\mathbb{P}(\tilde{A}) = 1$ tal que $\forall\omega\in\tilde{A}\ \tau_{ij}(\omega)<\infty$, y podemos escribir $N^{ij}_{\tau_{ij}(\omega)+n}(\omega) = 1 + N_n^{jj}(\omega)$, de donde

$$\lim_{n\to\infty}\frac{N_n^{ij}(\omega)}{n} = \lim_{n\to\infty}\frac{N^{ij}_{\tau_{ij}(\omega)+n}(\omega)}{\tau_{ij}(\omega)+n} = \lim_{n\to\infty}\frac{1+N_n^{jj}(\omega)}{n}\frac{n}{\tau_{ij}(\omega)+n} = \lim_{n\to\infty}\frac{N_n^{jj}(\omega)}{n}$$

y obtenemos la igualdad 4.

Ahora estamos en condiciones calcular $\lim\limits_{n\to\infty}\dfrac{N_n^{jj}}{n}$. Para esto, definimos la sucesión de variables $\{Y_k^j\}_{k\in\mathbb{N}}$, con

$$Y_k^j = \{\text{cantidad de pasos entre la } k-1\text{-ésima visita y la } k\text{-ésima visita}\}.$$

Es fácil ver que $Y_1^j = \tau_j$, y por la propiedad de Markov las variables $\{Y_k^j\}$ son i.i.d. Tenemos entonces que $Y_1^j+\cdots+Y^j_{N_n^{jj}} \leq n \leq Y_1^j+\cdots+Y^j_{N_n^{jj}}+Y^j_{N_n^{jj}+1}$, y

$$\frac{Y_1^j+\cdots+Y^j_{N_n^{jj}}}{N_n^{jj}} \leq \frac{n}{N_n^{jj}} \leq \frac{Y_1^j+\cdots+Y^j_{N_n^{jj}}+Y^j_{N_n^{jj}+1}}{N_n^{jj}+1}\frac{N_n^{jj}+1}{N_n^{jj}}$$

donde el primer y último termino en la cadena de desigualdades convergen a $\mathbb{E}\left[Y_1^j\right] = \mathbb{E}\left[\tau_j\right] = \mu_j$ por Ley de los Grandes Números.

Si j es recurrente positivo, entonces $\mu_j<\infty$ y por lo tanto $\lim\limits_{n\to\infty}\dfrac{n}{N_n^{jj}} = \mu_j$, de donde tenemos el límite deseado.

Si j es recurrente nulo, entonces $\mu_j = \infty$, de donde $\dfrac{n}{N_n^{jj}}$ diverge c.s. y por lo tanto $\lim\limits_{n\to\infty}\dfrac{N_n^{jj}}{n} = 0$.

□

Usando el Teorema de Convergencia Acotada, podemos deducir el siguiente resultado.

Corolario 33. *Dada una cadena irreducible, para cualesquiera $i, j \in S$ se cumple que*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{p_k(i,j)}{n} = \frac{1}{\mu_j}.$$

Dem. Por el teorema ergódico, tenemos que $\lim_{n\to\infty} \frac{N_n^{ij}}{n} = \frac{1}{\mu_j}$ c.s.. Notemos además que por definición $|\frac{N_n^{ij}}{n}| \leq 1$, por lo que podemos aplicar el Teorema de Convergencia Acotada [5] para concluir que

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{p_k(i,j)}{n} = \lim_{n\to\infty} \mathbb{E}\left[\frac{\sum_{k=1}^{n} \mathbb{1}_{\{X_k=j\}}}{n} \middle| X_0 = i\right] = \lim_{n\to\infty} \mathbb{E}\left[\frac{N_n^{ij}}{N}\right] = \frac{1}{\mu_j}.$$

□

Este corolario nos permite probar que la recurrencia positiva es una propiedad de clase.

Proposición 34. *La recurrencia positiva es una propiedad de las clases de comunicación, es decir*

i) si i es recurrente positivo e $i \Leftrightarrow j$, entonces j es recurrente positivo.

ii) si i es recurrente nulo e $i \Leftrightarrow j$, entonces j es recurrente nulo.

Dem. En la proposición 27 vimos que la recurrencia es una propiedad de clase, asi que para probar el resultado nos basta estudiar el tiempo medio de recurrencia.

i) Si $i \Leftrightarrow j$, existe n, m tal que $p_n(i,j) > 0$ y $p_m(j,i) > 0$. Entonces para todo $k \in \mathbb{N}$ se cumple

$$p_{m+k+n}(j,j) \geq p_m(j,i)p_k(i,i)p_n(i,j) \Rightarrow \lim_{N\to\infty} \frac{\sum_{k=1}^{N} p_{m+k+n}(j,j)}{N} \geq \lim_{N\to\infty} \frac{p_m(j,i)p_n(i,j)\sum_{k=1}^{\infty} p_k(i,i)}{N},$$

y ya que $\lim_{N\to\infty} \frac{\sum_{k=1}^{N} p_k(i,i)}{N} = \frac{1}{\mu_i}$ y

$$\lim_{N\to\infty} \frac{\sum_{k=1}^{N} p_{m+k+n}(j,j)}{N} \leq \lim_{N\to\infty} \frac{\sum_{k=1}^{N} p_k(j,j)}{N} = \frac{1}{\mu_j},$$

tenemos que si i es recurrente positivo entonces $\frac{1}{\mu_i} > 0$, y por lo tanto

$$0 < p_m(j,i)\frac{1}{\mu_i}p_n(j,i) \leq \frac{1}{\mu_j} \Rightarrow \mu_j < \infty.$$

[5] Corresponde al Teorema 67 en el apunte de Probabilidad II

ii) Análogamente a lo anterior e intercambiando los lugares de i y j, tenemos que

$$\frac{1}{\mu_i} = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} p_{n+k+m}(i,i)}{N} \geq \lim_{N\to\infty} \frac{p_n(i,j)p_m(j,i)\sum_{k=1}^{\infty} p_k(j,j)}{N} = p_n(i,j)p_m(j,i)\frac{1}{\mu_j}.$$

Si i es recurrente nulo, entonces $\frac{1}{\mu_i} = 0$, por lo que

$$0 \leq p_n(i,j)p_m(j,i)\frac{1}{\mu_j} \leq \frac{1}{\mu_i} = 0 \Rightarrow \frac{1}{\mu_j} = 0,$$

ya que $p_n(i,j)p_m(j,i) > 0$, de donde deducimos que $\mu_j = \infty$.

□

Como consecuencia, tenemos que

Proposición 35. *No existen estados recurrentes nulos en cadenas finitas.*

Dem. Supongamos que i es un estado recurrente. Entonces su clase de comunicación $\mathscr{C}_i$ es cerrada, es decir $p(j,l) = 0$ siempre que $j \in \mathscr{C}_i$ y $l \notin \mathscr{C}_i$, y por lo tanto para cada $k \in \mathbb{N}$, tenemos que

$$\sum_{j\in\mathscr{C}_i} p_k(i,j) = 1 \qquad \Rightarrow \qquad \sum_{k=1}^{n} \frac{\sum_{j\in\mathscr{C}_i} p_k(i,j)}{n} = 1 \quad \forall n.$$

Como $\mathscr{C}_i$ es finita, $\sum_{j\in\mathscr{C}_i} p_k(i,j)$ es una suma finita, y puedo intercambiar sumas para obtener

$$1 = \sum_{j\in\mathscr{C}_i} \frac{\sum_{k=1}^{n} p_k(i,j)}{n} \to \sum_{j\in\mathscr{C}_i} \frac{1}{\mu_j},$$

por lo que debe ser $\sum_{j\in\mathscr{C}_i} \frac{1}{\mu_j} = 1$, por lo que existe $j \in \mathscr{C}_i$ con $\frac{1}{\mu_j} > 0$, es decir $\mu_j < \infty$ y j es recurrente positivo, por lo que h es recurrente positivo $\forall h \in \mathscr{C}_i$. □

Nota. *Los resultados anteriores nos dicen que si una cadena de Markov es irreducible y recurrente positiva, entonces*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{p_k(i,j)}{n} = \frac{1}{\mu_j} > 0 \qquad \forall i,j \in S.$$

Esto sera de vital importancia para conocer el comportamiento de la cadena a largo plazo, tema en el que nos adentraremos a continuación.

3.3. Tiempos de Paro y Propiedad de Markov Fuerte

Los tiempos de primera visita son un caso particular de un tipo de variable aleatoria conocidos como **tiempos de paro**, que son tiempos aleatorios de gran interés en el estudio de procesos estocásticos, y particularmente en cadenas de Markov. Para introducir este concepto, necesitamos primero definir una familia de $\sigma-$álgebras especiales, que codifican la información (en general relacionada con ciertos procesos) que se tiene hasta cada instante de tiempo.

Definición 36. *Dado un espacio de probabilidad $(\Omega, \mathscr{F}, P)$, una* ***filtración*** *es una colección de σ-álgebras $\{\mathscr{F}_n\}_{n\in\mathbb{N}}$ tal que $\mathscr{F}_n \subseteq \mathscr{F}_m \subseteq \mathscr{F}$ para todo $n \leq m$. En particular, la* ***filtración natural*** *o canónica de un proceso $(X_n,\ n \in \mathbb{N}_0)$ es aquella sucesión de σ-álgebras generadas por las variables asociadas al proceso, es decir*

$$\mathscr{F}_n := \sigma\{X_k : 0 \leq k \leq n\} = \sigma\{\{B \in \mathscr{B} : X_k^{-1}(B) \in \mathscr{F}\} : 0 \leq k \leq n\}.$$

Nota. *La σ-álgebra $\mathscr{F}_n$ representa entonces la información que tenemos hasta el tiempo n, es decir los eventos que podemos saber si sucedieron o no a partir de los valores que toman las variables $(X_0, X_1, \ldots, X_n)$. En este sentido, la idea de filtración es que mientras más tiempo va pasando, mas información tenemos.*

Definición 37. *Un* ***tiempo de paro*** *para un proceso de Markov $(X_n,\ n \in \mathbb{N}_0)$ es una variable aleatoria $\tau : \Omega \to \mathbb{N}_0$ adaptada a la filtración natural del proceso, es decir que para todo $n \geq 0$, se cumple que*

$$\{\tau \leq n\} \in \mathscr{F}_n.$$

Nota. *Intuitivamente, podemos pensar que un tiempo de paro corresponde al tiempo de ocurrencia de un cierto evento relacionado con el proceso. En este sentido, la noción de "estar adaptado a la filtración natural" corresponde a decir que para saber si dicho evento ocurrió o no antes del tiempo n basta con la información del proceso hasta ese instante. Es fácil ver que los tiempos de primera visita o de recurrencia son en particular tiempos de paro, ya que para saber si el proceso visitó el estado j antes del tiempo n basta con ver la trayectoria entre los tiempos 0 y n. En cambio, si definimos la variable aleatoria*

$$\xi_{ij} = \sup\{n \geq 0 : X_n = j \text{ dado que } X_0 = i\},$$

este no será tiempo de paro, pues $\{\xi_{ij} \leq n\}$ corresponde a que el proceso no visite el estado j luego del tiempo n, y esa es información "del futuro", que no depende de la trayectoria del proceso hasta el tiempo n.

Los tiempos de paro nos permiten extender la propiedad de Markov a tiempos aleatorios.

Proposición 38. ***(Propiedad Fuerte de Markov).*** *Sea* $(X_n,\ n \in \mathbb{N}_0)$ *una cadena de Markov y* τ *un tiempo de paro con respecto a la cadena. Condicionando al evento* $\{\tau < \infty\}$*, el proceso* $\{X_{\tau+n}\}_{n\in\mathbb{N}}$ *es una cadena de Markov, es decir se cumple que* $\forall\ i, j, x_{n-1}, \ldots, x_0 \in S$*,*

$$\mathbb{P}\left[X_{\tau+n+1} = j | X_{\tau+n} = i, X_{\tau+n-1} = x_{n-1}, \ldots, X_\tau = x_0\right] = \mathbb{P}\left[X_{\tau+n+1} = j | X_{\tau+n} = i\right] = p(i,j).$$

Dem.

$$\begin{aligned}
\mathbb{P}\left[X_{\tau+n+1} = j | X_{\tau+n} = i, X_{\tau+n-1} = x_{n-1}, \ldots, X_\tau = x_0\right] &= \frac{\mathbb{P}\left[X_{\tau+n+1} = j, X_{\tau+n} = i, \ldots, X_\tau = x_0\right]}{\mathbb{P}\left[X_{\tau+n} = i, \ldots, X_\tau = x_0\right]} \\
= \sum_{k=0}^{\infty} \frac{\mathbb{P}\left[\tau = k, X_{\tau+n+1} = j, X_{\tau+n} = i, \ldots, X_\tau = x_0\right]}{\mathbb{P}\left[X_{\tau+n} = i, \ldots, X_\tau = x_0\right]} \frac{\mathbb{P}\left[\tau = k, X_{\tau+n} = i, \ldots, X_\tau = x_0\right]}{\mathbb{P}\left[\tau = k, X_{\tau+n} = i, \ldots, X_\tau = x_0\right]} \\
= \sum_{k=0}^{\infty} \mathbb{P}\left[X_{k+n+1} = j | X_{k+n} = i, X_{k+n-1} = x_{n-1}, \ldots, X_k = x_0, \tau = k\right] \frac{\mathbb{P}\left[\tau = k, X_{\tau+n} = i, \ldots, X_\tau = x_0\right]}{\mathbb{P}\left[X_{\tau+n} = i, \ldots, X_\tau = x_0\right]} \\
= p(i,j) \sum_{k=0}^{\infty} \frac{\mathbb{P}\left[\tau = k, X_{\tau+n} = i, \ldots, X_\tau = x_0\right]}{\mathbb{P}\left[X_{\tau+n} = i, \ldots, X_\tau = x_0\right]} = p(i,j),
\end{aligned}$$

dónde en el segundo renglón usamos probabilidad total y en el cuarto la propiedad de Markov usual, ya que el evento $\{\tau = k\} = \{\tau \le k\} \cap \{\tau \le k-1\}^C \in \mathscr{F}_k$, es decir de lo que pasa con el proceso entre los tiempos 0 y k. □

3.4. Distribuciones Estacionarias

Sabemos que a pesar de considerar cadenas de Markov homogéneas, es decir cuyas probabilidades de transición son las mismas en cada paso del tiempo, la distribución de la cadena en un cierto paso k no tiene por qué ser la misma que en un paso $k+1$. ¿Existirán condiciones bajo las cuales las variables $\{X_k\}_{k\in\mathbb{N}_0}$ están igualmente distribuídas?

Definición 39. *Una* ***distribución estacionaria*** *para una cadena de Markov* $\{X_n\}_{n\in\mathbb{N}}$ *con matriz de transición* P *es un vector* π *de entradas no negativas con* $\sum_{i\in S} \pi_i = 1$ *tal que*

$$\pi P = \pi, \quad \text{es decir} \quad \left[\sum_{i\in S} \pi_i p(i,j) = \pi_j \qquad \forall\, j \in S\right] \qquad \text{(Ecuación de balance general)}$$

Nota. *Es importante notar que podría existir un vector* $\tilde{\pi}$ *de entradas no negativas que satisfaga la Ecuación de balance general con* $\sum_{i\in S} \tilde{\pi}_i \neq 1$*. Dicho vector se conoce como una* ***medida estacionaria*** *(pero no será una distribución de probabilidad). Si* $\sum_{i\in S} \tilde{\pi}_i$ *converge*[6]*, entonces basta normalizar la medida*

[6]en particular esto siempre sucede si el cardinal del espacio de estados S es finito.

para obtener una distribución, es decir tomar $\pi_j = \tilde{\pi}_j / \sum_{i\in S} \tilde{\pi}_i$ *. Si* $\sum_{i\in S} \tilde{\pi}_i$ *diverge, no es posible obtener una distribución a partir de esta medida.*

En el caso en que consideramos una distribución inicial π que es una distribución estacionaria para la cadena, por la ecuación de balance general tenemos que

$$\mathbb{P}\left[X_1 = j\right] = \sum_{i\in S} \mathbb{P}\left[X_1 = j | X_0 = i\right] \mathbb{P}\left[X_0 = i\right] = \sum_{i\in S} p(i,j)\pi_i = \pi_j = \mathbb{P}\left[X_0 = j\right],$$

por lo que podemos deducir que si $X_0 \sim \pi$ entonces $X_1 \sim \pi$ y, recursivamente, $X_n \sim \pi\ \forall n \in \mathbb{N}$.

A partir de la definición anterior es razonable preguntarse: ¿siempre existe una distribución estacionaria asociada a una cadena de Markov? ¿puede haber más de una? Veremos algunos ejemplos, y luego trataremos de caracterizar los distintos casos usando las definiciones anteriores.

Ejemplo 40. Consideremos la caminata aleatoria en $\mathbb{Z}$ con probabilidades de transición dadas por

$$p(i,j) = \begin{cases} p & \text{si } j = i+1 \\ 1-p & \text{si } j = i-1 \\ 0 & \text{cualquier otro caso} \end{cases} \qquad i,j \in \mathbb{Z}.$$

para un $p \in (0,1)$. Si tomamos el vector infinito $\pi_i = 1\ \forall i \in \mathbb{Z}$, podemos probar que satisface la Ecuación de balance general, ya que

$$\sum_{i\in\mathbb{Z}} p(i,j)\pi_i = p(j-1,j)\pi_{j-1} + p(j+1,j)\pi_{j+1} = p\times 1 + (1-p)\times 1 = 1 = \pi_j,$$

por lo que es una medida estacionaria, pero $\sum_{i\in\mathbb{Z}} \pi_i$ diverge, por lo que no puede normalizarse para obtener una distribución estacionaria. Ahora, ¿este vector será la única solución de la ecuación de balance general? No! si tomamos $\tilde{\pi}_i = \left(\frac{p}{1-p}\right)^i$, vemos que

$$\sum_{i\in\mathbb{Z}} p(i,j)\tilde{\pi}_i = p\left(\frac{p}{1-p}\right)^{j-1} + (1-p)\left(\frac{p}{1-p}\right)^{j+1} = \left(\frac{p}{1-p}\right)^{j}.$$

Notemos que $\sum_{i\in\mathbb{Z}} \tilde{\pi}_i = \sum_{i\geq 1} \left(\frac{1-p}{p}\right)^i + 1 + \sum_{i\geq 1} \left(\frac{p}{1-p}\right)^i$ siempre diverge, pues solo una de las dos series puede ser convergente.

Ejemplo 41. En el caso del castillo de naipes, veamos ahora que tiene que cumplir un vector $(\pi_{2k} : k \in \mathbb{N}_0)$ para ser distribución estacionaria. Supongamos que $\sum_{k\in\mathbb{N}_0} \pi_{2k} = 1$. Si $j = 0$, tenemos $p(2k,0) = 1-p$, por lo que si se satisface la EBG tenemos

$$\pi_0 = \sum_{k\in\mathbb{N}_0} p(2k,0)\pi_{2k} = (1-p)\sum_{k\in\mathbb{N}_0} \pi_{2k} = 1-p.$$

Si $j = 2$, tenemos que $p(0,2) = p$, $p(2k,2) = 0$ para todo $k \in \mathbb{N}$, por lo que

$$\pi_2 = \sum_{k \in \mathbb{N}_0} p(2k,2)\pi_{2k} = p\pi_0 = p(1-p),$$

y si seguimos así obtenemos que el vector π dado por $\pi_{2k} = (1-p)p^k$ para cada $k \in \mathbb{N}$ es una distribución estacionaria, y es fácil ver que es el único vector que satisface la ecuación de balance general y cuyas entradas suman uno.

En resumen, la existencia de distribuciones estacionarias depende de la existencia de soluciones para el sistema

$$\begin{cases} \sum_{i \in S} \pi_i p(i,j) = \pi_j & \forall j \in S \\ \sum_{i \in S} \pi_i = 1 \end{cases}$$

Este sistema puede no tener ninguna solución (caminata aleatoria), una única solución (castillo de naipes), o infinitas soluciones (ruina del jugador).

3.4.1. Distribuciones Reversibles

Existe un tipo de distribuciones estacionarias especiales llamadas distribuciones **reversibles**.

Definición 42. *Una distribución π se dice **reversible** respecto a una Cadena de Markov si*

$$\pi_i p(i,j) = \pi_j p(j,i) \qquad \forall i, j \in S \qquad \text{(Ecuación de balance detallado)}$$

Proposición 43. *Una distribución reversible es una distribución estacionaria*

Dem. Para cualquier $j \in S$ tenemos que por reversibilidad

$$\sum_{i \in S} \pi_i p(i,j) = \sum_{i \in S} \pi_j p(j,i) = \pi_j \sum_{i \in S} p(j,i) = \pi_j,$$

por lo que se satisface la Ecuación de Balance General. □

OJO! No toda distribución estacionaria es reversible! Por ejemplo, en el caso del castillo de naipes

$$\pi_0 p(0,4) = 0 \neq (1-p)^2 p^2 = \pi_4 p(4,0)$$

El nombre de *reversible* viene del hecho que si tengo una cadena de Markov con este tipo de distribución inicial, la evolución de la cadena “para adelante en el tiempo” es igual a la evolución de la cadena reversada, es decir “para atrás ene el tiempo", como veremos a continuación.

Proposición 44. *Sea π una distribución inicial estacionaria para la cadena de Markov $(X_n, n \in \mathbb{N}_0)$. Entonces el* ***proceso backward -reversado-*** *$(Y_m, 0 \le m \le n)$, dado por $Y_m := X_{n-m}$, tiene probabilidades de transición dadas por*

$$p^R(i,j) = \frac{\pi_j p(j,i)}{\pi_i} \qquad \forall i,j \in S$$

y distribución inicial π.

Dem. Si $X_0 \sim \pi$, entonces $X_k \sim \pi$ para todo $k \in \mathbb{N}$ por ser estacionaria, y en particular $Y_0 = X_n \sim \pi$. Además, tenemos que

$$p^R(i,j) = \mathbb{P}\left[Y_1 = j | Y_0 = i\right] = \mathbb{P}\left[X_{n-1} = j | X_n = i\right] = \frac{\mathbb{P}\left[X_n = i | X_{n-1} = j\right]\mathbb{P}\left[X_{n-1} = j\right]}{\mathbb{P}\left[X_n = i\right]}$$

por el Teorema de Bayes, y por lo tanto $p^R(i,j) = \frac{p(j,i)\pi_j}{\pi_i}$ □

Corolario 45. *Bajo las hipótesis del teorema anterior, si π es reversible, entonces*

$$P^R = P.$$

Dem. Si π es reversible, por la ecuación de balance detallado tenemos que para todo $i,j \in S$,

$$p^R(i,j) = \frac{p(j,i)\pi_j}{\pi_i} = \frac{p(i,j)\pi_i}{\pi_i} = p(i,j).$$

□

Ejemplo 46 (**Cadena de Nacimiento y Muerte**). Definimos la cadena de Markov $(X_n, n \in \mathbb{N}_0)$ con espacio de estados $S = \mathbb{N}_0$ y probabilidades de transición

$$p(i,j) = \mathbb{P}\left[X_{n+1} = j | X_n = i\right] = \begin{cases} p_i & \text{si } j = i+1, i \in S \\ r_i & \text{si } j = i, i \in S \\ q_i & \text{si } j = i-1, i > 0 \\ 0 & \text{c.o.c} \end{cases}$$

¿Admite distribución reversible? Veamos si existe solución para la Ecuación de balance detallado.

$$\text{para } i = 0, j = 1 \qquad \pi_0 p(0,1) = \pi_1 p(1,0) \Rightarrow \pi_1 = \pi_0 \frac{p_0}{q_1}.$$

$$\text{para } i = 1, j = 2 \qquad \pi_1 p(1,2) = \pi_2 p(2,1) \Rightarrow \pi_2 = \pi_1 \frac{p_1}{q_2} = \pi_0 \frac{p_0 p_1}{q_1 q_2}.$$

$$\dots$$

$$\text{para } i = k, j = k+1 \qquad \pi_k p(k,k+1) = \pi_{k+1} p(k+1,k) \Rightarrow \pi_{k+1} = \pi_k \frac{p_k}{q_{k+1}} = \pi_0 \prod_{i=0}^{k} \frac{p_i}{q_{i+1}}.$$

¿Existe una distribución de probabilidad que cumpla con esto? Debería satisfacer

$$\sum_{k=0}^{\infty} \pi_k = \pi_0 \left[1 + \sum_{k=0}^{\infty} \prod_{i=0}^{k} \frac{p_i}{q_{i+1}} \right] = 1.$$

- Si $0 < \sum_{k=0}^{\infty} \prod_{i=0}^{k} \frac{p_i}{q_{i+1}} < \infty$, entonces basta tomar $\pi_0 = \frac{1}{1+\sum_{k=0}^{\infty} \prod_{i=0}^{k} \frac{p_i}{q_{i+1}}}$
- Si $\sum_{k=0}^{\infty} \prod_{i=0}^{k} \frac{p_i}{q_{i+1}}$ diverge es imposible encontrar π_0 tal que $\sum_{k=0}^{\infty} \pi_k = 1$.
- Si $p_0 = 0$ entonces $\prod_{i=0}^{k} \frac{p_i}{q_{i+1}} = 0$ y por lo tanto $\pi_0 = 1$, $\pi_k = 0$ para todo $k \neq 0$.

Teniendo en cuenta esto, estudiar que pasa si

1. $p_k = r_k = q_k = 1/3$ para todo $k \in \mathbb{N}, p_0 = 1/3$ y $r_0 = 2/3$. ¿Cambia algo si $p_0 = 2/3$ y $r_0 = 1/3$?
2. $p_k = r_k = 1/k, q_k = 1 - 2/k$ para todo $k > 2$, $p_i = r_i = q_i = 1/3$ para $i \in \{1,2\}$, y $p_0 = r_0 = 1/2$.
3. Dados p y q fijos, tomamos $p_k = p^k$, $q_k = q^k$ y $r^k = 1 - p^k - q^k$ para $k \in \mathbb{N}$ y $p_0 = 1$. ¿Para que valores de p y q existe distribución reversible?

3.4.2. Existencia y unicidad de Distribuciones Estacionarias

El teorema ergódico y su corolario 33 nos permiten probar el siguiente resultado

Teorema 47. *Toda cadena de Markov irreducible y recurrente positiva tiene una única distribución estacionaria, dada por*

$$\pi_j = \frac{1}{\mu_j} \qquad \forall j \in S.$$

Dem. Veamos primero que $\{1/\mu_j\}_{j\in S}$ es distribución de probabilidad.

$$\begin{aligned}\sum_{j\in S} \frac{1}{\mu_j} &= \sum_{j\in S} \lim_{n\to\infty} \frac{\sum_{k=1}^{n} p_k(i,j)}{n} = \lim_{n\to\infty} \sum_{j\in S} \frac{\sum_{k=1}^{n} p_k(i,j)}{n} \\ &= \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \sum_{j\in S} p_k(i,j)}{n} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} 1}{n} = 1.\end{aligned}$$

En esta cadena de igualdades podemos intercambiar el límite y la suma en la primera igualdad porque sabemos que el límite existe (por el Corolario 33); y podemos intercambiar las sumas en la segunda igualdad porque son acotadas.

Por otro lado tenemos que

$$\sum_{i\in S}\frac{1}{\mu_i}p(i,j)=\sum_{i\in S}\left[\lim_{n\to\infty}\frac{\sum\limits_{k=1}^{n}p_k(l,i)}{n}\right]p(i,j)=\lim_{n\to\infty}\frac{\sum\limits_{k=1}^{n}\sum\limits_{i\in S}p_k(l,i)p(i,j)}{n}$$
$$=\lim_{n\to\infty}\frac{\sum\limits_{k=1}^{n}p_{k+1}(l,j)}{n}=\frac{1}{\mu_j},$$

donde la última igualdad se debe a la unicidad del límite. Por lo tanto se cumple la Ecuación de balance general y entonces $\{1/\mu_j\}_{j\in S}$ es distribución estacionaria.

Veamos ahora que es única. Supongamos existe otra distribución estacionaria $\{\tilde{\pi}_j\}_{j\in S}$ distinta a π. Por la ecuación de balance general, tenemos que $\tilde{\pi}P^k=\tilde{\pi}$, y por lo tanto

$$\sum_{k=1}^{n}\frac{\sum\limits_{i\in S}\tilde{\pi}_i p_k(i,j)}{n}=\tilde{\pi}_j.$$

Como sabemos que las sumas y los límites existen,

$$\frac{1}{\mu_j}=\sum_{i\in S}\tilde{\pi}_i\frac{1}{\mu_j}=\sum_{i\in S}\tilde{\pi}_i\lim_{n\to\infty}\sum_{k=1}^{n}\frac{p_k(i,j)}{n}=\lim_{n\to\infty}\sum_{k=1}^{n}\frac{\sum\limits_{i\in S}\tilde{\pi}_i p_k(i,j)}{n}=\tilde{\pi}_j,$$

lo que es una contradicción. □

Como consecuencia del teorema anterior y la proposición 35 tenemos el siguiente corolario.

Corolario 48. *Toda cadena de Markov irreducible y finita tiene una única distribución estacionaria.*

3.5. Distribuciones Límites

Definición 49. *Dada una cadena de Markov con matriz de transición P y distribución inicial $\mu^{(0)}$, se llama **distribución límite** de la cadena al vector $\tilde{\mu}=\lim\limits_{n\to\infty}\mu^{(0)}P^n$, es decir*

$$\tilde{\mu}_j=\lim_{n\to\infty}\sum_{i\in S}\mu_i^{(0)}p_n(i,j)\qquad\forall j\in S,$$

siempre que existe y $\sum\limits_{j\in S}\tilde{\mu}_j=1$.

Nota. *Es importante notar que el límite, en principio, depende de la distribución inicial.*

Nota. *Si el límite $\tilde{\mu}$ existe pero $\sum\limits_{j\in S}\tilde{\mu}_j<1$, decimos que hubo "pérdida de masa".*

Es fácil ver que por la misma definición de límite, una distribución límite es en particular una distribución estacionaria, ya que

$$\tilde{\mu}P = \left[\lim_{n\to\infty} \mu^{(0)} P^n\right] P = \lim_{n\to\infty} \mu^{(0)} P^{n+1} = \tilde{\mu}.$$

Bajo esta definición (dependiente de la distribución inicial) una distribución estacionaria π es una distribución límite, ya que para todo $j \in S$

$$\lim_{n\to\infty} \sum_{i\in S} \pi_i p_n(i,j) = \lim_{n\to\infty} \pi_j = \pi_j.$$

Tenemos entonces el siguiente resultado que nos da condiciones para la convergencia de una cadena a una distribución estacionaria.

Teorema 50. *Dada una cadena de Markov irreducible, aperiódica y con distribución estacionaria π, entonces para todo $i, j \in S$*

$$\lim_{n\to\infty} p_n(i,j) = \pi_j.$$

Nota. *Dado $\mu^{(0)}$ cualquier distribución inicial, bajo las hipótesis del teorema anterior tenemos que para cada $j \in S$*

$$\lim_{n\to\infty} \sum_{i\in S} \mu_i^{(0)} p_n(i,j) = \sum_{i\in S} \mu_i^{(0)} \lim_{n\to\infty} p_n(i,j) = \left(\sum_{i\in S} \mu_i^{(0)}\right) \pi_j = \pi_j,$$

por lo que en este caso π es una distribución límite (independiente de la distribución inicial).

Dem.(Teorema 50) Sean $\{X_n\}_{n\in\mathbb{N}}$ una cadena de Markov con matriz de transición P empezando de una distribución $\mu^{(0)}$; y $\{Y_n\}_{n\in\mathbb{N}}$ una cadena de Markov con matriz de transición P empezando de la distribución (estacionaria) π. Definimos entonces la cadena de Markov (bidimensional) $\{Z_n = (X_n, Y_n)\}_{n\in\mathbb{N}}$ con probabilidades de transición

$$p\big((i,j);(k,l)\big) = p(i,k)p(j,l) \qquad \forall\, i,j,k,l \in S.$$

Veamos primero que la aperiodicidad asociada a la matriz P implica que Z es irreducible. Para esto utilizaremos el siguiente resultado de Teoría de Números:

Lema 51. *Dado $A = \{a_1, a_2, \dots\}$ un conjunto de enteros positivos tal que*

1. $\operatorname{mcd} A = 1$*; y*

2. *A es cerrado para la suma, es decir para todo $a, a' \in A$, $a + a' \in A$.*

Entonces existe $N \in \mathbb{N}$ tal que para todo $n \geq N$, $n \in A$.

Dem. Si mcd $A = 1$ tenemos dos opciones:

- Si $[1 \in A]$, tenemos que por propiedad *2.*, $2 = 1 + 1 \in A$; $3 = 2 + 1 \in A$, y así sucesivamente, por lo que $A = \mathbb{N}$.
- Si $\left[\text{mcd}(a, b) = 1 \text{ para algún par } a, b \in A\right]$, tenemos que por la *Identidad de Bézout*

$$\exists c_a, c_b \in \mathbb{Z} \text{ tal que } 1 = c_a a + c_b b,$$

y por lo tanto, si tomamos $N = |c_a|a + |c_b|b \in A$ (por propiedad 2.), se cumple que $N + k = (|c_a| + kc_a)a + (|c_b| + kc_b)b \in A$ para todo $k \in \mathbb{N}$.

$\square$

En nuestro caso para cada estado $i \in S$, $A_i = \left\{k \geq 1 : p_k(i, i) > 0\right\}$ satisface ambas condiciones, por lo que el lema anterior me asegura que existe $N_i \in \mathbb{N}$ tal que $p_n(i, i) > 0$ para todo $n \geq N_i$.

Entonces, para ir del estado (i, j) a (k, l), dado que existe n_{ik} tal que $p_{n_{ik}}(i, k) > 0$ por irreducibilidad de la cadena original, por el lema anterior tenemos que para todo $\tilde{n} > N_i + n_{ik}$ se tiene $p_{\tilde{n}}(i, k) > p_n(i, i)p_{n_{ik}}(i, k) > 0$. Análogamente, existe N_j tal que si $n \geq N_j$ se cumple que $p_n(j, j) > 0$; y n_{jl} tal que $p_{n_{jl}}(j, l) > 0$, por lo que para todo $\tilde{n} > N_j + n_{jl}$ se tiene que $p_{\tilde{n}}(j, l) > p_n(j, j)p_{n_{jl}}(j, l) > 0$. Tomando $\tilde{N} = \text{máx}\{N_i + n_{ik}, N_j + n_{jl}\}$, para todo $n \geq \tilde{N}$ tenemos que $p_n((i, j); (k, l)) = p_n(i, k)p_n(j, l) > 0$. Como puede aplicar el mismo razonamiento para ir de (k, l) a (i, j), tenemos que estos estados están comunicados, y como supusimos estados arbitrarios, tenemos que la cadena Z es irreducible [7].

Ahora, tenemos que Z es irreducible y además tiene una distribucion estacionaria $\bar{\pi}$ dada por

$$\bar{\pi}_{(i,j)} = \pi_i \pi_j \qquad \forall (i, j) \in S \times S.$$

En efecto,

$$\sum_{(i,j) \in S \times S} \bar{\pi}_{(i,j)} p((i, j); (k, l)) = \sum_{i \in S} \sum_{j \in S} \pi_i \pi_j p(i, k) p(j, l) = \sum_{i \in S} \pi_i p(i, k) \sum_{j \in S} \pi_j p(j, l) = \pi_k \pi_l = \bar{\pi}_{(k,l)},$$

y obviamente $\sum_{(i,j) \in S \times S} \bar{\pi}_{(i,j)} = 1$.

Dado que Z es irreducible y tiene distribución estacionaria, debe ser recurrente positiva. De hecho, si es irreducible, y todos son transientes o recurrentes nulos, por el Teorema Ergódico $\frac{1}{\mu_j} = 0$

[7] ¡La hipótesis de aperiodicidad es necesaria! Esto puede verse por ejemplo en el caso de la caminata del borracho, donde si parto de $(1, 2)$ nunca podré llegar a $(1, 1)$

para todo $j \in S$, y por lo tanto no puede existir distribución estacionaria (por un argumento análogo al usado para demostrar unicidad en la prueba del Teorema 47). [8]

Queremos entonces probar que las X y Y se juntan en tiempo finito, es decir, si definimos el tiempo de paro

$$\bar{\tau} = \text{mín}\{n \geq 1 : X_n = Y_n\}$$

esta variable es finita casi seguramente. En efecto, dados $i, j, k \in S$

$$\left\{\tau_{(i,j);(k,k)} < \infty\right\} \subseteq \{\bar{\tau} < \infty\},$$

y como $\mathbb{P}\left[\tau_{(i,j);(k,k)} < \infty\right] = f_{(i,j);(k,k)} = 1$ para todo $i, j, k \in S$ por ser Z recurrente positiva, tenemos que $\mathbb{P}[\bar{\tau} < \infty] = 1$. Como ambas cadenas están construidas conjuntamente (con las mismas realizaciones) y tienen las mismas probabilidades de transición, a partir de ese tiempo aleatorio no se vuelven a separar. Es decir, por la propiedad de Markov fuerte tenemos que para todo $k \in S$

$$\mathbb{P}[X_n = k, \bar{\tau} \leq n] = \mathbb{P}[Y_n = k, \bar{\tau} \leq n],$$

y por lo tanto para cada $n \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}[X_n = k] &= \mathbb{P}[X_n = k, \bar{\tau} \leq n] + \mathbb{P}[X_n = k, \bar{\tau} > n] \\ &= \mathbb{P}[Y_n = k, \bar{\tau} \leq n] + \mathbb{P}[X_n = k, \bar{\tau} > n] \leq \mathbb{P}[Y_n = k] + \mathbb{P}[\bar{\tau} > n]\end{aligned}$$

y análogamente

$$\mathbb{P}[Y_n = k] \leq \mathbb{P}[Y_n = k] + \mathbb{P}[\bar{\tau} > n],$$

por lo que

$$|\mathbb{P}[X_n = k] - \mathbb{P}[Y_n = k]| \leq \mathbb{P}[\bar{\tau} > n] \underset{n\to\infty}{\longrightarrow} 0.$$

Dado un estado $i \in S$ podemos tomar $\mu^{(0)} = \delta_i$, entonces $\mathbb{P}[X_n = k] = p_n(i,k)$ y $\mathbb{P}[Y_n = k] = \pi_k$ por ser π distribución estacionaria. Lo anterior implica que para todo $i, k \in S$

$$\lim_{n\to\infty} |p_n(i,k) - \pi_k| = 0 \quad \Longrightarrow \quad \lim_{n\to\infty} p_n(i,k) = \pi_k.$$

□

Como consecuencia del Teorema 47 y el Teorema 50, tenemos el siguiente resultado

Teorema 52. *Dada una cadena de Markov irreducible, recurrente positiva y aperiódica, entonces existe una única distribución estacionaria π que satisface*

$$\lim_{n\to\infty} p_n(i,j) = \pi_j = \frac{1}{\mu_j} \qquad \forall i, j \in S.$$

[8] Si la cadena no es irreducible puede tener distribuciones estacionarias sin ser recurrente positiva.

3.6. Cadenas de Markov Finitas

3.6.1. Cadenas Regulares

Definición 53. *Una cadena finita se dice* ***regular*** *si existe* $n > 0$ *tal que* $p_n(i, j) > 0$ *para todo* $i, j \in S$.

Proposición 54. *Una cadena finita es regular si y solo si es irreducible y aperiódica.*

Dem.

($\Rightarrow$) Es claro de la definición que si una cadena es regular entonces es irreducible. Veamos ahora que es aperiódica. Dado un estado $i \in S$, debe existir un estado $k \in S$ tal que $p(i,k) > 0$ (pues $\sum_{j\in S} p(i, j) = 1$). Tenemos entonces por regularidad que $p_n(k,k) > 0$ y $p_{n+1}(k,k) \geq p_n(k,i)p(i,k) > 0$, por lo que $d(k)|\text{mcd}\{n, n+1\} = 1$, por lo que el estado k es aperiódico, y por irreducibilidad toda la cadena lo será.

($\Leftarrow$) Si una cadena finita es irreducible y aperiódica, por el Lema 51 tenemos que para todo $i \in S$ existe N_i tal que para todo $n \geq N_i$ se cumple que $p_n(i,i) > 0$. Usando que la cadena es irreducible, basta tomar

$$N = \max_{i\in S} \max_{j\in S}\{N_i + n_{ij}\},$$

que existe pues la cadena es finita, y tendremos que para todo $n > N$ se cumple que $p_n(i, j) > 0$ para todo $i, j \in S$.

$\square$

De la proposición anterior junto con los Teoremas 47 y 50 podemos deducir el siguiente resultado

Proposición 55. *Sea* X *una cadena finita*

i.) Si la cadena es regular, tiene como distribución límite la única distribución estacionaria π *dada por* $\pi_j = 1/\mu_j$, $j \in S$.

ii.) Si la cadena es regular y doblemente estocástica, es decir $\sum_{i\in S} p(i, j) = 1$, *entonces la distribución límite es la distribución uniforme, correspondiente a* $\pi_j = 1/|S|$ *para todo* $j \in S$.

iii.) Si la cadena es irreducible, aperiódica y doblemente estocástica, entonces la unica distribución límite es la uniforme.

Demostración. ii.) Como la cadena es regular, existe una única distribución estacionaria dada por $\pi_j = 1/\mu_j$, con $\sum_{j\in S} 1/\mu_j = 1$. Además, por ser doblemente estocástica, por inducción tenemos que

$$\sum_{i\in S} p_k(i,j) = 1 \quad \Rightarrow \quad \sum_{i\in S} \frac{\sum_{k=1}^{n} p_k(i,j)}{n} = 1.$$

Tomando límites cuando n tiende a infinito, por el Teorema ergódico tenemos que para todo $j \in S$,

$$\sum_{i\in S} \pi_j = 1 \Rightarrow |S|\pi_j = 1 \Rightarrow \pi_j = 1/|S|.$$

□

3.6.2. Distribuciones de Salida

Dada una cadena de Markov a estados finitos no irreducible, muchas veces tenemos estados absorbentes. ¿Cómo podemos estudiar el tiempo que tarda en absorberse la cadena?¿Cuál es la probabilidad de que termine en un estado determinado? A continuación veremos algunos ejemplos.

Ejemplo 56. Supongamos que en una cierta preparatoria, se sabe que 60 % de los alumnos de primer año pasan a segundo, 20 % de ellos repiten y el resto abandona. De los de segundo año, 70 % pasan a tercer año, 20 % repiten y el resto abandona; y en el último año 80 % termina, el 10 % repite y el resto abandona. ¿Que fracción de los alumnos que ingresan se recibe en algún momento?

Modelaremos este problema como una cadena de Markov $(X_n)_{n\in\mathbb{N}}$con espacio de estados $S = \{1,2,3,G,A\}$, donde G corresponde a graduados y A a los que abandonaron. Entonces, tenemos que las transiciones de mi cadena estan dadas por la matriz

$$P = \begin{pmatrix} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{G} & \mathbf{A} \\ \mathbf{1} & 0{,}2 & 0{,}6 & 0 & 0 & 0{,}2 \\ \mathbf{2} & 0 & 0{,}2 & 0{,}7 & 0 & 0{,}1 \\ \mathbf{3} & 0 & 0 & 0{,}1 & 0{,}8 & 0{,}1 \\ \mathbf{G} & 0 & 0 & 0 & 1 & 0 \\ \mathbf{A} & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Es claro que G y A son estados absorbentes. Dado $z \in S$, definimos $V_z = \text{mín}\{n \geq 1 : X_n = z\}$. Si tomamos $\tau = \text{mín}\{V_G, V_A\}$, tenemos que $\mathbb{P}[\tau < \infty] = 1$. Denotemos por $g_G(z)$ la probabilidad de graduarse estando en el estado z, es decir $g_G(z) = \mathbb{P}[V_G < V_A | X_0 = z]$. Es claro que $g_G(G) = 1$ y $g_G(A) = 0$. Veamos que pasa con los otros casos. Por teorema de probabilidad total y propiedad de Markov,

$$g_G(z) = \mathbb{P}[V_G < V_A | X_0 = z] = \sum_{y\in S} \frac{\mathbb{P}\left[V_G < V_A, X_0 = z, X_1 = y\right]}{\mathbb{P}[X_0 = z]}$$

$$= \sum_{y\in S} \mathbb{P}\left[V_G < V_A | X_1 = y, X_0 = z\right] \mathbb{P}\left[X_1 = y | X_0 = z\right] = \sum_{y\in S} g_G(y) p(z,y).$$

Teniendo en cuenta la matriz de transición, tenemos entonces el sistema de ecuaciones

$$g_G(1) = g_G(1)p(1,1) + g_G(2)p(1,2) = 0{,}2g_G(1) + 0{,}6g_G(2)$$

$$g_G(2) = g_G(2)p(2,2) + g_G(3)p(2,3) = 0{,}2g_G(2) + 0{,}7g_G(3)$$

$$g_G(3) = g_G(3)p(3,3) + g_G(G)p(3,G) = 0{,}1g_G(3) + 0{,}8.$$

Resolviendo el sistema tenemos que $g_G(3) = 8/9$, $g_G(2) = 7/9$ y $g_G(1) = 7/12$, de donde podemos deducir que de los alumnos que comienzan, alrededor de un 58 % se graduará de la prepa y un 42 % abandonará.

Ejemplo 57. En el ejemplo 4 correspondiente al apostador, vemos que la cadena no es irreducible ya que tiene dos estados absorbentes, 0 y N. Tenemos entonces que $\tau = V_0 \wedge V_N$ es un tiempo de paro finito casi seguramente. Podemos calcular así la probabilidad de retirarse con un capital N empezando de un estado $k \in \{0,1,2,\dots,N\}$, que denotamos por $g(k)$. Es claro que $g(0) = \mathbb{P}[V_N < V_0 | X_0 = 0] = 0$, $g(N) = \mathbb{P}[V_N < V_0 | X_0 = N] = 1$; y tenemos el sistema de ecuaciones

$$g(1) = g(2)p(1,2) + g(0)p(1,0) = 0{,}5g(2)$$

$$g(2) = g(1)p(2,1) + g(3)p(2,3) = 0{,}5g(1) + 0{,}5h(3)$$

$$\dots\dots$$

$$g(N-2) = g(N-3)p(N-2,N-3) + g(N-1)p(N-2,N-1) = 0{,}5g(N-3) + 0{,}5h(N-1)$$

$$g(N-1) = g(N-2)p(N-1,N-2) + g(N)p(N-1,N) = 0{,}5g(N-2) + 0{,}5.$$

Este sistema tiene como solución $g(k) = k/N$ para todo $k \in \{0,1,2,\dots,N\}$, lo que quiere decir que si empiezo a jugar con un capital k, la probabilidad de que llegue a reunir un capital N antes de perder todo mi dinero es k/N.

3.6.3. Tiempos de Salida

En los ejemplos anteriores calculamos las probabilidades de terminar en cierto estado absorbente. Ahora ¿cuánto tiempo (en media) tardará mi cadena en absorberse?

Supongamos que en el ejemplo de la preparatoria queremos saber cuál es el tiempo medio que pasa un estudiante en la prepa a partir de un estado x. Queremos calcular entonces la siguiente función

$$\begin{aligned} h(x) &= \text{"tiempo medio que le resta a un estudiante en la preparatoria que se encuentra en el estado } x\text{"} \\ &= E\left[\tau | X_0 = x\right]. \end{aligned}$$

Es claro que $h(A) = h(G) = 0$, y

$$\begin{aligned} h(1) &= \sum_{k\in\mathbb{N}_0} kP[\tau = k|X_0 = 1] = \sum_{k\in\mathbb{N}_0}\sum_{y\in S} kP[\tau = k|X_1 = y, X_0 = 1]P[X_1 = y|X_0 = 1] \\ &= \sum_{k\in\mathbb{N}_0}\sum_{y\in S} kP[1+\tau_{X_1} = k|X_1 = y, X_0 = 1]p(1,y) = \sum_{y\in S}\left(\sum_{k\in\mathbb{N}_0} kP[\tau_{X_1} = k-1|X_1 = y]\right)p(1,y) \\ &= \sum_{y\in S}\big(1+h(y)\big)p(1,y) = (1+h(1))p(1,1) + (1+h(2))p(1,2) + (1+h(A))p(1,A) \\ &= 1 + 0{,}2h(1) + 0{,}6h(2), \end{aligned}$$

donde usamos la propiedad de Markov, ya que

$$\begin{aligned} \sum_{k\in\mathbb{N}_0} kP[\tau_{X_1} = k-1|X_1 = y] &= \sum_{k\in\mathbb{N}_0} kP[\tau = k-1|X_0 = y] = \sum_{v\in\mathbb{N}_0}(v+1)P[\tau = v|X_0 = y] \\ &= \sum_{v\in\mathbb{N}_0} vP[\tau = v|X_0 = y] + \sum_{v\in\mathbb{N}_0} P[\tau = v|X_0 = y] \\ &= E\left[\tau|X_0 = y\right] + 1 = h(y) + 1. \end{aligned}$$

De la misma manera tenemos que

$$h(2) = 1 + 0{,}2h(2) + 0{,}7h(3) \qquad \text{y} \qquad h(3) = 1 + 0{,}1h(3),$$

y podemos deducir que el tiempo medio que se quedarán en la prepa los que se encuentran en el tercer año será de $1.\widehat{11}$ $(10/9)$ años; los que se encuentran en el segundo año se quedarán $2.\widehat{22}$ $(20/9)$ años; y los que se encuentran en el primero se quedarán una media de $2{,}9\widehat{16}$ $(35/12)$ años. [9]

Tenemos entonces que se cumple la siguiente igualdad

$$h(i) = E_i[\tau] = E_i[V_A \mathbb{1}_{\{V_A < V_G\}}] + E_i[V_G \mathbb{1}_{\{V_G < V_A\}}]$$

Si queremos calcular $h_A(i) := E_i[V_A \mathbb{1}_{\{V_A < V_G\}}]$, podemos pensar en el mismo procedimiento que

[9] notar que estamos teniendo en cuenta también a los que abandonan antes de terminar.

antes, teniendo en cuenta que $h_A(A) = h_A(G) = 0$. Tenemos entonces que

$$\begin{aligned}
h_A(i) &= \sum_{k=0}^{\infty} kP\left[V_A \mathbb{1}_{\{V_A<V_G\}} = k|X_0 = i\right] = \sum_{k=1}^{\infty} kP\left[V_A = k, V_A < V_G|X_0 = i\right] \\
&= \sum_{y\in S}\sum_{k=1}^{\infty} kP\left[V_A = k, V_A < V_G|X_1 = y, X_0 = i\right] P[X_1 = y|X_0 = i] \\
&= \sum_{y\in S/\{G\}}\sum_{k=1}^{\infty} kP\left[V_A = k-1, V_A < V_G|X_0 = y\right] p(i,y) \\
&= \sum_{y\in S/\{G\}}\sum_{r=0}^{\infty} (r+1)P\left[V_A = r, V_A < V_G|X_0 = y\right] p(i,y) \\
&= \sum_{y\in S/\{G\}} h_A(y)p(i,y) + \sum_{y\in S/\{G\}} P\left[V_A < V_G|X_0 = y\right] p(i,y) \\
&= \sum_{y\in S/\{G\}} h_A(y)p(i,y) + \sum_{y\in S} g_A(y)p(i,y) \\
&= \sum_{y\in S/\{G\}} h_A(y)p(i,y) + g_A(i)
\end{aligned}$$

lo que da como resultado el sistema de ecuaciones

$$\begin{cases}
h_A(1) & = h_A(1)0{,}2 + h_A(2)0{,}6 + 5/12 \\
h_A(2) & = h_A(2)0{,}2 + h_A(3)0{,}7 + 2/9 \\
h_A(3) & = h_A(3)0{,}1 + 1/9
\end{cases}$$

de donde $h_A(3) = 0{,}123(10/81)$, $h_A(2) = 0{,}386(125/324)$ y $h_A(1) = 0{,}81(175/216)$. Análogamente, tenemos que $h_G(G) = h_G(A) = 0$ y

$$h_G(i) := E_i[V_G \mathbb{1}_{\{V_G<V_A\}}] = \sum_{y\in S/\{A\}} h_G(y)p(i,y) + g_i(G),$$

por lo que tenemos el sistema de ecuaciones

$$\begin{cases}
h_G(1) & = h_G(1)0{,}2 + h_G(2)0{,}6 + 7/12 \\
h_G(2) & = h_G(2)0{,}2 + h_G(3)0{,}7 + 7/9 \\
h_G(3) & = h_G(3)0{,}1 + 8/9
\end{cases}$$

que resulta en $h_G(3) = 0{,}988(80/81)$, $h_G(2) = 1{,}836(595/324)$ y $h_G(1) = 2{,}106(455/216)$.

4. Proceso de Poisson

Uno de los procesos estocásticos a tiempo continuo más importantes es el proceso de Poisson, y puede definirse de varias maneras alternativas. Una de sus posibles definiciones es la siguiente:

Definición 58. ***Un proceso de Poisson*** *$N = (N_t)_{t\geq 0}$ es un proceso càdlàg no decreciente tal que*

(i) $N_t \in \mathbb{N}_0$, y $N_0 = 0$ c.s.

(ii) N tiene incrementos estacionarios e independientes.

(iii) las trayectorias $(N_t(\omega))_{t\geq 0}$ tienen discontinuidades de tamaño 1 c.s.

Nota. *De la definición anterior podemos deducir algunas cosas. La primera es que el proceso de Poisson es un proceso a tiempo continuo constante por pedazos con saltos de tamaño 1, y tiene a lo más una cantidad finita de saltos en un intervalo acotado, ya que por definición no puede "explotar" (tomar un valor infinito) en tiempo finito. Además, el último ítem me dice que $\lim_{h\to 0} P[N_{t+h} - N_t \geq 2] = 0$, y por lo tanto en cada instante solo podemos quedarnos en donde estamos o dar un salto positivo de tamaño 1. Como los incrementos son estacionarios, podemos aproximar las probabilidades de transición de este proceso en cualquier instante $t \in \mathbb{R}_0^+$ por*

$$P\left[N_{t+h} = j | N_t = i\right] = P\left[N_{t+h} - N_t = j - i\right] = \begin{cases} \lambda h + o_1(h) & \text{si } j = i+1 \\ 1 - \lambda h + o_2(h) & \text{si } j = i \\ o_3(h) & \text{si } j > i+1 \end{cases},$$

dónde λ corresponde a la ***tasa de salto*** *en cada instante de tiempo y las funciones $o_j(h)$ satisface que $\lim_{h\to 0} \frac{o_j(h)}{h} = 0$ para $j = 1,2,3$.*

¿Porque se llama proceso de Poisson? Porque sus distribuciones marginales corresponden a distribuciones de Poisson, como probaremos en el siguiente resultado.

Proposición 59. *Dado un proceso de Poisson N, existe un parámetro λ tal que para cada tiempo $t \geq 0$ se cumple que $N_t \sim \mathcal{P}(\lambda t)$.*

Dem. Queremos demostrar que

$$p_k(t) := P[N_t = k] = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \qquad k \in \mathbb{N}_0.$$

Para esto usaremos nuestra aproximación a las probabilidades de transición, definiendo una ecuación diferencial (conocida como *ecuación forward*) y calculando así la función $(p_k(t), k \geq 0)$. Primero que nada, si $k = 0$, tenemos que

$$\begin{aligned} P[N_{t+h} = 0] &= \sum_{i \in \mathbb{N}_0} P[N_{t+h} = 0 | N_t = i] P[N_t = i] = P[N_{t+h} = 0 | N_t = 0] P[N_t = 0] \\ &= [1 - \lambda h - o(h)] P[N_t = 0], \end{aligned}$$

ya que por ser el proceso no decreciente y $N_0 \equiv 0$ tenemos que $P[N_{t+h} = 0 | N_t = i] = 0$ para todo $i \in \mathbb{N}$. Por lo tanto

$$p_0'(t) = \lim_{h \to 0} \frac{P[N_{t+h} = 0] - P[N_t = 0]}{h} = \lim_{h \to 0} \frac{-\lambda h - o(h)}{h} P[N_t = 0] = -\lambda p_0(t).$$

Si $k \geq 1$, tenemos que

$$\begin{aligned} P[N_{t+h} = k] &= \sum_{i \in \mathbb{N}_0} P[N_{t+h} = k | N_t = i] P[N_t = i] \\ &= [\lambda h + o(h)] P[N_t = k-1] + [1 - \lambda h - o(h)] P[N_t = k] + \sum_{i \neq k, k-1} o(h) P[N_t = i] \end{aligned},$$

de dónde

$$\begin{aligned} p_k'(t) &= \lim_{h \to 0} \frac{P[N_{t+h} = k] - P[N_t = k]}{h} \\ &= \lim_{h \to 0} \frac{\lambda h + o(h)}{h} P[N_t = k-1] - \lim_{h \to 0} \frac{\lambda h + o(h)}{h} P[N_t = k] + \sum_{i \neq k, k-1} \lim_{h \to 0} \frac{o(h)}{h} P[N_t = i] \\ &= \lambda \left[p_{k-1}(t) - p_k(t) \right]. \end{aligned}$$

Entonces tenemos el sistema de ecuaciones

$$\begin{cases} p_0'(t) &= -\lambda p_0(t) \\ p_k'(t) &= \lambda \left[p_{k-1}(t) - p_k(t) \right], \quad k \geq 1 \end{cases}$$

La primera ecuación tiene solución $p_0(t) = Ce^{-\lambda t}$ para alguna constante C, que por la condición inicial $p_0(0) = P[N_0 = 0] = 1$ sabemos que debe ser $C = 1$. Entonces, para $k = 1$ tenemos que $p_1(t)$ debe satisfacer

$$p_1'(t) + \lambda p_1(t) = \lambda e^{-\lambda t},$$

y utilizando técnicas estándar de factor integrante para resolver ecuaciones diferenciales no homogéneas tenemos que

$$p_1(t) = e^{-\lambda t} \left[\int_0^t \lambda e^{-\lambda z} e^{\lambda z} dz \right] = e^{-\lambda t} \lambda.$$

Por inducción podemos probar que

$$p_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

para cualquier $k \in \mathbb{N}$, que es lo deseado. □

La proposición anterior nos dice que podemos dar la siguiente definición alternativa

Definición 60 (Definición Alternativa)**.** *Un proceso de Poisson* $(N_t, t \geq 0)$ *de parámetro* λ *es un proceso càdlàg con incrementos estacionarios e independientes tal que* $N_t \sim \mathscr{P}(\lambda t)$.

Es claro que de esta definición podemos deducir la definición 58, ya que en este caso

$$P\left[N_{t+h} = j | N_t = i\right] = P\left[N_{t+h} - N_t = j - i\right] = \begin{cases} \lambda h e^{-\lambda h} = \lambda h \sum_{k=0}^{\infty} \frac{(-\lambda h)^k}{k!} = \lambda h + o_1(h) & \text{si } j = i+1 \\ e^{-\lambda h} = \sum_{k=0}^{\infty} \frac{(-\lambda h)^k}{k!} = 1 - \lambda h + o_2(h) & \text{si } j = i \\ 1 - e^{-\lambda h} - \lambda h e^{-\lambda h} = o_3(h) & \text{si } j > i+1 \end{cases},$$

Nota. *Teniendo en cuenta esta definición, podemos pensar en muchas aplicaciones de un proceso de Poisson : las llamadas que recibe el Centro de Atención de cierta empresa, la cantidad de personas que llegan a una sucursal de banco, los accidentes de carro, los desastres naturales... si bien la suposición de "incrementos estacionarios e independientes" no parece realista, es muy útil en muchos ejemplos para obtener información de la evolución de los procesos.*

4.1. Tiempos de llegada y tiempos entre llegadas.

Podemos caracterizar también los procesos de Poisson a través de los tiempos entre llegadas sucesivas. Definamos los tiempos de llegada de un proceso de Poisson como

$$S_n := \text{ínf}\{t \geq 0 : N_t = n\}, \quad S_0 = 0,$$

y a partir de allí denotamos $T_n := S_n - S_{n-1}$ las variables de tiempos entre llegadas (o *tiempos interarribos*). Claramente $S_n = \sum_{i=1}^{n} T_i$, y podemos reconstruir el proceso $(N_t, t \geq 0)$ definiendo

$$N_t = \sup\{n \geq 0 : S_n \leq t\},$$

es decir el índice del último tiempo de llegada antes que t.

Es importante notar la siguiente igualdad de eventos, que es directa de las definiciones anteriores

$$\{N_t \geq n\} = \{S_n \leq t\} = \left\{\sum_{i=1}^{n} T_i \leq t\right\}, \qquad \forall t \geq 0,\ n \in \mathbb{N}_0. \tag{5}$$

A partir de la definición original y de la Prop.59 (o alternativamente a partir de la Def. 60) podemos probar entonces el siguiente resultado.

Proposición 61. *Las variables aleatorias $\{T_i\}_{i\in\mathbb{N}}$ correspondientes a los tiempos interarribos de un proceso de Poisson de parámetro λ son i.i.d., con $T_i \sim Exp(\lambda)$.*

Dem. Por la proposición 59 tenemos que para todo $t \geq 0$, $N_t \sim \mathscr{P}(\lambda t)$. Entonces, usando la igualdad de eventos anterior tenemos que

$$1 - F_{T_1}(t) = P[T_1 > t] = P[S_1 > t] = P[N_t < 1] = P[N_t = 0] = e^{-\lambda t},$$

por lo que $T_1 \sim Exp(\lambda)$.

Por otro lado, tenemos que para todo par $s, t \geq 0$

$$\begin{aligned} P[T_1 > s, T_2 > t] &= \int_s^\infty P[T_2 > t | T_1 = u] f_{T_1}(u) du = \int_s^\infty P[S_2 > t + u | S_1 = u] \lambda e^{-\lambda u} du \\ &= \int_s^\infty P[N_{t+u} < 2 | N_u = 1] \lambda e^{-\lambda u} du = \int_s^\infty P[N_{t+u} - N_u < 1 | N_u = 1] \lambda e^{-\lambda u} du \\ &= \int_s^\infty P[N_{t+u} - N_u < 1] \lambda e^{-\lambda u} du = e^{-\lambda t} e^{-\lambda s} \end{aligned}$$

por incrementos estacionarios e independientes, y como $P[T_2 > t | T_1 > s] = e^{-\lambda t}$ no depende de s, podemos deducir que $T_1 \perp\!\!\!\perp T_2$ y $T_2 \sim Exp(\lambda)$. Trabajando por inducción de manera análoga tenemos el resultado deseado. □

Recordando que la distribución de la suma de variables i.i.d exponenciales corresponde a una distribución Gamma, tenemos el siguiente corolario

Corolario 62. *La variable aleatoria S_n correspondiente al tiempo de la n-ésima ocurrencia en un proceso de Poisson de parámetro λ tiene distribución $Gamma(\lambda, n)$.*

La caracterización a partir de los tiempos interarribo nos permite dar una tercera definición del proceso de Poisson.

Definición 63 (Definición Alternativa)**.** *Dada una sucesión $\{T_i\}_{i\in\mathbb{N}}$ de variables aleatorias i.i.d. con distribución exponencial de parámetro λ, el proceso de Poisson de parámetro λ es el proceso $(N_t, t \geq 0)$ dado por*

$$N_t := \sup\{n \geq 1 : \sum_{i=1}^n T_i \leq t\}.$$

Nota. *De lo anterior se deduce que la Def. 58 implica la Def. 60 (como propiedad); y ésta a su vez implica la Def. 63. Para probar la equivalencia de las definiciones bastaría probar que la Def. 63 implica la Def. 58 (como propiedad). Se deja esta prueba como ejercicio para el lector.*

Usando la equivalencia de eventos (5), podemos probar el siguiente resultado.

Proposición 64. *Dado un tiempo $t > 0$, el tiempo entre la última ocurrencia antes de t y la primera ocurrencia después de t tiene distribución $Gamma(\lambda, 2)$.*

Demostración. Queremos estudiar entonces la distribución de $S_{N_t+1} - S_{N_t}$, y para eso calculemos $S_{N_t+1} - t$ y $t - S_{N_t}$. Es fácil deducir que

$$P\left[S_{N_t+1} - t > s\right] = P\left[S_{N_t+1} > s + t\right] = P\left[N_{s+t} < N_t + 1\right] = P\left[N_{s+t} - N_t < 1\right] = e^{-\lambda s},$$

por lo que $(S_{N_t+1} - t) \sim Exp(\lambda)$. Por otro lado,

$$P\left[t - S_{N_t} > s\right] = P\left[S_{N_t} < t - s\right] = P\left[N_{t-s} \geq N_t\right] = P\left[N_{t-s} - N_t = 0\right] = e^{-\lambda s},$$

ya que $\{N_{t-s} \geq N_t\} = \{N_{t-s} - N_t = 0\}$ por ser el proceso no decreciente. Entonces

$$S_{N_t+1} - S_{N_t} = (S_{N_t+1} - t) + (t - S_{N_t}) \sim Gamma(2, \lambda),$$

ya que $(S_{N_t+1} - t)$ depende de $\{N_{s+t} - N_t : s \geq 0\}$ y $(t - S_{N_t})$ depende de $\{N_t - N_{t-s} : 0 \leq s \leq t\}$, por lo que $(S_{N_t+1} - t) \perp\!\!\!\perp (t - S_{N_t})$. $\square$

Nota. *El resultado anterior puede parecer contra-intuitivo, porque estamos hablando de un intervalo interarribo que es el "doble" de un intervalo interarribo típico, pero esto es por que no estoy mirando hacia el futuro, sino desde cierta información intermedia, ya que $S_{N_t+1} - S_{N_t} \not\perp\!\!\!\perp N_t$.*

Probaremos a continuación el siguiente resultado: *Si sabemos que en el intervalo $[0, t]$ hubo n llegadas, entonces la cantidad de llegadas en el intervalo $[0, s]$ con $s < t$ tiene distribución Binomial con probabilidad de éxito ("que una llegada ocurrida antes del tiempo t haya sucedido antes del tiempo s") la proporción de tamaño entre los intervalos.*

Proposición 65. *Dados $s, t \in \mathbb{R}_+$, con $s < t$, entonces $N_s|_{N_t=n} \sim Bi(n, s/t)$ para todo $n \in \mathbb{N}$.*

Demostración. Tenemos entonces que para todo $k \in \{0, 1, \ldots, n\}$

$$\begin{aligned}
P\left[N_s = k | N_t = n\right] &= \frac{P\left[N_s = k, N_t = n\right]}{P\left[N_t = n\right]} = \frac{P\left[N_s = k, N_t - N_s = n - k\right]}{P\left[N_t = n\right]} \\
&= \frac{P\left[N_s = k\right] P\left[N_t - N_s = n - k\right]}{P\left[N_t = n\right]} = \frac{\frac{(\lambda s)^k}{k!} e^{-\lambda s} \frac{(\lambda(t-s))^{n-k}}{(n-k)!} e^{-\lambda(t-s)}}{\frac{(\lambda t)^n}{n!} e^{-\lambda t}} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{s}{t}\right)^k \left(\frac{t-s}{t}\right)^{n-k},
\end{aligned}$$

por lo que tenemos la distribución deseada. $\square$

Otra propiedad que es de importancia para simular procesos de Poisson es la siguiente: *Si sabemos cuantas llegadas hubo en un intervalo* $[0,t]$, *puedo obtener los tiempos correspondientes a estas llegadas a partir una muestra de n vs.as. i.i.d. uniformes en* $[0,t]$ *simplemente ordenándola.*

Proposición 66. *Dado el evento* $\{N_t = n\}$, *tenemos que*

$$(S_1, S_2, \ldots, S_n) \stackrel{\mathscr{D}}{=} (Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}),$$

dónde $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ *corresponde al vector de estadísticos de orden de una muestra* $(Y_1, Y_2, \ldots, Y_n)$ *de vs. as. i.i.d. con* $Y_i \sim \mathscr{U}([0,t])$, *es decir*

$$f_{(S_1,S_2,\ldots,S_n)|_{N_t=n}}(s_1, s_2, \ldots, s_n) = \begin{cases} \frac{n!}{t^n} & \text{si } 0 < s_1 < \ldots < s_n < t, \\ 0 & \text{cualquier otro caso.} \end{cases}$$

Demostración. Calcularemos la FDA de $(S_1, S_2, \ldots, S_n)|_{N_t=n}$ y luego derivaremos. Si $0 < t_1 < \ldots < t_n < t$,

$$\begin{aligned} P[S_1 \leq t_1, \ldots, S_n \leq t_n | N_t = n] =& P\left[N_{t_1} \geq 1, \ldots, N_{t_n} \geq n | N_t = n\right] \\ =& \frac{P\left[N_{t_1} \geq 1, \ldots, N_{t_n} \geq n, N_t = n\right]}{P[N_t = n]} \\ = \sum_{(j_1,\ldots,j_n):\sum j_i = n} & \frac{P\left[N_{t_1} = j_1, N_{t_2} - N_{t_1} = j_2, \ldots, N_{t_{n-1}} - N_{t_n} = j_n, N_t - N_{t_n} = 0\right]}{P[N_t = n]} \\ = \sum_{(j_1,\ldots,j_n):\sum j_i = n} & \frac{P\left[N_{t_1} = j_1\right] P\left[N_{t_2} - N_{t_1} = j_2\right] \ldots P\left[N_{t_{n-1}} - N_{t_n} = j_n\right] P\left[N_t - N_{t_n} = 0\right]}{P[N_t = n]} \\ = \sum_{(j_1,\ldots,j_n):\sum j_i = n} & \frac{\frac{(\lambda t_1)^{j_1}}{(j_1)!} \frac{(\lambda(t_2 - t_1))^{j_2}}{(j_2)!} \ldots \frac{(\lambda(t_n - t_{n-1}))^{j_n}}{(j_n)!}}{\frac{(\lambda t)^n}{n!}}. \end{aligned}$$

Si derivamos la expresión anterior, tenemos que

$$\begin{aligned} &\frac{\partial^n}{\partial t_1 \ldots \partial t_n} \sum_{(j_1,\ldots,j_n):\sum j_i = n} \frac{P\left[N_{t_1} = j_1\right] P\left[N_{t_2} - N_{t_1} = j_2\right] \ldots P\left[N_{t_{n-1}} - N_{t_n} = j_n\right] P\left[N_t - N_{t_n} = 0\right]}{P[N_t = n]} \\ &= \frac{\partial^n}{\partial t_1 \ldots \partial t_n} \frac{P\left[N_{t_1} = 1\right] P\left[N_{t_2} - N_{t_1} = 1\right] \ldots P\left[N_{t_{n-1}} - N_{t_n} = 1\right] P\left[N_t - N_{t_n} = 0\right]}{P[N_t = n]}, \end{aligned}$$

ya que en cualquier otro caso habrá un término del tipo $P\left[N_{t_{k+1}} - N_{t_k} = 0\right]$ para algún k, de donde la variable k no aparecerá cuando quiera derivarla, haciendo que todo el producto sea cero. Así, tenemos que

$$\begin{aligned} \frac{\partial^n}{\partial t_1 \ldots \partial t_n} P[S_1 \leq t_1, \ldots, S_n \leq t_n | N_t = n] &= \frac{\partial^n}{\partial t_1 \ldots \partial t_n} \frac{(\lambda t_1)(\lambda(t_2 - t_1)) \ldots (\lambda(t_n - t_{n-1})}{\frac{(\lambda t)^n}{n!}} \\ &= \frac{n!}{t^n} \frac{\partial^n}{\partial t_1 \ldots \partial t_n} (t_n - t_{n-1}) \ldots (t_2 - t_1) t_1 = \frac{n!}{t^n} \end{aligned}$$

□

4.2. Superposición (o Suma) de procesos de Poisson independientes

Recordemos que dadas dos variables $X \sim \mathscr{P}(\lambda_1)$, $Y \sim \mathscr{P}(\lambda_2)$ con $X \perp\!\!\!\perp Y$, entonces $X+Y \sim \mathscr{P}(\lambda_1 + \lambda_2)$.

Es natural entonces, dados N^1 y N^2 procesos de Poisson independientes de parámetro λ_1 y λ_2 respectivamente, definir el proceso N dado por la suma de N^1 y N^2 como

$$N_t = N_t^1 + N_t^2 \qquad \forall\, t \geq 0. \tag{6}$$

Es inmediato deducir que cada $t \geq 0$, como $N_t^1 \sim \mathscr{P}(\lambda_1 t) \perp\!\!\!\perp N_t^2 \sim \mathscr{P}(\lambda_2 t)$, entonces $N_t \sim \mathscr{P}((\lambda_1 + \lambda_2)t)$.

Nota. *Una forma alternativa de definirlo es la siguiente: dados $\{S_n^1\}_{n\in\mathbb{N}}$ y $\{S_n^2\}_{n\in\mathbb{N}}$ los tiempos de llegada de N^1 y N^2 respectivamente, tomamos*

$$\{S_n\}_{n\in\mathbb{N}} = \{S_k^1\}_{k\in\mathbb{N}} \cap \{S_j^2\}_{j\in\mathbb{N}},$$

ordenados de menor a mayor, es decir la superposición de las llegadas de N^1 y N^2. Tenemos entonces

$$N_t = \text{máx}\{n \geq 0 : S_n \leq t\}. \tag{7}$$

Claramente

$$P[T_1 > t] = P[S_1 > t] = P\left[S_1^1 > t, S_1^2 > t\right] = P\left[N_t^1 < 1, N_t^2 < 1\right] = P\left[N_t^1 < 1\right] P\left[N_t^2 < 1\right] = e^{-(\lambda_1+\lambda_2)t},$$

Por lo que el tiempo de la primera llegada $T_1 \sim Exp(\lambda_1 + \lambda_2)$ y podemos probar de manera análoga que $\{T_i\}_{i\in\mathbb{N}}$ es una sucesión de variables i.i.d. Es importante notar que $T_1 = \text{mín}\{T_1^1, T_1^2\}$ de donde podemos deducir también su distribución.

Proposición 67.

(i) $\{S_n^1\}_{n\in\mathbb{N}}$ *y* $\{S_n^2\}_{n\in\mathbb{N}}$ *no tienen puntos en común, es decir*

$$P\left[\text{ exista un evento múltiple en } (0,a]\right] = 0 \qquad \forall\, a \geq 0.$$

(ii) $(N_t, t \geq 0)$ *es un proceso de Poisson de parámetro $\lambda_1 + \lambda_2$.*

Dem.

(i) Probaremos el caso $a=1$, ya que la demostración es análoga para cualquier otro intervalo. Para eso, utilizaremos las particiones diádicas del intervalo $(0,1]$, de la forma

$$(0,1]=\bigcup_{i=1}^{2^k-1}\left(\frac{i}{2^k},\frac{i+1}{2^k}\right].$$

Para cada $k\in\mathbb{N}$, sea $C_k=\left\{N_{\frac{i+1}{2^k}}-N_{\frac{i}{2^k}}\geq 2 \text{ para algún } i\in\{1,2,\ldots,2^k-1\}\right\}$. Tenemos entonces que

$$\begin{aligned}P\left[\text{ exista un evento múltiple en } (0,1]\right]&=P\left[\forall n\in\mathbb{N},\exists i\in\{1,2,\ldots,2^n-1\}: N_{\frac{i+1}{2^n}}-N_{\frac{i}{2^n}}\geq 2\right]\\&=P\left[\bigcap_{n\in\mathbb{N}}C_n\right]=1-P\left[\bigcup_{n\in\mathbb{N}}C_n^c\right]=1-\lim_{n\to\infty}P\left[C_n^c\right],\end{aligned}$$

ya que $\{C_n^c\}_{n\in\mathbb{N}}$ es una sucesión creciente de eventos. Tenemos que

$$\begin{aligned}P\left[C_n^c\right]&=P\left[N_{\frac{i+1}{2^n}}-N_{\frac{i}{2^n}}\leq 1\quad\forall i\in\{1,2,\ldots,2^n-1\}\right]=\prod_{i=1}^{2^n-1}P\left[N_{\frac{i+1}{2^n}}-N_{\frac{i}{2^n}}\leq 1\right]\\&=\prod_{i=1}^{2^n-1}\left[1+(\lambda_1+\lambda_2)2^{-n}\right]e^{-(\lambda_1+\lambda_2)2^{-n}}=\left[1+(\lambda_1+\lambda_2)2^{-n}\right]^{2^n}e^{-(\lambda_1+\lambda_2)},\end{aligned}$$

por lo que

$$P\left[\text{ exista un evento múltiple en } (0,1]\right]=1-e^{-(\lambda_1+\lambda_2)}\lim_{n\to\infty}\left[1+(\lambda_1+\lambda_2)2^{-n}\right]^{2^n}=0.$$

(ii) Usaremos la definición alternativa 60. Ya notamos que $N_t\sim\mathscr{P}((\lambda_1+\lambda_2)t))$, por lo que basta mostrar que los incrementos son estacionarios e independientes. Tenemos que por definición

$$N_{t+s}-N_t=\left(N_{t+s}^1+N_{t+s}^2\right)-\left(N_t^1+N_t^2\right)=\left(N_{t+s}^1-N_t^1\right)+\left(N_{t+s}^2-N_t^2\right)\sim\mathscr{P}((\lambda_1+\lambda_2)s)$$

ya que corresponde a suma de vs.as. de Poisson independientes, por la estacionaridad de los incrementos de N^1 y N^2. Por otro lado, dados los tiempos $t_0=0\leq t_1\leq t_2\leq\cdots\leq t_n$, tenemos que

$$\left\{N_{t_{i+1}}-N_{t_i}\right\}_{i=0}^{n-1}=\left\{\left(N_{t_{i+1}}^1-N_{t_i}^1\right)+\left(N_{t_{i+1}}^2-N_{t_i}^2\right)\right\}_{i=0}^{n-1},$$

por lo que la independencia entre incrementos en intervalos disjuntos se deduce del hecho que $N^1\perp\!\!\!\perp N^2$ y los incrementos de cada proceso son independientes.

□

Proposición 68. *Sean N^1 y N^2 dos procesos de Poisson independientes de parámetros λ_1 y λ_2 respectivamente. Para todo $n\in\mathbb{N}$, $t\geq 0$, se cumple que*

$$N_t^1\Big|_{N_t^1+N_t^2=n}\sim Bi(n,\tfrac{\lambda_1}{\lambda_1+\lambda_2}).$$

Dem. Dado $k \in \{0, 1, \ldots, n\}$, por Teorema de Bayes tenemos que

$$\begin{aligned}
P\left[N_t^1 = k | N_t^1 + N_t^2 = n\right] &= P\left[N_t^1 + N_t^2 = n | N_t^1 = k\right] \frac{P\left[N_t^1 = k\right]}{P\left[N_t^1 + N_t^2 = n\right]} \\
&= \frac{P\left[N_t^2 = n-k\right] P\left[N_t^1 = k\right]}{P\left[N_t^1 + N_t^2 = n\right]} \\
&= \frac{n!}{(n-k)!k!} \left(\tfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(1 - \tfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^{(n-k)}
\end{aligned}$$

□

4.3. Adelgazamiento (o coloreo) de un proceso de Poisson

Dado $(N_t, t \geq 0)$ un proceso de Poisson de parámetro λ, usaremos los tiempos de llegada $\{S_n\}_{n\in\mathbb{N}}$ asociados y una sucesión$\{B_n\}_{n\in\mathbb{N}}$ de vs. as. i.i.d con $B_n \sim Be(p)$ independientes de los tiempos de llegada para construir un nuevo proceso. Para esto, definimos la sucesión $\{\tilde{S}_k\}_{k\in\mathbb{N}}$ de manera inductiva : $\tilde{S}_0 \equiv 0$; y dado $\tilde{S}_{k-1}$,

$$\tilde{S}_k := \text{mín}\{S_n > \tilde{S}_{k-1} : S_n B_n > 0\}.$$

Proposición 69. *El proceso* $(M_t, t \geq 0)$ *definido como*

$$M_t = \text{máx}\{n > 0 : \tilde{S}_n < t\} = \text{máx}\left\{n > 0 : \sum_{i=1}^{n} \tilde{T}_i < t\right\},$$

dónde $\tilde{T}_i = \tilde{S}_{i+1} - \tilde{S}_i$ *es un proceso de Poisson de parámetro* λp.

Dem. Por la definición 63, basta probar que $\{\tilde{T}_i\}_{i\in\mathbb{N}}$ es una sucesión de vs. as. i.i.d. con $T_i \sim Exp(\lambda p)$. Tenemos entonces que

$$\begin{aligned}
P\left[\tilde{T}_1 > t\right] = P\left[\tilde{S}_1 > t\right] &= \sum_{n=0}^{\infty} P\left[\tilde{S}_1 > t | N_t = n\right] P\left[N_t = n\right] \\
&= \sum_{n=0}^{\infty} P\left[B_i = 0 \, \forall i \in \{1, 2, ..., n\}\right] P\left[N_t = n\right] \\
&= \sum_{n=0}^{\infty} (1-p)^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = e^{-\lambda p t}.
\end{aligned}$$

Por otro lado, dado $s, t > 0$

$$\begin{aligned}
P\left[\tilde{T}_1 > s, \tilde{T}_2 > t\right] &= \int_s^{\infty} P\left[\tilde{T}_2 > t | \tilde{T}_1 = u\right] f_{\tilde{T}_1}(u) du = \int_s^{\infty} P\left[\tilde{N}_{t+u} - \tilde{N}_u < 1\right] \lambda e^{-\lambda p u} du \\
&= \int_s^{\infty} \sum_{k=0}^{\infty} P\left[\tilde{N}_{t+u} - \tilde{N}_u = 0 | N_{t+u} - N_u = k\right] P\left[N_{t+u} - N_u = k\right] \lambda e^{-\lambda p u} du \\
&= \int_s^{\infty} \sum_{k=0}^{\infty} (1-p)^k \frac{(\lambda t)^k}{k!} e^{-\lambda t} \lambda e^{-\lambda p u} du = e^{-\lambda p t} e^{-\lambda p s},
\end{aligned}$$

de dónde $\tilde{T}_2 \sim Exp(\lambda p)$ y $\tilde{T}_1 \perp\!\!\!\perp \tilde{T}_2$. Siguiendo con esta idea de manera recursiva tenemos el resultado deseado. □

4.4. Procesos de Poisson no homogéneo

Definición 70. *Un **proceso de Poisson no homogéneo** $(X_t, t \geq 0)$ es un proceso càdlàg no decreciente tal que*

(i) $X_0 \equiv 0$, $X_t \in \mathbb{N}_0$,

(ii) X *tiene incrementos independientes, y*

(iii) *Existe una función integrable* $\lambda(t) \geq 0$ *para todo* $t \geq 0$*, llamada **intensidad**, tal que*

$$\begin{cases} P\left[X_{t+h} - X_t = 1\right] & = \lambda(t)h + o(h) \\ P\left[X_{t+h} - X_t = 0\right] & = 1 - \lambda(t)h + o(h) \\ P\left[X_{t+h} - X_t \geq 2\right] & = o(h) \end{cases}$$

Claramente, en el caso $\lambda(t) := \lambda$ recuperamos el proceso de Poisson homogéneo original.

Proposición 71. *Dado* $t \geq 0$*, tenemos que* $X_t \sim \mathscr{P}(\Lambda(t))$*, dónde*

$$\Lambda(t) := \int_0^t \lambda(s) ds.$$

La demostración de esta proposición es análoga a la de la Prop. 59 y se deja como ejercicio.

Si $\lambda(t)$ es una función continua, tenemos que $\Lambda(t)$ es una función diferenciable con $\Lambda'(t) = \lambda(t)$.

Proposición 72. *Para todo par* $s, t \geq 0$ $X_{t+s} - X_t \sim \mathscr{P}(\Lambda(t+s) - \Lambda(t))$.

Dem. Para probar el resultado deseado, usaremos la independencia de los incrementos y las distribuciones marginales del proceso. Tenemos entonces que la Función generadora de momentos de la variable $X_{t+s} - X_t$ está dada por

$$\begin{aligned} M_{X_{t+s}-X_t}(\theta) = E\left[e^{\theta(X_{t+s}-X_t)}\right] \frac{E\left[e^{\theta X_t}\right]}{E\left[e^{\theta X_t}\right]} &= \frac{E\left[e^{\theta X_t}\right] E\left[e^{\theta(X_{t+s}-X_t)}\right]}{E\left[e^{\theta X_t}\right]} = \frac{E\left[e^{\theta X_t} e^{\theta(X_{t+s}-X_t)}\right]}{E\left[e^{\theta X_t}\right]} \\ &= \frac{E\left[e^{\theta X_{t+s}}\right]}{E\left[e^{\theta X_t}\right]} = \frac{e^{\Lambda(t+s)(e^\theta - 1)}}{e^{\Lambda(t)(e^\theta-1)}} = e^{[\Lambda(t+s)-\Lambda(t)](e^\theta-1)}, \end{aligned}$$

donde usamos que la FGM de $N \sim Po(\ell)$ está dada por

$$M_N(\theta) = \mathbb{E}\left[e^{\theta X}\right] = e^{\ell(e^\theta - 1)} \qquad \forall \theta \in \mathbb{R}.$$

Del hecho que la FGM caracteriza la distribución, tenemos que el incremento es una variable aleatoria de Poisson de parámetro $\Lambda(t+s) - \Lambda(t)$. $\square$

Nota. *Es importante notar que los tiempos interarribos de este proceso no son independientes ni igualmente distribuídos. En particular* $P[\tau_1 > t] = P[X_t = 0] = e^{-\Lambda(t)}$*, y*

$$P[\tau_1 > s, \tau_2 > t] = \int_s^{\infty} P[X_{t+u} - X_u = 0]\,\lambda(u) e^{-\Lambda(u)} du = \int_s^{\infty} e^{-\Lambda(t+u)} \lambda(u) du,$$

que no puede factorizarse si $\lambda(t)$ *no es constante.*

Dado un proceso de Poisson no homogéneo, podemos recuperar un proceso homogéneo haciendo un cambio de tiempo conveniente, como veremos a continuación.

Proposición 73. *Sea* $(X_t, t \geq 0)$ *un Proceso de Poisson no homogéneo asociado a la función* $\lambda(\cdot)$*. Dada*

$$\Lambda^{-1}(t) = \text{ínf}\{u \geq 0 : \Lambda(u) = t\}$$

la inversa por derecha de la función $\Lambda(\cdot)$*, el proceso* $(\tilde{X}_t, t \geq 0)$ *definido por* $\tilde{X}_t := X_{\Lambda^{-1}(t)}$ *es un Proceso de Poisson de parámetro* $\lambda = 1$.

Dem. Dada una función $\lambda(t) > 0$, entonces $\Lambda(t) = \int_0^t \lambda(r) dr$ es una función continua y no decreciente. Por lo tanto, la función Λ^{-1} es estrictamente creciente con $\Lambda(\Lambda^{-1}(t)) = t$ [10]

Dada una partición $t_0 < t_1 < \ldots < t_n$, tenemos que $\Lambda^{-1}(t_0) \leq \Lambda^{-1}(t_1) \leq \cdots \leq \Lambda^{-1}(t_n)$, y por lo tanto

$$\left\{\tilde{X}_{t_{i+1}} - \tilde{X}_{t_i}\right\}_{i=0}^{n-1} = \left\{X_{\Lambda^{-1}(t_{i+1})} - X_{\Lambda^{-1}(t_i)}\right\}_{i=0}^{n-1},$$

por lo que el proceso $\tilde{X}$ hereda los incrementos independientes del proceso X. Además, como

$$\tilde{X}_{t+s} - \tilde{X}_s = X_{\Lambda^{-1}(t+s)} - X_{\Lambda^{-1}(s)} \sim \mathscr{P}\left(\Lambda(\Lambda^{-1}(t+s)) - \Lambda(\Lambda^{-1}(s))\right) = \mathscr{P}(t),$$

tenemos que los incrementos son estacionarios. Como $\tilde{X}_t = X_{\Lambda^{-1}(t)} \sim \mathscr{P}\left(\Lambda(\Lambda^{-1}(t))\right) = \mathscr{P}(t)$, tenemos que $\tilde{X}$ es un PP(1) por la Definición 60. □

4.5. Proceso de Poisson Compuesto

Definición 74. *Dado* $(N_t, t \geq 0)$ *un proceso de Poisson de parámetro* λ *y sean* $\{Y_n\}_{n \in \mathbb{N}}$ *una sucesión de variables i.i.d. independientes de* N*. El proceso* $(X_t, t \geq 0)$ *dado por*

$$X_t := \sum_{n=1}^{N_t} Y_n$$

se denomina ***Proceso de Poisson Compuesto****.*

[10] los intervalos de constancia de $\Lambda(\cdot)$ corresponden a saltos en $\Lambda^{-1}(\cdot)$, y en general $\Lambda^{-1}(\Lambda(t)) \neq t$.

Este proceso sirve para modelar diversas situaciones, por ejemplo la cantidad de dinero extraída por los individuos que visitan un cajero electrónico o el tamaño de los archivos enviados a un determinado servidor.

Proposición 75. *Un proceso de Poisson compuesto* $(X_t, t \geq 0)$ *cumple las siguientes propiedades*

(i) $M_{X_t}(\theta) = E\left[e^{\theta X_t}\right] = e^{\lambda t[M_{Y_1}(\theta)-1]}$.

(ii) X *tiene incrementos estacionarios e independientes.*

(iii) $E[X_t] = \lambda t E[Y_1], \quad Var[X_t] = \lambda t E[Y_1^2], \quad Cov[X_t, X_s] = \lambda E[Y_1^2] \operatorname{mín}\{s,t\}$.

Dem.

(i)

$$E\left[e^{\theta X_t}\right] = E\left[E\left[e^{\theta \sum_{n=1}^{N_t} Y_n} \Big| N_t\right]\right] = E\left[E\left[e^{\theta Y_n}\right]^{N_t}\right] = E\left[M_{Y_1}(\theta)^{N_t}\right]$$
$$= E\left[e^{\log(M_{Y_1}(\theta))N_t}\right] = e^{\lambda t\left[M_{Y_1}(\theta)-1\right]}$$

(ii) Dado que

$$\left\{X_{t_{i+1}} - X_{t_i}\right\}_{i=0}^{n-1} = \left\{\sum_{n=N_{t_i}+1}^{N_{t_{i+1}}} Y_n\right\}_{i=0}^{n-1},$$

podemos deducir la independencia de los incrementos de X de la independencia de los incrementos de N y las variables $\{Y_n\}$. Por otro lado, usando la independencia de los incrementos tenemos que

$$E\left[e^{\theta(X_{t+s}-X_s)}\right] = \frac{E\left[e^{\theta(X_{t+s}-X_s)}e^{\theta X_s}\right]}{E\left[e^{\theta X_s}\right]} = \frac{e^{\lambda(t+s)\left[M_{Y_1}(\theta)-1\right]}}{e^{\lambda s\left[M_{Y_1}(\theta)-1\right]}} = e^{\lambda t\left[M_{Y_1}(\theta)-1\right]},$$

y por lo tanto los incrementos son estacionarios.

(iii) Recordemos que $E[X_t] = \frac{\partial M_{X_t}(\theta)}{\partial\theta}\Big|_{\theta=0}$, y por lo tanto

$$E[X_t] = \lambda t \frac{\partial M_{Y_1}(\theta)}{\partial\theta}\Big|_{\theta=0} = \lambda t E[Y_1].$$

Análogamente

$$E\left[X_t^2\right] = \left[(\lambda t)^2\left(\frac{\partial M_{Y_1}(\theta)}{\partial\theta}\right)^2 + \lambda t\frac{\partial^2 M_{Y_1}(\theta)}{\partial\theta^2}\right]\Big|_{\theta=0} = (\lambda t)^2 E[Y_1]^2 + \lambda t E[Y_1^2],$$

de dónde

$$Var[X_t] = E\left[X_t^2\right] - E[X_t]^2 = \lambda t E\left[Y^2\right].$$

Por otro lado, si $s < t$ tenemos que

$$Cov[X_t - X_s + X_s, X_s] = Cov[X_t - X_s, X_s] - Cov[X_s, X_s] = Var[X_s],$$

dónde en la primera igualdad usamos la linealidad de la covarianza y en la segunda el hecho que los incrementos son independientes.

□

4.5.1. Modelo Clásico de Cramer-Lundberg

Queremos modelar la evolución del capital de una empresa de seguros a lo largo del tiempo. Dado un capital inicial u y una cuota c, usaremos un proceso de Poisson compuesto para representar los siniestros que debe cubrir la empresa hasta un tiempo determinado. Dada $(N_t, t \geq 0)$ un proceso de Poisson de parámetro λ (frecuencia), y $\{Y_n\}_{n\in\mathbb{N}}$ (reclamaciones) vs. as. i.i.d. con $E[Y_i] = \mu$ (severidad), tenemos

$$C_t = u + ct - \sum_{n=1}^{N_t} Y_n, \qquad t \geq 0.$$

El valor medio del capital al tiempo t está dado por

$$E[C_t] = u + ct - E\left[\sum_{n=1}^{N_t} Y_n\right] = u + ct - \lambda t \mu,$$

y si queremos que esta media sea siempre positiva, debe ser $(c - \lambda\mu)t > 0$, es decir $c > \lambda\mu$, que se conoce como la Condición de Ganancia Neta.

Otra noción a estudiar es la probabilidad de ruina en un horizonte finito en función del capital inicial, es decir

$$\psi(u,t) = P[\tau < t | C_0 = u],$$

donde $\tau = \text{ínf}\{t > 0 : C_t < 0\}$ es el momento de entrada a $(-\infty, 0)$. Conocer la distribución de la variable aleatoria τ es aún un problema abierto en muchos casos, y es el principal objeto de estudio en Teoría del Riesgo.

5. Martingalas

En esta sección estudiaremos las definiciones y resultados básicos asociados a la teoría de martingalas a tiempos discretos. El nombre se debe a la estrategia de apuestas que se usaba en Francia en el siglo XVIII, pero el primero en usarlo formalmente en probabilidad fue Paul Lévy en 1934.

Definición 76. *El proceso* $\{X_n\}_{n\in\mathbb{N}}$ *es una* ***martingala*** *si para todo* $n \geq 0$ *se cumple que*

(i) $E[|X_n|] < \infty$, *y*

(ii) $E[X_{n+1}|X_0, X_1, \dots, X_n] = X_n$

La definición anterior es un caso particular de la definición general de martingala que se presenta a continuación, dada en términos de una *filtración* (que es una familia de $\sigma-$álgebras no decreciente, revisar Def. 36).

Definición 77. *Dado un espacio de probabilidad filtrado* $(\Omega, \mathscr{F}, \{\mathscr{G}_n\}_{n\in\mathbb{N}}, P)$ *y un proceso estocástico* $\{X_n\}_{n\in\mathbb{N}}$ *en* $(\Omega, \mathscr{F}, P)$ *es una* ***martingala con respecto a la filtración*** $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ *si para cada* $n \in \mathbb{N}$ *satisface*

(i) $E[|X_n|] < \infty$*;*

(ii) X_n *es* $\mathscr{G}_n$*-medible, es decir* $\forall B \in \mathscr{B}$, $X^{-1}(B) \in \mathscr{G}_n$*; y*

(iii) $E[X_{n+1}|\mathscr{G}_n] = X_n$.

La definición 76 corresponde a una martingala con respecto a su filtración natural $\{\mathscr{F}_n\}_{n\in\mathbb{N}}$, dónde $\mathscr{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$.

La condición (iii) puede interpretarse de la siguiente manera: *la mejor aproximación del proceso al momento* $n+1$*, dada la información que tengo hasta el momento* n*, es el comportamiento del proceso en el instante* n*.*[11]

Nota. *Un proceso que cumple la propiedad* (ii) *en la definición anterior se dice* ***adaptado a la filtración*** $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$.

[11] Para repasar la definición de esperanza condicional, su interpretación en el espacio L^2 y alguna de sus propiedades, revise la sección 2.6 del Apunte de Probabilidad II.

Ejemplo 78. Sea $\{S_n\}_{n\in\mathbb{N}}$ la caminata aleatoria simple, dada por $S_0 = 0$ y

$$S_n = \sum_{i=1}^{n} Y_i, \quad \{Y_i\} \text{ i.i.d. , con } Y_i = \begin{cases} 1 & \text{con probabilidad } p \\ -1 & \text{con probabilidad } q = 1-p \end{cases}.$$

Si $p = q$, este proceso es una martingala, ya que

$$E[|S_n|] \leq \sum_{i=1}^{n} E[|Y_i|] = n < \infty \quad \forall n \in \mathbb{N},$$

y

$$E[S_{n+1}|S_0, S_1, \dots, S_n] = E[S_{n+1}|S_n] = E[S_n + Y_{n+1}|S_n] = S_n + E[Y_{n+1}] = S_n + (p-q),$$

donde en la primera igualdad usamos que el proceso es Markoviano, y en las dos últimas las propiedades de la esperanza condicional. Claramente si $p \neq q$, la caminata aleatoria no es una martingala, ya que el proceso tiene una *deriva* $p-q$. ¿Qué podemos hacer para obtener una martingala asociada? simplemente corregir esta deriva. Si tomamos ahora el proceso $\{\tilde{S}_n\}_{n\in\mathbb{N}}$ dado por

$$\tilde{S}_n = S_n - n(p-q),$$

tenemos que $E\left[|\tilde{S}_n|\right] \leq E[|S_n|] + 2n < \infty$, y

$$E\left[\tilde{S}_{n+1}|\tilde{S}_1, \dots, \tilde{S}_n\right] = E\left[\tilde{S}_{n+1}|\tilde{S}_n\right] = E\left[\tilde{S}_n + Y_{n+1} - (p-q)|\tilde{S}_n\right] = \tilde{S}_n + (p-q) - (p-q) = \tilde{S}_n.$$

Ejemplo 79. Es importante notar que una martingala no tiene por que cumplir la propiedad de Markov. Por ejemplo, sea $\{X_n\}_{n\in\mathbb{N}}$ el proceso dado por

$$X_{n+1} = X_n + \xi_{n+1} \max\{X_1, \dots, X_n\}, \quad X_0 := 0, X_1 := \xi_1,$$

con $\{\xi_n\}_{n\in\mathbb{N}}$ i.i.d. con $E[\xi_n] = 0$ y $E[|\xi_n|] < M$. Es claro que $(X_n, n \in \mathbb{N}_0)$ no es un proceso de Markov, pero probemos que es una martingala. Por construcción, tenemos que

$$E[|X_n|] \leq \sum_{i=1}^{n} iM^i < \infty \qquad \forall n \in \mathbb{N}, \text{ y}$$

$$\begin{aligned} E[X_{n+1}|X_1, \dots, X_n] &= E[X_n + \xi_{n+1} \max\{X_1, \dots, X_n\}|X_1, \dots, X_n] \\ &= X_n + \max\{X_1, \dots, X_n\} E[\xi_{n+1}] = X_n. \end{aligned}$$

Ejemplo 80. Pensemos en la estrategia de juego “Doblar la apuesta”. Dado un juego en el que puedo ganar o perder con probabilidad $1/2$, empiezo apostando 1\$. Si gano, me retiro, y si pierdo, apuesto 2\$. En el siguiente paso, si gano, me retiro, y si pierdo apostaré 4\$ y asi sucesivamente. Una vez que me retiré me quedo con lo obtenido hasta ese momento. Lo podemos escribir de la siguiente manera: $G_0 = 0$, y

$$G_n = G_{n-1} + \left(2^{n-1} Y_n\right) \mathbb{1}_{\{G_{n-1}<0\}},$$

donde $\{Y_n\}$ son i.i.d. que toman el valor 1 ó -1 con probabilidad $1/2$. Es importante notar que si bien podríamos pensar que el proceso no es Markoviano (porque detenerme o no depende de si gané en algún momento en el pasado), la definición anterior nos da una manera Markoviana de escribirlo, ya que con esta estrategia la ganancia al retirarnos siempre será 1\$. Entonces, el proceso $\{G_n\}_{n\in\mathbb{N}}$ es una martingala, ya que $E[|G_n|] \leq \sum_{k=1}^{n} 2^{k-1} E[|Y_k|] = 2^n - 1 < \infty$ para todo $n \in \mathbb{N}$; y por otro lado

$$E[G_{n+1}|G_0, G_1, \ldots, G_n] = E[G_{n+1}|G_n] = G_n + \mathbb{1}_{\{G_n<0\}} 2^n E[Y_{n+1}] = G_n.$$

Sin embargo, veamos que esta estrategia no es viable en la práctica. Definamos

$$\tau = \text{ínf}\{n > 0 : G_n = 1\}$$

el tiempo en el que nos detenemos, es decir la primera vez que ganamos. Dado que $\{G_n\}_{n\in\mathbb{N}}$ es una cadena de Markov irreducible y absorbente en 1, podemos probar que $P[T < \infty] = 1$. Sin embargo, como por construcción $T \sim Geom(1/2)$, tenemos que la ganancia media antes de retirarnos (es decir el dinero que habremos apostado en total justo antes de ganar) es

$$E[G_{T-1}] = E\left[-\sum_{k=0}^{T-2} 2^k\right] = 1 - E\left[2^{T-1}\right] = 1 - \sum_{n=0}^{\infty} 2^n P[T = n+1] = -\infty,$$

pues $P[T = n+1] = (1/2)^{n+1}$. Es decir que si bien el juego es "justo", en el sentido de que en algún momento puedo recuperar el dinero perdido, el tiempo que tengo que esperar para que eso pase puede ser tan grande como quiera. [12]

Ejercicio: Pruebe que el proceso de ramificación $\{Z_n\}_{n\in\mathbb{N}}$ definido en el Ejemplo 19 no es una martingala, pero el proceso $\{M_n\}_{n\in\mathbb{N}}$ dado por $M_n = Z_n/\mu^n$, con $E[\xi] = \mu$ sí lo es.

Ejemplo 81. Dada U una variable aleatoria con distribución uniforme en el intervalo $[0,1]$, definimos el proceso $\{X_n\}_{n\in\mathbb{N}}$ dado por $X_n = 2^n \mathbb{1}_{\{U\leq 2^{-n}\}}$. Veamos que este proceso es martingala con respecto a su filtración natural, ya que se cumple

- $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = 1 \; \forall n \in \mathbb{N}$; y

[12] otra forma de comprobar esto es mostrar que $E[T] = +\infty$, y la cadena es recurrente pero nula.

- $$\mathbb{E}[X_{n+1}|X_1,\dots,X_n] = 0\mathbb{1}_{\{X_n=0\}} + \mathbb{1}_{\{X_n=2^n\}}\mathbb{E}\left[X_{n+1}|X_n = 2^n\right]$$
$$= 0\mathbb{1}_{\{U>2^{-n}\}} + 2^n\mathbb{1}_{\{U\leq 2^{-n}\}} = X_n \qquad \forall n\in\mathbb{N}.$$

Por otro lado, si definimos la filtración $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ dada por $\mathscr{G}_n := \sigma_U = \sigma\{U^{-1}(B) : B\in\mathscr{B}\}$ para todo $n\in\mathbb{N}$, es claro que X_n es $\mathscr{G}_n$-medible por ser una función determinista de la variable U, pero se tiene que $\mathbb{E}[X_{n+1}|\mathscr{G}_n] = X_{n+1}$ pues X_{n+1} también es $\mathscr{G}_n$-medible. Es decir, $\{X_n\}_{n\in\mathbb{N}}$ no es martingala con respecto a $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$, pues dada σ_U, el proceso esta totalmente determinado y no necesito pensar en *aproximaciones.*

Dada la definición de martingala, tenemos las siguientes definiciones relacionadas

Definición 82. *Dado un espacio de probabilidad filtrado $(\Omega,\mathscr{F},\{\mathscr{G}_n\}_{n\in\mathbb{N}},P)$ y un proceso estocástico $\{X_n\}_{n\in\mathbb{N}}$ en $(\Omega,\mathscr{F},P)$ tal que para cada $n\in\mathbb{N}$ X_n es $\mathscr{G}_n$-medible con $E[|X_n|]<\infty$, decimos que*

*(i) $\{X_n\}_{n\in\mathbb{N}}$ es una **submartingala** con respecto a la filtración $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ si*

$$E[X_{n+1}|\mathscr{G}_n] \geq X_n.$$

*(ii) $\{X_n\}_{n\in\mathbb{N}}$ es una **supermartingala** con respecto a la filtración $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ si*

$$E[X_{n+1}|\mathscr{G}_n] \leq X_n.$$

Nota. *Notemos que una martingala es en particular una sub/supermartingala. Además, tenemos que si $\{X_n\}_{n\in\mathbb{N}}$ es una <u>martingala</u>, entonces*

$$E[X_n] = E[E[X_n|\mathscr{G}_{n-1}]] = E[X_{n-1}] = \cdots = E[X_0],$$

por lo que tiene <u>esperanza constante</u>. De manera análoga, si $\{X_n\}_{n\in\mathbb{N}}$ es una <u>submartingala</u>, entonces

$$E[X_n] = E[E[X_n|\mathscr{G}_{n-1}]] \geq E[X_{n-1}] \geq \cdots \geq E[X_0],$$

por lo que tiene <u>esperanza creciente</u>; y si $\{X_n\}_{n\in\mathbb{N}}$ es una <u>supermartingala</u>, entonces

$$E[X_n] = E[E[X_n|\mathscr{G}_{n-1}]] \leq E[X_{n-1}] \leq \cdots \leq E[X_0],$$

por lo que tiene <u>esperanza decreciente</u>.

Proposición 83. *Si $\{X_n\}_{n\in\mathbb{N}}$ es una martingala respecto a la filtración $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ y $\phi:\mathbb{R}\to\mathbb{R}$ es una función convexa tal que $E\left[|\phi(X_n)|\right]<\infty$, entonces $\{\phi(X_n)\}_{n\in\mathbb{N}}$ es una submartingala respecto a la misma filtración.*

Dem. Como cualquier función convexa (sobre $\mathbb{R}$) es continua, entonces $\phi(X_n)$ es $\mathscr{G}_n$-medible, y por hipótesis satisface $E\left[|\phi(X_n)|\right]<\infty$. Finalmente, por la desigualdad de Jensen para la esperanza condicional,

$$\mathbb{E}\left[\phi(X_{n+1})|\mathscr{G}_n\right]\geq\phi(\mathbb{E}\left[X_{n+1}|\mathscr{G}_n\right]=\phi(X_n).$$

□

5.0.1. Convergencia de martingalas

Podemos pensar en una martingala $\{X_n\}_{n\in\mathbb{N}}$ como una sucesión de variables aleatorias definidas en el mismo espacio. En ese sentido, nos interesa estudiar bajo que condiciones de acotamiento podemos asegurar la convergencia de una martingala, es decir la existencia de una variable aleatoria X_∞ tal que $X_n \underset{n\to\infty}{X_\infty}$ en alguna de las nociones conocidas. Para abordar este resultado, es importante extender la igualdad $\mathbb{E}[X_n]=E[X_0]$ no sólo a tiempos deterministas como n, sino también a los tiempos de paro, definidos en 37. Tenemos entonces la siguiente definición

Definición 84. *Dado un proceso estocástico $\{X_n\}_{n\in\mathbb{N}}$ adaptado a la fitración $\{\mathscr{F}_n\}_{n\in\mathbb{N}}$, y τ un tiempo de paro con respecto a la misma filtración, definimos el* ***proceso detenido al tiempo*** *τ, denotado por $\{X_n^\tau\}_{n\in\mathbb{N}}$, por*

$$X_n^\tau=X_{n\wedge\tau}=\begin{cases} X_n & si\ n\leq\tau\\ X_\tau & si\ n>\tau.\end{cases}$$

Ejemplo 85. En el Ejemplo 80, puedo considerar el proceso $\{X_n\}_{n\in\mathbb{N}}$ que corresponde jugar eternamente doblando la apuesta dado por $X_n=X_{n-1}+2^{n-1}Y_n$, con $X_0=0$. Entonces el proceso $\{G_n\}_{n\in\mathbb{N}}$ se obtiene como el proceso detenido dado por $G_n=X_n^\tau$, con $\tau=\text{ínf}\{n\geq 0: X_n=1\}$, que es t.d.p. con respecto a la filtración natural del proceso X.

Ejemplo 86. Dada la caminata aleatoria del Ejemplo 78, si tenemos el tiempo de paro $\tau:=\text{ínf}\{n:|S_n|=k\}$, el proceso detenido $\{S_n^\tau\}_{n\in\mathbb{N}}$ es la caminata aleatoria absorbida en k y $-k$.

Proposición 87. *Si $\{X_n\}_{n\in\mathbb{N}}$ es una martingala (o sub/supermg. respectivamente) con respecto a la filtración $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$ el proceso detenido $\{X_n^\tau\}_{n\in\mathbb{N}}$ también lo es.*

Dem. Veamos que cumple con las tres propiedades de la Definición 77.

(i) Sabemos por la definición de martingala que $E[|X_k|] < \infty$ para todo $k \in \mathbb{N}$. Entonces tenemos que

$$\begin{aligned} E\left[|X_n^\tau|\right] &= \sum_{k=1}^{\infty} E\left[|X_n^\tau| \mathbb{1}_{\{\tau=k\}}\right] = \sum_{k=1}^{n} E\left[|X_k| \mathbb{1}_{\{\tau=k\}}\right] + \sum_{k=n+1}^{\infty} E\left[|X_n| \mathbb{1}_{\{\tau=k\}}\right] \\ &\leq \sum_{k=1}^{n} E[|X_k|] + E[|X_n|] < \infty. \end{aligned}$$

(ii) Dado $B \in \mathscr{B}$, tenemos que

$$\left\{X_n^\tau \in B\right\} = \left(\bigcup_{k=1}^{n} \{X_k \in B, \tau = k\}\right) \bigcup (X_n \in B, \tau > n) \in \mathscr{G}_n,$$

ya que

$$\{X_k \in B, \tau = k\} = \{X_k \in B\} \cap \{\tau = k\} \in \mathscr{G}_k \subseteq \mathscr{G}_n,$$

y la σ-álgebra $\mathscr{G}_n$ es cerrada para uniones e intersecciones numerables.

(iii)

$$\begin{aligned} E\left[X_{n+1}^\tau | \mathscr{G}_n\right] &= \sum_{k=1}^{n} E\left[X_k \mathbb{1}_{\{\tau=k\}} | \mathscr{G}_n\right] + E\left[X_{n+1} \mathbb{1}_{\{\tau>n\}} | \mathscr{G}_n\right] \\ &= \sum_{k=1}^{n} X_k \mathbb{1}_{\{\tau=k\}} + \mathbb{1}_{\{\tau>n\}} E[X_{n+1} | \mathscr{G}_n] = \sum_{k=1}^{n} X_k \mathbb{1}_{\{\tau=k\}} + \mathbb{1}_{\{\tau>n\}} X_n \\ &= X_n^\tau. \end{aligned}$$

El caso sub/supermg es análogo, trabajando con la desigualdad correspondiente. $\square$

5.1. Teorema del Paro Opcional de Doob

Teorema 88. *Dada una martingala $\{X_n\}_{n\in\mathbb{N}}$ y un t.d.p. finito, ambos con respecto a la misma filtración $\{\mathscr{G}_n\}_{n\in\mathbb{N}}$, si se cumple*

(i) $E[|X_\tau|] < \infty$, *y*

(ii) $\lim_{n\to\infty} E\left[X_n \mathbb{1}_{\{\tau>n\}}\right] = 0$

entonces $E[X_\tau] = E[X_n]$ *para todo* $n \in \mathbb{N}$*, en particular*

$$E[X_\tau] = E[X_1].$$

Dem. Para cada $n \in \mathbb{N}$ podemos escribir

$$
\begin{aligned}
X_\tau &= X_\tau \mathbb{1}_{\{\tau \le n\}} + X_\tau \mathbb{1}_{\{\tau > n\}} + X_n \mathbb{1}_{\{\tau > n\}} - X_n \mathbb{1}_{\{\tau > n\}} \\
&= \left(X_\tau \mathbb{1}_{\{\tau \le n\}} + X_n \mathbb{1}_{\{\tau > n\}} \right) + (X_\tau - X_n) \mathbb{1}_{\{\tau > n\}} \\
&= X_n^\tau + (X_\tau - X_n) \mathbb{1}_{\{\tau > n\}}.
\end{aligned}
$$

Tenemos entonces que

$$
E[X_\tau] = E\left[X_n^\tau\right] + E\left[(X_\tau - X_n) \mathbb{1}_{\{\tau > n\}}\right] = E\left[X_1^\tau\right] + \lim_{n\to\infty} E\left[X_\tau \mathbb{1}_{\{\tau > n\}}\right] + \lim_{n\to\infty} E\left[X_n \mathbb{1}_{\{\tau > n\}}\right] = E[X_1],
$$

ya que $E\left[X_n^\tau\right] = E\left[X_1^\tau\right] = E[X_1]$ por ser martingala, y

$$
\lim_{n\to\infty} E\left[X_\tau \mathbb{1}_{\{\tau > n\}}\right] = \lim_{n\to\infty} \sum_{k=n+1}^{\infty} E\left[X_k \mathbb{1}_{\{\tau = k\}}\right] = 0,
$$

pues por hipótesis $P[\tau < \infty] = 1$ y

$$
\left| \sum_{k=0}^{\infty} E\left[X_k \mathbb{1}_{\{\tau = k\}}\right] \right| = |E[X_\tau]| \le E[|X_\tau|] < \infty,
$$

por lo que el límite anterior es la cola de una serie absolutamente convergente. □

Ejemplo 89. Juan y María juegan a apostar a los volados. Juan empieza con a\$ y María con b\$. Tenemos entonces que la ganancia de María luego de n pasos está dada por

$$
X_n = S_n + b,
$$

dónde S_n es la caminata aleatoria simétrica. Ya vimos que en este caso $\{S_n\}$ es martingala con respecto a su filtración natural, por lo que tiene esperanza constante igual a 0.

Estamos interesados en calcular la probabilidad de que María se quede con todo el capital antes que Juan, es decir en calcular $P[T_a < T_{-b}]$, dónde $T_z = \inf\{n : S_n = z\}$.

Claramente $\tau = T_a \wedge T_{-b}$ es un t.d.p. con respecto a la filtración natural de $\{S_n\}_{n\in\mathbb{N}}$, con $E[|S_\tau|] \le \max\{a, b\} < \infty$, y

$$
0 = \lim_{n\to\infty} -bP[\tau > n] \le \lim_{n\to\infty} E\left[S_n \mathbb{1}_{\{\tau > n\}}\right] \le \lim_{n\to\infty} aP[\tau > n] = 0,
$$

por lo que por el teorema de paro opcional tenemos que $E[S_\tau] = E[S_1] = 0$, y entonces

$$
\begin{aligned}
0 = E[S_\tau] &= E\left[S_{T_a} \mathbb{1}_{\{T_a < T_{-b}\}}\right] + E\left[S_{T_{-b}} \mathbb{1}_{\{T_a \ge T_{-b}\}}\right] \\
&= aP[T_a < T_{-b}] - b(1 - P[T_a < T_{-b}]) = (b + a)P[T_a < T_{-b}] - b,
\end{aligned}
$$

por lo que

$$P[T_a < T_{-b}] = \frac{b}{b+a}.$$

Ahora queremos saber cual es la duración media del juego, es decir queremos calcular $E[\tau]$. Para esto, usaremos el proceso $\{S_n^2 - n\}_{n\in\mathbb{N}}$, que es martingala con respecto a la filtración natural de $\{S_n\}_{n\in\mathbb{N}}$ (la prueba se deja como ejercicio para el lector). Tenemos entonces que $\{S_{n\wedge\tau}^2 - (n\wedge\tau)\}_{n\in\mathbb{N}}$ también es una martingala, por lo que

$$E\left[S_{n\wedge\tau}^2 - (n\wedge\tau)\right] = E\left[S_{1\wedge\tau}^2 - (1\wedge\tau)\right] = E[S_1^2] - 1 = 0.$$

Nuevamente por el Teorema de paro opcional aplicado a esta martingala (¡revisar que se satisfacen las hipótesis!), tenemos que

$$0 = E\left[S_\tau^2 - \tau\right] \Rightarrow E[\tau] = E\left[S_\tau^2\right] = a^2 P[T_a < T_{-b}] + b^2(1 - P[T_a < T_{-b}]),$$

de dónde $E[\tau] = b \times a$.

Ejemplo 90. Si revisitamos el Ejemplo 85, vemos que

$$\mathbb{E}[X_\tau] = 1 \neq 0 = \mathbb{E}[X_1].$$

¿Contradice esto el Teorema de Paro Opcional de Doob? No! Pues vemos que

$$\mathbb{E}[X_n \mathbb{1}_{\tau>n}] = \mathbb{E}\left[-(2^n-1)\mathbb{1}_{\tau>n}\right] = (1-2^n)\mathbb{P}[\tau>n] = (1-2^n)(1-\tfrac{1}{2})^n = \frac{1}{2^n} - 1 \underset{n\to\infty}{\longrightarrow} -1 \neq 0,$$

y por lo tanto no cumple las hipotesis del Teorema. Esto está relacionado con el hecho de que $\mathbb{E}[X_{\tau-1}]$ explota, lo que quiere decir que la media de mi proceso original diverge *más rapido* que el tiempo en el que me detengo.

5.2. Desigualdades de Doob

Los siguientes resultados nos permite encontrar cotas para submartingalas estudiando el máximo valor que toma, y nos brindan un criterio para verificar si una martingala converge, no solo casi seguramente sino también en el espacio $L^2 = L^2(\Omega, \mathscr{F}, P)$, que es el espacio de variables aleatorias cuadrado-integrables.

Proposición 91 (Desigualdad maximal de Doob)**.** *Dada $\{X_n\}_{n\geq 1}$ una (sub)martingala no negativa, definimos $M_n = \max\{X_1, X_2, \dots, X_n\}$. Entonces para todo $\lambda > 0$,*

$$P[M_n \geq \lambda] \leq \frac{E\left[X_n \mathbb{1}_{\{M_n\geq\lambda\}}\right]}{\lambda},$$

y de aquí es directo que $P[M_n \geq \lambda] \leq \dfrac{E[X_n]}{\lambda}$.

Nota. *La cota obvia, obtenida al aplicar la desigualdad de Markov, es que*

$$P[M_n \geq \lambda] \leq \frac{E[M_n]}{\lambda},$$

pero la importancia de la desigualdad maximal de Doob es que de hecho ¡podemos acotar la cola de la distribución del supremo de la (sub)martingala simplemente por la (sub)martingala misma!

Dem. Dado que $\{X_n\}_{n\in\mathbb{N}}$ es una (sub)martingala no negativa, tenemos que

$$E[|X_n|] = E[X_n] < \infty \qquad \text{y} \qquad E[X_{n+1}|X_1,\dots,X_n] \geq X_n.$$

Entonces, dado que $\{M_n \geq \lambda\} = \{\exists k \in \{1,\dots,n\} : X_k \geq \lambda\}$, definimos el tiempo de paro (con respecto a la filtración natural del proceso)

$$\tau := \inf\{k : X_k \geq \lambda\} \qquad \Rightarrow \qquad X_\tau \geq \lambda.$$

Tenemos entonces que

$$\begin{aligned} E[X_n] &= \sum_{k=1}^{n} E\left[X_n \mathbb{1}_{\{\tau=k\}}\right] + E\left[X_n \mathbb{1}_{\{\tau>n\}}\right] = \sum_{k=1}^{n} E\left[\mathbb{1}_{\{\tau=k\}} E[X_n|X_1,\dots,X_k]\right] + E\left[X_n \mathbb{1}_{\{\tau>n\}}\right] \\ &\geq \sum_{k=1}^{n} E\left[\mathbb{1}_{\{\tau=k\}} X_k\right] + E\left[X_n \mathbb{1}_{\{\tau>n\}}\right] = E\left[X_n^\tau\right]. \end{aligned}$$

Por otro lado, se cumple que

$$\begin{aligned} E[X_n] \geq E\left[X_n^\tau\right] &= E\left[X_n^\tau \mathbb{1}_{\{M_n\geq\lambda\}}\right] + E\left[X_n^\tau \mathbb{1}_{\{M_n<\lambda\}}\right] = E\left[X_\tau \mathbb{1}_{\{M_n\geq\lambda\}}\right] + E\left[X_n \mathbb{1}_{\{M_n<\lambda\}}\right] \\ &\geq \lambda E\left[\mathbb{1}_{\{M_n\geq\lambda\}}\right] + E\left[X_n \mathbb{1}_{\{M_n<\lambda\}}\right], \end{aligned}$$

de donde

$$E\left[X_n \mathbb{1}_{\{M_n\geq\lambda\}}\right] = E[X_n] - E\left[X_n \mathbb{1}_{\{M_n<\lambda\}}\right] \geq \lambda P[M_n \geq \lambda],$$

de donde obtenemos la desigualdad deseada. □

Nota. *En particular, si* $\{X_n\}$ *es una martingala no negativa tenemos que*

$$P[M_n \geq \lambda] \leq \frac{E\left[X_n \mathbb{1}_{\{M_n\geq\lambda\}}\right]}{\lambda} \leq \frac{E[X_n]}{\lambda} = \frac{E[X_0]}{\lambda} \quad \forall n \in \mathbb{N}.$$

El resultado anterior puede usarse para encontrar cotas en el espacio L^2, como las siguientes

Proposición 92 (Desigualdad L^2 de Doob)**.** *Dada* $\{X_n\}$ *una submartingala no negativa tal que* $E\left[X_n^2\right] < \infty$ *para todo* $n \in \mathbb{N}$*, se tiene que*

$$E\left[M_n^2\right] \leq 4E\left[X_n^2\right].$$

Dem. Usando el Teorema de Fubini, es fácil probar que para una variable aleatoria no negativa se cumple que

$$E\left[X^2\right] = 2\int_0^\infty y\left(1 - F_X(y)\right) dy.$$

Usando la desigualdad de la Proposición 91 tenemos entonces que

$$\begin{aligned} E\left[M_n^2\right] &= 2\int_0^\infty yP\left[M_n \geq y\right] dy \leq 2\int_0^\infty y\frac{E\left[X_n \mathbb{1}_{\{M_n \geq y\}}\right]}{y} dy \\ &\leq 2\int_0^\infty E\left[X_n \mathbb{1}_{\{M_n \geq x\}}\right] dx = 2E\left[X_n \int_0^\infty \mathbb{1}_{\{M_n \geq x\}} dx\right] \\ &\leq 2E\left[X_n M_n\right], \end{aligned}$$

donde podemos usar desigualdades e intercambiar esperanza e integrales pues todas las funciones son positivas y acotadas. Por la desigualdad de Cauchy-Schwartz en el espacio L^2 tenemos que

$$E\left[M_n^2\right] \leq 2E\left[X_n M_n\right] \leq 2\sqrt{E\left[X_n^2\right]}\sqrt{E\left[M_n^2\right]},$$

de donde

$$E\left[M_n^2\right]^2 \leq 4E\left[X_n^2\right] E\left[M_n^2\right] \Rightarrow E\left[M_n^2\right] \leq 4E\left[X_n^2\right].$$

□

5.3. Convergencia de submartingalas

A partir de las desigualdades anteriores, podemos obtener los principales resultados de convergencia de submartingalas de Doob.

Teorema 93. *Sea* $\{X_n\}$ *una submartingala con respecto a la filtración* $\{\mathcal{G}_n\}$ *tal que existe* $M \in \mathbb{R}_+$ *que satisface*

$$\sup_{n \in \mathbb{N}} E\left[|X_n|\right] \leq M.$$

Entonces existe una variable aleatoria X_∞ *tal que* $X_n \xrightarrow{c.s.} X_\infty$*.*

La demostración de este teorema puede encontrarse en [7, Sección 7.9]

Corolario 94 (Teorema de convergencia casi segura de martingalas)**.** *Dada $\{X_n\}$ una martingala no negativa con respecto a la filtración $\{\mathcal{G}_n\}$, existe una variable aleatoria X_∞ tal que $X_n \xrightarrow{c.s.} X_\infty$.*

Dem. Una martingala es en particular una submartingala, y si es no negativa satisface que $\sup_{n\in\mathbb{N}} E[|X_n|] = E[X_0]$, por lo que tenemos la existencia del límite casi seguro por el teorema anterior. □

Corolario 95. *Dada $\{Y_n\}$ una submartingala no positiva con respecto a la filtración $\{\mathcal{G}_n\}$ (o una supermartingala no negativa), existe una variable aleatoria Y_∞ tal que $Y_n \xrightarrow{c.s.} Y_\infty$.*

Dem. Si $\{Y_n\}$ es una submartingala no positiva, por tener esperanza no decreciente tenemos que

$$E[|Y_n|] = -E[Y_n] \geq -E[Y_{n+1}] = E[|Y_{n+1}|] \qquad \forall n \in \mathbb{N},$$

por lo que $\sup E[|Y_n|] \leq E[|Y_0|]$ y tenemos convergencia por el Teorema anterior. Por otro lado si $\{Y_n\}_{n\in\mathbb{N}}$ es una supermartingala no negativa entonces $\{-Y_n\}$ es una submartingala no positiva, y tenemos que $-Y_n \xrightarrow{c.s.} -Y_\infty$, y por lo tanto $Y_n \xrightarrow{c.s.} Y_\infty$. □

5.4. Modelo Binomial

5.4.1. Modelo Binomial en un paso

Supongamos que tenemos un cierto activo en el mercado, cuyo precio sigue un modelo binomial, es decir dado un valor inicial S_0, tenemos que

$$S_1 = \begin{cases} uS_0 & \text{con probabilidad } p^*, \\ dS_0 & \text{con probabilidad } (1-p^*) \end{cases},$$

lo que podemos interpretar de la siguiente manera: se lanza una moneda, y si sale *águila* - con probabilidad p^* - el precio se incrementa por una razón u (por *"up"*), y si sale *sol*, el precio se decrementa por una razón d (por *"down"*). Además, consideraremos que hay un interés r en el mercado, por lo que para poder comparar el precio al tiempo 1 con el precio al tiempo 0 tenemos que tener en cuenta el *precio descontado*, es decir $\widetilde{S}_1 := \frac{S_1}{(1+r)}$.

Queremos decidir cual es el valor/precio "justo" para un cierto derivado de este activo, por ejemplo una **opción**, es decir un derecho sobre su compra o su venta. Por ejemplo, en el caso de una **opción *call*** en un período, estamos pagando una cierta cantidad V_0 por el derecho/opción a comprar el activo a precio K al tiempo 1. Entonces, el valor de mi opción a tiempo 1 será

$$V_1 = (S_1 - K)^+ \begin{cases} S_1 - k & \text{si } S_1 > K \\ 0 & \text{si } S_1 \leq K \end{cases},$$

porque si $S_1 > K$, me conviene ejercer la opción y comprar el activo a precio K y luego venderla al valor del mercado, ganando la diferencia; pero si $S_1 \leq K$ comprar el activo a precio K es más caro que comprarla directamente a valor del mercado, por lo que decido no hacerlo y me quedo sin nada.

Entonces, nos gustaría encontrar el precio V_0 justo, es decir la constante v_0 tal que

$$\mathbb{E}[V_1] - v_0(1+r) = 0 \qquad \Rightarrow \qquad v_0 = \mathbb{E}\left[\frac{V_1}{(1+r)}\right] = \frac{\mathbb{E}[V_1]}{(1+r)}. \tag{8}$$

Notemos que si el valor descontado $\left(\frac{V_k}{(1+r)^k}, k \in \mathbb{N}_0\right)$ del derivado es una martingala (con respecto a alguna filtración), esto claramente se cumplirá.

Otra forma de pensar en precios justos, es pensar en **estrategias**. Supongamos que yo cuento con una cantidad de dinero v_0 al tiempo cero. Puedo decidir comprar Δ_0 unidades del activo en cuestión, por lo que tendré $\Delta_0 S_0$ de valor en mi activo y $(v_0 - \Delta_0 S_0)$ en líquido. ¿Cual será el valor de lo que poseo al tiempo 1, considerando lo que tengo en líquido al valor presente? Es fácil deducir que tendré una ganancia $v_1 = \Delta_0 S_1 + (v_0 - \Delta_0 S_0)(1+r)$. Quiero entonces encontrar el valor (V_0, Δ_0) que hace que nuestra inversión (en opciones) reproduzca en un paso la *ganancia* del mercado. Para saber que necesitamos para lograr esto, tomemos en cuenta el siguiente resultado.

Proposición 96. *Dado el proceso $(S_k, k \in \mathbb{N}_0)$ y $\{\mathscr{F}_k\}_{k\in\mathbb{N}_0}$ su filtración natural, consideremos una estrategia $(\Delta_k, k \in \mathbb{N}_0)$, tal que Δ_k es $\mathscr{F}_k$-medible. Tenemos entonces que el proceso $\left(\widetilde{V}_k, k \in \mathbb{N}_0\right)$ dado por $\widetilde{V}_k = \frac{V_k}{(1+r)^k}$, dónde*

$$V_{k+1} := \Delta_k S_{k+1} + (V_k - \Delta_k S_k)(1+r) = V_k + \Delta_k \left[S_{k+1} - S_k(1+r)\right],$$

es una martingala con respecto a $\{\mathscr{F}_k\}$ siempre que $E[|V_k|] < \infty$ para todo $k \in \mathbb{N}_0$ y $(\widetilde{S}_k, k \in \mathbb{N}_0)$ sea una martingala con respecto a $\{\mathscr{F}_k\}$, con $\widetilde{S}_k = \frac{S_k}{(1+r)^k}$ el proceso descontado.

Dem. Notemos que V_k es una función de $(S_0, \ldots, S_k)$ y $(\Delta_0, \ldots, \Delta_k)$, por lo que es $\mathscr{F}_k$-medible. Además

$$\begin{aligned}\mathbb{E}\left[V_{k+1}|\mathscr{F}_k\right] &= \mathbb{E}\left[\Delta_k S_{k+1}|\mathscr{F}_k\right] + (1+r)\, E\left[(V_k - \Delta_k S_k)|\mathscr{F}_k\right] \\ &= \Delta_k\, E\left[S_{k+1}|\mathscr{F}_k\right] + (V_k - \Delta_k S_k)(1+r) \qquad , \\ &= \Delta_k\, S_k(1+r) + (V_k - \Delta_k S_k)(1+r) = V_k(1+r)\end{aligned}$$

Pues $\mathbb{E}\left[\frac{S_{k+1}}{(1+r)^{k+1}}\middle|\mathscr{F}_k\right] = \frac{\mathbb{E}[S_{k+1}|\mathscr{F}_k]}{(1+r)^{k+1}} = \frac{S_k}{(1+r)^k}$, y por lo tanto

$$\mathbb{E}\left[\frac{V_{k+1}}{(1+r)^{k+1}}\middle|\mathscr{F}_k\right] = \frac{\mathbb{E}[V_{k+1}|\mathscr{F}_k]}{(1+r)^{k+1}} = \frac{V_k}{(1+r)^k},$$

de donde deducimos lo deseado. $\square$

Entonces, pedir que el valor de mi opción sea justo y reproduzca el comportamiento del mercado equivale a pedir que el precio descontado $(\widetilde{S}_k, k \in \mathbb{N}_0)$ sea una martingala. ¿Cuando este precio será una martingala? Una condicion necesaria para que lo sea es pedir que

$$\mathbb{E}\left[\widetilde{S}_k\right] = \widetilde{S}_0 = S_0,$$

en particular pedir que $E[S_1] = S_0(1+r)$. Pensando que u y d están fijos, para que esto se cumpla debe ser que

$$S_0(1+r) = E[S_1] = uS_0 p^* + dS_0(1-p^*) \quad \Rightarrow \quad S_0(1+r-d) = (u-d)p^* S_0 \quad \Rightarrow \quad p^* = \frac{1+r-d}{u-d},$$

y por lo tanto $1-p^* = \frac{u-(1+r)}{u-d}$. Es importante notar que para que $p^* \in (0,1)$ debe ser que $d < (1+r) < u$. En otro caso no es posible encontrar una probabilidad p^* que haga el precio descontado martingala, y existirá una estrategia que permita ganar sin correr ningún riesgo, lo cual consideramos injusto. En este último caso, decimos que hay **arbitraje**.

En el caso $d < (1+r) < u$, consideremos que tomamos p^* tal que $E[S_1] = S_0(1+r)$. Tenemos entonces que el valor descontado es martingala, y

$$\mathbb{E}[V_1] = \Delta_0 \mathbb{E}[S_1] + (v_0 - \Delta_0 S_0)(1+r) = v_0(1+r),$$

de donde (8) se satisface, y podemos calcular el precio justo v_0 como

$$v_0 = \frac{p^* V_1(A) + (1-p^*)V_1(S)}{1+r}.$$

En el caso de la opción *call*, donde $V_1 = (S_1 - K)^+$, tenemos tres escenarios posibles

1. Cuando $K < dS_0 < uS_0$, entonces
$$v_0 = \frac{p^*(uS_0-K) + (1-p^*)(dS_0-K)}{1+r} = \frac{u-d}{1+r}p^* S_0 + \frac{dS_0}{1+r} - \frac{K}{1+r} = S_0 - \frac{K}{1+r}.$$

2. Cuando $dS_0 \leq K < uS_0$, entonces
$$v_0 = \frac{p^*(uS_0-K)}{1+r} = \frac{(1+r)-d}{1+r}\frac{uS_0-K}{u-d}.$$

3. Cuando $K \geq uS_0$ entonces $v_0 = 0$.

La estrategia que replica el comportamiento del mercado entonces será Δ_0 tal que

$$\begin{cases} V_1(A) &= \Delta_0 S_1(A) + (v_0 - \Delta_0 S_0)(1+r) \\ V_1(S) &= \Delta_0 S_1(S) + (v_0 - \Delta_0 S_0)(1+r), \end{cases}$$

de donde restando las dos ecuaciones anteriores obtenemos que

$$\Delta_0 = \frac{V_1(A) - V_1(S)}{S_1(A) - S_1(S)} = \frac{(uS_0 - K)^+ - (dS_0 - K)^+}{uS_0 - dS_0},$$

que podemos calcular en los tres casos anteriores. El caso de la opción *put* es análogo, teniendo en cuenta que en este caso $V_1 = (K - S_1)^+$.

5.4.2. Modelo Binomial en N pasos

Consideremos ahora el precio de nuestro activo hasta un tiempo N. Dado S_0, definimos nuevamente

$$S_{n+1} = X_{n+1} S_n, \qquad n \in \{0, 1, \ldots, N-1\}.$$

dónde $\{X_k\}_{k\in\mathbb{N}}$ son variables i.i.d. tal que $p^* = p_X(u) = 1 - p_X(d)$. Nuevamente, consideramos una tasa de interés del mercado de r, y por lo tanto el precio descontado de nuestro activo a tiempo n será $\widetilde{S}_n = \frac{S_n}{(1+r)^n}$.

Si tenemos ahora una opción a plazo N, como una *call*o una *put*, conocemos cual será el valor de nuestra opción a tiempo N como función del precio (aleatorio) S_N. Por ejemplo, sabemos que $V_N = (S_N - K)^+$ en el caso de una opción call. ¿Cual será el valor de la opción en un tiempo cualquiera, es decir el valor del proceso $\{V_n\}_{n=0}^N$? Pensemos el problema en reversa. Si sabemos que sucedió con los precios hasta tiempo n, queremos que el valor de la opción al tiempo n sea justo, es decir que

$$\mathbb{E}\left[\frac{V_N}{(1+r)^N}\middle|\mathscr{F}_n\right] = \frac{V_n}{(1+r)^n} \quad \Rightarrow \quad \mathbb{E}\left[V_N\middle|\mathscr{F}_n\right] = V_n(1+r)^{N-n}, \quad \{0, 1, \ldots, N-1\},$$

por lo que buscamos que el valor descontado sea una martingala. En este caso es claro que

$$v_0 = \mathbb{E}\left[\frac{V_N}{(1+r)^N}\right].$$

¿En que escenario el precio descontado es martingala? Como en el caso en un paso, vemos que si al tiempo n estamos en el evento $E \in \mathscr{F}_n$, la cantidad de unidades de nuestro activo que necesitamos para replicar la ganancia de la opción (y financiar nuestra situación) es $\Delta_n(E)$, y se cumple que

$$\begin{cases} V_{n+1}(EA) & = \Delta_n(E) S_{n+1}(EA) + (V_n(E) - \Delta_n(E) S_n(E))(1+r) \\ V_{n+1}(ES) & = \Delta_n(E) S_{n+1}(ES) + (V_n(E) - \Delta_n(E) S_n(E))(1+r) \end{cases}.$$

En general, si definimos el valor de la opción al tiempo $n+1$ como

$$V_{n+1} = V_n + \Delta_n \left[S_{n+1} - S_n(1+r)\right], \qquad n \in \{0, 1, \ldots, N-1\},$$

por Proposición 96 este proceso (descontado) será martingala si el precio (descontado) lo es. Debe ser entonces que si definimos $p_n^*(E) := \mathbb{E}\left[EA\middle|E\right]$, tenemos que el precio es martingala siempre que

$$S_n(E)(1+r) = \mathbb{E}\left[S_{n+1}\middle|E\right] = p_n^*(E)S_{n+1}(EA)+(1-p_n^*(E))S_{n+1}(ES) \quad \Rightarrow \quad p_n^*(E) = \frac{(1+r)S_n(E) - S_{n+1}(ES)}{S_{n+1}(EA) - S_{n+1}(ES)}.$$

En el caso particular del modelo binomial con el que estamos trabajando, es claro que $p_n^*(E) = p^*$ para todo $E \in \mathscr{F}_n$. Se cumple entonces que

$$\mathbb{E}\left[V_{n+1}\middle|E\right] = V_n(E)(1+r) \quad \Rightarrow \quad V_n(E) = \frac{p^* V_{n+1}(EA) - (1-p^*)V_{n+1}(ES)}{1+r},$$

lo que nos da un sistema que podemos resolver de manera recursiva en reversa, empezando con $n = N-1$, con $E \in \mathscr{F}_n$ definidos a partir de una variable aleatoria binomial $Y_{N-1} \sim Bi(N-1, p^*)$, donde $Y_{N-1} = \#\{n \in \{1, \dots, N-1\} : X_n = u\}$. En particular, queremos que $v_0 = E[V_N]$, por lo que debe ser

$$v_0 = \sum_{i=0}^{N} \binom{N}{i} {p^*}^i (1-p^*)^{(1-i)} V_N(s_i), \qquad \text{donde } s_i := u^i d^{N-i}.$$

6. Movimiento Browniano

Definición 97. *Un **movimiento browniano estándar** $(B_t, t \geq 0)$ es un proceso estocástico a tiempo continuo que satisface*

I. $B_0 = 0$,

II. $t \mapsto B_t$ *es una función continua c.s.,*

III. *para toda sucesión de tiempos* $0 < t_1 < t_2 < \cdots < t_n$ *las variables* $\{B_{t_i+1} - B_{t_i}\}_{i=1}^{n-1}$ *son independientes.*

IV. *para todo* $t, s \geq 0$ $B_{t+s} - B_s \sim \mathcal{N}(0, t)$.

Una generalización de la definición anterior es la siguiente.

Definición 98. *Un **movimiento browniano** $(W_t, t \geq 0)$ es un proceso estocástico a tiempo continuo dado por*

$$W_t = w_0 + \mu t + \sigma B_t, \qquad t \geq 0,$$

dónde $(B_t, t \geq 0)$ es un MB estándar.

En esta última definición, w_0 es el valor inicial del proceso, μ es la *deriva* o drift y σ^2 es la varianza o *volatilidad*. Podemos deducir entonces que $W_t \sim \mathcal{N}(w_0 + \mu t, \sigma^2 t)$, y mantiene la propiedad ser un proceso continuo con incrementos estacionarios e independientes.

Nota. *La familia de procesos estocásticos cádlág*

Proposición 99. *Dado $(B_t, t \geq 0)$ un MB estándar, se cumple que para todo par de tiempos $0 < t_1 < t_2$ y par de borelianos A_1, A_2 integrables, entonces*

$$P\left[B_{t_1} \in A_1, B_{t_2} \in A_2\right] = \int_{A_1}\int_{A_2} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(y-x)^2}{2(t_2-t_1)}} dy \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{x^2}{2t_1}} dx,$$

Dem. Gracias a las condiciones III. y IV., tenemos que

$$\begin{aligned}
P\left[B_{t_1} \in A_1, B_{t_2} \in A_2\right] &= \int_{A_1} P\left[(B_{t_2} - B_{t_1}) + B_{t_1} \in A_2 \middle| B_{t_1} = x\right] f_{B_{t_1}}(x) dx \\
&= \int_{A_1} P\left[x + (B_{t_2} - B_{t_1}) \in A_2\right] \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{x^2}{2t_1}} dx \\
&= \int_{A_1}\int_{A_n} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(y-x)^2}{2(t_2-t_1)}} dy \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{x^2}{2t_1}} dx,
\end{aligned}$$

donde usamos que si $B_{t_2} - B_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$, entonces $x + \left(B_{t_2} - B_{t_1}\right) \sim \mathcal{N}(x, t_2 - t_1)$. $\square$

<u>**Ejercicio**</u>: Usando el resultado anterior, probar que las condiciones III. y IV. en la definición de MB estándar son equivalentes a

V. para toda sucesión de tiempos $0 < t_1 < t_2 < \cdots < t_n$ y borelianos $A_1, A_2, \ldots, A_n$ integrables, se cumple que

$$P\left[B_{t_1} \in A_1, \cdots, B_{t_n} \in A_n\right] = \int_{A_1} \cdots \int_{A_n} p(t_1, 0, x_1) \ldots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_n \ldots dx_1,$$

dónde $p(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}$.

6.1. Caracterización de Paul Lévy

En el caso de los procesos a tiempo y estados continuos, podemos definir una martingala de manera análoga a la definición 76, solo que en este caso la filtración natural será $\{\mathcal{F}_t\}_{t\geq 0}$, dada por $\mathcal{F}_t = \sigma\{X_s, s \leq t\}$.

Proposición 100. *Un MB estándar es una martingala continua (con respecto a su filtración natural). Además, el proceso $(M_t, t \geq 0)$ definido por $M_t = B_t^2 - t$ para todo $t \geq 0$ es también una martingala con respecto a la filtración natural del MB.*

Dem.

Claramente

$$E\left[|B_s|\right] = 2E\left[B_s \mathbb{1}_{B_s > 0}\right] = \frac{2\sqrt{s}}{\sqrt{2\pi}} < \infty$$

por ser $B_s \sim \mathcal{N}(0, s)$ simétrica. Por otro lado

$$E\left[B_t | \mathcal{F}_s\right] = E\left[B_t - B_s + B_s | \mathcal{F}_s\right] = E\left[B_t - B_s\right] + B_s = B_s,$$

dónde usamos los incrementos independientes del MB y el hecho que $B_t - B_s \sim \mathcal{N}(0, t-s)$. Que el proceso $(M_t, t \geq 0)$ es martingala con respecto a la filtración natural del MB se deja de ejercicio para el lector. $\square$

Esta última proposición <u>caracteriza</u> a un MB estándar.

Teorema 101 (Caracterización de Paul Lévy)**.** *Un proceso estócastico $(X_t, t \geq 0)$ que satisface las siguientes condiciones* *i.* $t \mapsto X_t$ *es continua c.s.,*

ii. $X_0 = 0$, *y*

iii. $(X_t, t \geq 0)$ *y* $(X_t^2 - t, t \geq 0)$ *son martingalas con respecto a* $\{\mathscr{F}_t\}$ *la filtración natural del proceso*

es un Movimiento Browniano estándar.

Este resultado está relacionado también con como se obtiene el Movimiento Browniano a partir de convergencia de variables aleatorias. Dada $\{Y_k\}_{k\in\mathbb{N}}$ una sucesión de variables independientes con esperanza μ y varianza σ^2 finita. Entonces, sabemos por el Teorema Central del Límite que $\frac{\sum_{k=1}^{n} Y_k - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathscr{D}} Z \sim \mathcal{N}(0,1)$. En particular, para cada $t > 0$, si tomamos la subsucesión de variables $\{Y_{-\lfloor kt \rfloor}\}_{k\in\mathbb{N}}$, nuevamente por TCL, tenemos que

$$\frac{\sum_{k=1}^{\lfloor nt \rfloor} Y_k - \lfloor nt \rfloor \mu}{\sigma\sqrt{\lfloor nt \rfloor}} \xrightarrow{\mathscr{D}} Z \sim \mathcal{N}(0,1) \qquad \Rightarrow \qquad \frac{\sum_{k=1}^{\lfloor nt \rfloor} Y_k - \lfloor nt \rfloor \mu}{\sigma\sqrt{n}} \xrightarrow{\mathscr{D}} Z \sim \mathcal{N}(0,1)$$

Si queremos ver que pasa como proceso, tenemos el siguiente resultado

Teorema 102 (Principio de invariancia de Donsker)**.**

$$\frac{\sum_{i=1}^{\lfloor nt \rfloor} Y_i - \lfloor nt \rfloor \mu}{\sigma\sqrt{n}} \xrightarrow{\mathscr{D}} B_t \qquad \forall t \geq 0,$$

dónde $(B_t, t \geq 0)$ *es un MB estándar, y esta convergencia es localmente uniforme en* t.

6.2. Principio de Reflexión

Estamos interesados ahora en estudiar la probabilidad de que el Movimiento Browniano sobrepase cierta “barrera”. Probaremos que la distribución del máximo del MB hasta un tiempo dado coincide con la de su valor absoluto en ese instante. Esto se debe a que la simetría del MB nos permite usar cualquier barrera como un espejo, ya que por cada trayectoria que pasa por arriba de dicha barrera en un tiempo dado hay otra que pasa por abajo y es exactamente la reflexión de la anterior.

Teorema 103 (Principio de Reflexión)**.** *Dado* $(W_t^a, t \geq 0)$ *un MB empezando en* $a > 0$,

$$P\left[\max_{s\in[0,t]} W_s^a \geq b\right] = P\left[W_s^a \geq b \text{ para algún } s \in [0,t]\right] = 2P\left[W_t^a \geq b\right], \qquad b > a, \quad t \geq 0.$$

Como $W_t^a = B_t + a$ *para algún* B *MB estándar, lo anterior equivale a*

$$P\left[\max_{s\in[0,t]} B_s \geq b - a\right] = 2P\left[B_t \geq b - a\right] = P\left[|B_t| \geq b - a\right]$$

Dem. Sea $\tau_b = \text{ínf}\{t \geq 0 : W_s^a = b\}$ [13], es fácil ver que es un tiempo de paro finito con respecto a la filtración natural del proceso. Tenemos entonces que el proceso $\left(W_{\tau_b+t}^a - W_{\tau_b}^a, t \geq 0\right)$ es un Movimiento Browniano estándar independiente de $\mathscr{F}_{\tau_b}$, y por lo tanto

$$\begin{aligned}
P\left[W_s^a \geq b \text{ para algún } s \in [0,t]\right] &= P[\tau_b \leq t] \\
&= P\left[\tau_b \leq t, W_t^a < b = W_{\tau_b}^a\right] + P\left[\tau_b \leq t, W_t^a \geq b = W_{\tau_b}^a\right] \\
&= P\left[\tau_b \leq t, W_t^a - W_{\tau_b}^a < 0\right] + P\left[\tau_b \leq t, W_t^a - W_{\tau_b}^a \geq 0\right] \\
&= 2P\left[\tau_b \leq t, W_t^a - W_{\tau_b}^a \geq 0\right] = 2P\left[\tau_b \leq t, W_t^a \geq b\right] \\
&= 2P\left[W_t^a \geq b\right],
\end{aligned}$$

donde la ultima igualdad se debe a que $\{\tau_b \leq t\} \subseteq \{W_t^a \geq b\}$. □

Ejercicio: Probar que

$$P[|B_t| > b-a] = \int_0^t \frac{(b-a)}{\sqrt{2\pi y^3}} e^{-(b-a)^2/2y} dy \qquad \text{para todo } b > a > 0.$$

6.3. Recurrencia y Transitoriedad

Usando el principio de reflexión se puede probar que dado $0 < t_1 < t_2$ cualquier par de reales positivos, entonces

$$P\left[B_s = 0 \text{ para algún } s \in [t_1, t_2]\right] = 1 - P[B_s \neq 0\, \forall s \in [t_1, t_2]] = 1 - \frac{2}{\pi} \arctan\left(\frac{\sqrt{t_1}}{\sqrt{t_2 - t_1}}\right).$$

En efecto

$$\begin{aligned}
P\left[B_s = 0 \text{ para algún } s \in [t_1, t_2]\right] &= 2\int_0^\infty P\left[B_s \leq -u \text{ para algún } s \in [0, t_2 - t_1]\right] f_{B_{t_1}}(u) du \\
&= 2\int_0^\infty P\left[\tilde{B}_s \geq u \text{ para algún } s \in [0, t_2 - t_1]\right] f_{B_{t_1}}(u) du \\
&= 2\int_0^\infty \int_0^{t_2 - t_1} \frac{u}{\sqrt{2\pi y^3}} e^{-u^2/2y} dy \frac{du}{\sqrt{2\pi t_1}} e^{-u^2/2t_1} \\
&= \frac{1}{\pi} \int_0^{t_2 - t_1} \frac{1}{\sqrt{y^3 t_1}} \frac{t_1 y}{t_1 + y} dy = \frac{2}{\pi} \int_0^{\frac{\sqrt{t_2 - t_1}}{\sqrt{t_1}}} \frac{dw}{1 + w^2},
\end{aligned}$$

donde en la ultima igualdad hicimos el cambio de variable $w = \sqrt{y/t_1}$. Se cumple entonces que

$$P\left[B_s = 0 \text{ para algún } s \in [t_1, t_2]\right] = \frac{2}{\pi} \int_0^{\frac{\sqrt{t_2 - t_1}}{\sqrt{t_1}}} \frac{dw}{1 + w^2} = \frac{2}{\pi} \arctan\left(\frac{\sqrt{t_2 - t_1}}{\sqrt{t_1}}\right),$$

de donde concluimos lo deseado usando la igualdad trigonométrica $\arctan x + \arctan 1/x = \pi/2$.

[13] podemos tomar la igualdad por ser un proceso continuo c.s.

De aquí deducimos que el proceso "toca" infinitamente el cero una vez que sale de él, como afirmamos en el siguiente resultado.

Teorema 104. *Dado $t > 0$, tenemos que*

$$P\left[B_s = 0 \text{ para algún } s \in (0,t]\right] = 1 \qquad \forall\, t > 0,$$

de dónde $\tau_0 = \text{ínf}\{t > 0 : B_t = 0\} = 0$ c.s.

Dem. Sea $\{\varepsilon_k\}_{k\in\mathbb{N}}$ una sucesión de reales menores que t tal que $\varepsilon_k \to 0$. Si notamos que los eventos $\{B_s = 0 \text{ para algún } s \in [\varepsilon_k, t]\}$ son crecientes en k, usando continuidad de la probabilidad y el resultado anterior tenemos que

$$P\left[B_s = 0 \text{ para algún } s \in (0,t]\right] = \lim_{k\to\infty} P\left[B_s = 0 \text{ para algún } s \in [\varepsilon_k, t]\right] = 1 - \lim_{k\to\infty} \tfrac{2}{\pi}\arctan\left(\tfrac{\sqrt{\varepsilon_k}}{\sqrt{t-\varepsilon_k}}\right),$$

de donde concluimos lo deseado. □

Nota. *Notemos que lo anterior implica que*

$$P\left[W_s^a = a \text{ para algún } s \in (0,t]\right] = 1 \qquad \forall\, t > 0,$$

pues $(W_t^a - a, t \geq 0)$ también es un MB estándar, y podemos afirmar que el resultado anterior vale para cualquier nivel que visita el proceso.

6.4. Propiedades Trayectoriales del MB

Revisaremos ahora algunas propiedades que tienen las trayectorias del Movimiento Browniano, es decir, cual es su comportamiento como función de t. Primero probaremos que en cualquier intervalo real, por más pequeño que sea, las trayectorias del MB tienen *subidas* y *bajadas* casi seguramente.

Proposición 105. *Dado $[a,b] \subset (0,+\infty)$, con probabilidad 1 las trayectorias $(B_t, a \leq t \leq b)$ no son monótonas.*

Dem. Si el MB es monótono en el intervalo $[a,b]$, debe ser que

$$B_t - B_s \geq 0 \quad \forall\, a \leq s \leq t \leq b, \qquad \text{ó} \qquad B_t - B_s \leq 0 \quad \forall\, a \leq s \leq t \leq b.$$

Esto implica que para toda partición $a = a_0 < a_1 < ... < a_n = b$ del intervalo $[a, b]$ de paso $\delta_i = 1/n$, debe cumplirse que

$$A_n := \left\{B_{a_{i+1}} - B_{a_i} \geq 0 \quad \forall\, i \in \{1, ..., n\}\right\}, \qquad \text{ó} \qquad A_n^{\star} := \left\{B_{a_{i+1}} - B_{a_i} \leq 0 \quad \forall\, i \in \{1, ..., n\}\right\}.$$

Tenemos entonces que

$$P[A_n] = P\left[B_{a_{i+1}} - B_{a_i} \geq 0 \quad \forall\, i \in \{1, ..., n\}\right] = \left(P\left[B_{1/n} \geq 0\right]\right)^n = \tfrac{1}{2^n},$$

donde usamos incrementos estacionarios e independientes. Por simetría tenemos también que $P[A_n^{\star}] = \frac{1}{2^n}$

Como queremos que el evento A_n (resp. $A_n^{\star}$) se cumpla para todo tamaño de partición $1/n$ con $n \in \mathbb{N}$, calcularemos $P\left[\bigcap_{n\in\mathbb{N}} A_n\right]$. Como se cumple que $A_{n+1} \subset A_n$, por continuidad de la proba tenemos que

$$P\left[\bigcap_{n\in\mathbb{N}} A_n\right] = \lim_{n\to\infty} P[A_n] = 0,$$

y análogamente $P\left[\bigcap_{n\in\mathbb{N}} A_n^{\star}\right] = 0$, por lo que, con probabilidad 1, el MB no puede ser monótono en $[a, b]$.

□

El resultado anterior nos lleva a intuir que las trayectorias del MB no pueden ser suaves. En este sentido, probaremos que, con probabilidad 1, estas trayectorias no pueden ser localmente Lipschitz en ningún punto, y por lo tanto no pueden ser diferenciables. [14]

Proposición 106. *Con probabilidad* 1, *las trayectorias del MB no son localmente Lipschitz en ningún punto* $t_0 \in [0, \infty]$. *En particular, esto implica que casi seguramente las trayectorias del MB no son diferenciables en ningún punto.*

Dem. Sin perder generalidad, probaremos que $P[E_n^c] = 0$ para todo $c \in \mathbb{R}_+$, $n \in \mathbb{N}$, donde E_n^c corresponde al evento

$$E_n^c := \{\exists s \in [0,1] : |B_t - B_s| \leq c|t - s| \forall\, t \in B(s, 3/n)\}.^{15}$$

Empezaremos por probar que, para ciertos eventos M_n^c, $E_n^c \subset M_n^c$ para todo $c \in \mathbb{R}_+$, $n \in \mathbb{N}$, con $\lim_{n\to\infty} P\left[M_n^c\right] = 0$.

[14]Recuerde que si una función es localmente Lipschitz en un punto, entonces será continua allí; y si una función es diferenciable en un punto, debe ser localmente Lipschitz allí.

[15]Esto coincide con probar lo que queremos solo en el intervalo $[0, 1]$, pero esta demostración se extiende directamente a cualquier otro subintervalo de $[0, \infty]$.

Si definimos

$$M_n^c := \left\{\exists k \in \{1,\dots,n-2\} : \text{máx}\left\{|B_{k/n} - B_{(k-1)/n}|, |B_{(k+1)/n} - B_{k/n}|, |B_{(k+2)/n} - B_{(k+1)/n}|\right\} \le \tfrac{5c}{n}\right\},$$

podemos ver que si existe $s \in [0,1]$ tal que $|B_t - B_s| \le c|t-s|$ para todo $t \in B(s, 3/n)$, entonces existe un $k_s \in \{1,\dots,n-2\}$ que satisface $s \in [\frac{k_s-1}{n}, \frac{k_s}{n}]$. Se cumple entonces que $[\frac{k_s-1}{n}, \frac{k_s+2}{n}] \subset B(s,3/n)$ por lo que

$$|B_{k_s/n} - B_{(k_s-1)/n}| \le |B_{k_s/n} - s| + |B_s - B_{(k_s-1)/n}| \le c2/n,$$
$$|B_{(k_s+1)/n} - B_{k_s/n}| \le |B_{(k_s+1)/n} - B_s| + |B_{k_s/n} - B_s| \le c3/n,$$
$$|B_{(k_s+2)/n} - B_{(k_s+1)/n}| \le |B_{(k_s+2)/n} - B_s| + |B_{(k_s+1)/n} - B_s| \le c5/n,$$

y por lo tanto

$$\text{máx}\left\{|B_{k_s/n} - B_{(k_s-1)/n}|, |B_{(k_s+1)/n} - B_{k_s/n}|, |B_{(k_s+2)/n} - B_{(k_s+1)/n}|\right\} \le \tfrac{5c}{n},$$

que implica que $E_n^c \subset M_n^c$ y por lo tanto $P\left[E_n^c\right] \le P\left[M_n^c\right]$. Por definición, tenemos que

$$\begin{aligned}
P\left[M_n^c\right] &= P\left[\bigcup_{k=1}^{n-2} \text{máx}\left\{|B_{k/n} - B_{(k-1)/n}|, |B_{(k+1)/n} - B_{k/n}|, |B_{(k+2)/n} - B_{(k+1)/n}|\right\} \le \tfrac{5c}{n}\right] \\
&= \sum_{k=1}^{n-2} P\left[|B_{1/n}| \le \tfrac{5c}{n}\right]^3 = (n-2)P\left[\sqrt{n}|B_{1/n}| \le \sqrt{n}\tfrac{5c}{n}\right]^3 \\
&\le nP\left[|Z| \le \tfrac{5c}{\sqrt{n}}\right] = 2\Phi(\tfrac{5c}{\sqrt{n}}) - 1
\end{aligned}$$

para $Z \sim \mathcal{N}(0,1)$, donde hemos usado que los incrementos $\{B_{k/n} - B_{(k-1)/n}\}_{k=1}^{n-2}$ son estacionarios e independientes con distribución $\mathcal{N}(0,1/n)$. Tenemos entonces que

$$\lim_{n\to\infty} P\left[M_n^c\right] \le \lim_{n\to\infty} 2\Phi(\tfrac{5c}{\sqrt{n}}) - 1 = 2\Phi(0) - 1 = 0,$$

para todo $c > 0$, y por lo tanto

$$P\left[\bigcup_{n=1}^{\infty} E_n^c\right] = \lim_{n\to\infty} P\left[E_n^c\right] = 0,$$

ya que la sucesión $\{E_n^c\}_{n\in\mathbb{N}}$ es creciente, ya que si una trayectoria es localmente lipschitz en un entorno $B(s,3/n)$ para algún $s \in [0,1]$, entonces también lo será en $B(s,3/(n+1))$. Podemos concluir entonces que, para todo $n \in \mathbb{N}$

$$P\left[E_n^c\right] \le P\left[\bigcup_{n=1}^{\infty} E_n^c\right] = 0,$$

y obtenemos el resultado deseado. □

6.5. No diferenciabilidad del MB

Proposición 107. *Dado $t_0 \in \mathbb{R}_+$ un tiempo fijo, el MB $(B_t, t \geq 0)$ no es diferenciable en t_0 casi seguramente.*

Dem. Dado que el proceso tiene incrementos estacionarios e independientes, es suficiente con probar el resultado para $t_0 = 0$. Queremos probar que con probabilidad 1 no existe $\frac{dB}{dt}|_{t=0}$, en particular

$$\limsup_{n\to\infty} \frac{B_{1/n}}{1/n} = +\infty \qquad \text{y} \qquad \liminf_{n\to\infty} \frac{B_{1/n}}{1/n} = -\infty.$$

Para ver lo anterior basta ver que si definimos $A_n := \left\{ |\frac{B_t}{t}| > n \text{ para algún } t \in [0, n^{-2}] \right\}$, tenemos que $P[A_n] = 1$ para todo $n \in \mathbb{N}$.

Claramente $A_{n+1} \subseteq A_n$, por lo que $P\left[\bigcap_{n\in\mathbb{N}} A_n\right] = \lim_{n\to\infty} P[A_n]$. Dado $n \in \mathbb{N}$, tomemos $t = n^{-4}$. Entonces

$$P[A_n] \geq P\left[|\frac{B_{n^{-4}}}{n^{-4}}| > n\right] = P\left[|B_{n^{-4}}| > n^{-3}\right] = P\left[|n^2 B_{n^{-4}}| > n^{-1}\right] = P\left[|B_1| > n^{-1}\right],$$

y por lo tanto

$$1 \geq \lim_{n\to\infty} P[A_n] \geq \lim_{n\to\infty} P\left[|B_1| > n^{-1}\right] = 1,$$

de dónde $P\left[\bigcap_{n\in\mathbb{N}} A_n\right] = 1$, y entonces $P[A_n] = 1$ para todo $n \in \mathbb{N}$. □

6.6. Variación del Movimiento Browniano

Si queremos calcular sumas del tipo Riemann-Stieltjes con respecto a cierta función continua, es importante conocer que pasa con los incrementos de la misma. En esa dirección, definimos

Definición 108. *Dada $g : [a, b] \to \mathbb{R}$ una función continua, la* ***variación total*** *de g en $[a, b]$ está dada por*

$$V_a^b(g) := \limsup_{n\to 0} \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)|,$$

donde el límite se toma sobre $\{\mathscr{P}_n, n \geq 1\}$, el conjunto de particiones finitas de $[a, b]$.

Notemos que, dado que g es continua, es claro que $|g(t_{i+1}) - g(t_i)| \xrightarrow[n\to\infty]{} 0$, pero no podemos asegurar que la serie no explote. Si $V_a^b(g) < \infty$, decimos que g tiene **variación finita** en $[a, b]$.

Si $V_a^b(g) = \infty$ decimos que g no tiene variación finita, pues sus incrementos no son sumables. En este caso tiene sentido estudiar los incrementos cuadráticos.

Definición 109. *Dada $g : [a,b] \to \mathbb{R}$ una función continua, la* ***variación cuadrática*** *de g en $[a,b]$ está dada por*

$$V_a^{b^2}(g) := \limsup_{n\to 0} \sum_{i=0}^{n-1} \left(g(t_{i+1}) - g(t_i)\right)^2,$$

donde el límite se toma sobre $\{\mathscr{P}_n, n \geq 1\}$, el conjunto de particiones finitas de $[a,b]$.

Si pensamos en las trayectorias del MB como una función del tiempo, podemos estudiar su variación y su variación cuadrática. Empezaremos con la variación cuadrática y, dado que esta es finita, deduciremos que el MB no tiene variación finita.

Proposición 110. *Para todo intervalo $[a,b]$*

$$V_a^{b^2}(B) := \limsup_{n\to\infty} \sum_{i=1}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2 = b - a,$$

en L^2, es decir la ***variación cuadrática*** *del movimiento browniano sobre un intervalo es finita y coincide con la longitud del intervalo.*

Dem. Dado el espacio L^2, queremos probar que

$$E\left[|\sum_{i=1}^{n-1} (\Delta B_i)^2 - (b-a)|^2\right]^{1/2} \underset{n\to 0}{\longrightarrow} 0.$$

Tenemos entonces que

$$\begin{aligned}
E\left[|\sum_{i=1}^{n-1} (\Delta B_i)^2 - (b-a)|^2\right] &= E\left[\left(\sum_{i=1}^{n-1} (\Delta B_i)^2\right)^2\right] - 2\,(b-a)\,E\left[\sum_{i=1}^{n-1} (\Delta B_i)^2\right] + (b-a)^2 \\
&= \sum_{i=1}^{n-1} E\left[(\Delta B_i)^4\right] + \sum_{i=1}^{n-1}\sum_{j\neq i} E\left[(\Delta B_i)^2\right] E\left[\left(\Delta B_j\right)^2\right] - 2\,(b-a) \sum_{i=1}^{n-1} E\left[(\Delta B_i)^2\right] + (b-a)^2 \\
&= \sum_{i=1}^{n-1} 3(\Delta t_i)^2 + \sum_{i=1}^{n-1}\sum_{j\neq i} \Delta t_i \Delta t_j - 2(b-a) \sum_{i=1}^{n-1} \Delta t_i + (b-a)^2 \\
&= \sum_{i=1}^{n-1} 2(\Delta t_i)^2 + \left(\sum_{i=1}^{n-1} (\Delta t_i)^2 + \sum_{i=1}^{n-1}\sum_{j\neq i} \Delta t_i \Delta t_j\right) - 2(b-a)^2 + (b-a)^2 = 2\sum_{i=1}^{n-1} (\Delta t_i)^2,
\end{aligned}$$

dónde usamos que $\Delta B_i \sim \mathcal{N}(0, \Delta t_i)$, y por lo tanto

$$E\left[|\sum_{i=1}^{n-1} (\Delta B_i)^2 - (b-a)|^2\right] = 2\sum_{i=1}^{n-1} (\Delta t_i)^2 \leq 2\,(\text{máx}\,\Delta t_i) \sum_{i=1}^{n-1} (\Delta t_i) = 2\left(\max_{i=1\ldots n-1} \Delta t_i\right)(b-a) \underset{n\to\infty}{\longrightarrow} 0.$$

□

El resultado anterior implica que la **variación** es infinita, como establecemos a continuación.

Proposición 111.

$$\limsup_{\Delta t\to\infty} \sum_{i=1}^{n-1} |B_{t_{i+1}} - B_{t_i}| = \infty \qquad c.s.$$

Dem. Tenemos que

$$\sum_{i=1}^{n-1} (\Delta B_i)^2 \le \left(\max_{i=1\ldots n-1} |\Delta B_i|\right) \sum_{i=1}^{n-1} |\Delta B_i|,$$

y como $\sum_{i=1}^{n-1} (\Delta B_i)^2 \xrightarrow{L^2} b-a$, entonces existe una subsucesión de particiones $\{\mathscr{P}_{n_k}\}$ tal que $\sum_{i=1}^{n_k-1} (\Delta B_i)^2 \xrightarrow{c.s.} b-a$. Además, por continuidad c.s. del MB tenemos que $\lim_{k\to\infty} \left[\max_{n_k} |\Delta B_i|\right] = 0$, de dónde $\sum_{i=1}^{n_k-1} |\Delta B_i|$ no puede converger a una constante, y por lo tanto el límite superior de la sucesión debe ser infinito. □

Nota. *En la demostración anterior usamos el hecho de que la convergencia en L^2 implica convergencia en probabilidad, y por lo tanto implica convergencia c.s. de una subsucesión. La primera implicación se obtiene simplemente utilizando la Desigualdad de Markov, y la segunda tomando una subsucesión diagonal conveniente (pueden consultar la Proposición 79 del Apunte de Probabilidad II).*

6.7. Modelo de Black-Scholes para valoración de opciones

Pensemos ahora en el análogo a tiempo continuo del modelo Binomial. Supongamos que tenemos un proceso de precios $(X_t, t \ge 0)$ donde X_t representa el precio de un cierto activo a tiempo t, con precio inicial $X_0 = x_0$. Si asumimos que existe un factor α de descuento (continuo), tenemos que el *valor presente* del precio del activo al tiempo t está dado por $e^{-\alpha t} X_t$ [16]. Supongamos nuevamente que estamos interesados en pagar por el derecho de comprar/vender a cierto valor fijo K dicho activo en un tiempo futuro (fijo) T ¿Cual será el precio justo a pagar por ese derecho? En el caso de una *opción call*, el derecho es a comprar, por lo que el valor presente de nuestra ganancia a tiempo T estará dada por

$$e^{-\alpha T} V_T = e^{-\alpha T}(X_T - K)^+ = e^{-\alpha T}(X_T - K)\mathbb{1}_{\{X_T \ge K\}}$$

Para que el precio v_0 al tiempo 0 sea justo, debe ser que

$$\mathbb{E}\left[e^{-\alpha T} V_T - v_0\right] = 0 \quad \Rightarrow \quad v_0 = \mathbb{E}\left[e^{-\alpha T}(X_T - K)^+\right]$$

Como en el caso discreto, si el proceso descontado $(e^{-\alpha t} V_t, t \ge)$ es una martingala con respecto a $(\mathscr{F}_t, t \ge 0)$ (la filtración natural del proceso de precios) tenemos que

$$v_0 = \mathbb{E}\left[e^{-\alpha T} V_T\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\alpha T} V_T \middle| \mathscr{F}_t\right]\right] = \mathbb{E}\left[e^{-\alpha t} V_t\right] \qquad \forall t \in [0, T],$$

[16] esto corresponde al límite del proceso descontado a tiempo caso discreto: si α es la tasa de interés por unidad de tiempo, y tomamos una partición de tamaño $1/n$ es claro que para t unidades de tiempo se cumplirá $(1 + \frac{\alpha}{n})^{-nt} \underset{n\to\infty}{\longrightarrow} e^{-\alpha t}$

lo que nos permite sacar el precio justo de la opción en cualquier instante intermedio.

Como en el caso discreto, queremos que este valor v_0 cumpla que no hay ninguna estrategia que no lleve a una ganancia segura. Asumiendo el **Teorema del arbitraje** (buscar la referencia) tenemos que no hay ganancia segura si y solo si en el espacio de probabilidad en el que estamos trabajando el proceso de precios traído al valor presente es martingala, en particular

$$\mathbb{E}\left[e^{-\alpha T}X_T\middle|\mathscr{F}_t\right] = e^{-\alpha t}X_t \qquad \text{para todo} t \le T.$$

Pensando en el límite de escala del caso Binomial, teníamos que $S_n = x_0 \prod_{k=1}^{n} X_k$, con $\{X_k\}$ variables aleatorias i.i.d. Podemos escribir entonces $S_n = x_0 e^{\sum_{k=1}^{n} Y_k}$, con $\{Y_k = lnX_k\}$ vs. as. i.i.d. Si estas variables tienen segundo momento finito, por el Teorema 102 tenemos que

$$\sum_{k=1}^{\lfloor nt \rfloor} \frac{Y_k}{\sqrt{n}} \xrightarrow{\mathscr{D}} \mu t + \sigma B_t, \qquad t \ge 0,$$

para un cierto $\mu \in \mathbb{R}$, que llamaremos *deriva o drift* y un cierto $\sigma > 0$, que llamaremos *volatilidad*. Entonces podemos modelar el proceso de precios por

$$X_t = x_0 e^{\mu t + \sigma B_t}, \qquad \forall t \ge 0,$$

donde $(B_t, t \ge 0)$ es un Movimiento Browniano estándar. Si α es mi factor de descuento, podemos encontrar el valor de μ^* que hace que el precio descontado sea una martingala, ya que si tomamos $(\mathscr{F}_t, t \ge 0)$ del MB [17], tenemos que

$$\begin{aligned}\mathbb{E}\left[e^{-(\alpha-\mu)T}X_T\middle|\mathscr{F}_t\right] &= e^{-(\alpha-\mu)T}x_0\mathbb{E}\left[e^{\sigma(B_T-B_t+B_t)}\middle|\mathscr{F}_t\right] = e^{-(\alpha-\mu)T}x_0e^{\sigma B_t}\mathbb{E}\left[e^{\sigma(B_T-B_t)}\right] \\ &= e^{-(\alpha-\mu)T}x_0e^{\sigma B_t}e^{\frac{\sigma^2}{2}(T-t)} = e^{-(\alpha-\mu)t}x_0X_te^{-(\alpha-\mu-\frac{\sigma^2}{2})(T-t)},\end{aligned}$$

por lo que el precio descontado solo será martingala si se satisface que $\mu^* = \alpha - \frac{\sigma^2}{2}$. En ese caso, podemos calcular explícitamente el precio de una opción call europea., simplemente calculando $v_0 =$

Teorema 112. *El precio de una opción call con ganancia final $(X_T - K)^+$ está dado por*

$$v_0 = \mathbb{E}\left[e^{-\alpha t}V_T\right] = x_0\Phi(d_1) - e^{\alpha T}K\Phi(d_2),$$

donde $\Phi(\cdot)$ es la FDA de una normal estándar, $d_1 := \frac{\ln\left(\frac{x_0}{K}\right)+(\alpha+\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$ y $d_2 := d_1 - \sigma\sqrt{T}$.

[17] es fácil ver que será la misma que la del proceso de precios por ser una función continua

Dem. Notemos que la esperanza que queremos calcular, usando la definición del proceso de precios y la linealidad de la esperanza, corresponde a

$$\mathbb{E}\left[e^{-\alpha t}(X_T-K)\mathbb{1}_{X_T\geq K}\right]=e^{-\alpha t}x_0\mathbb{E}\left[e^{\mu^* t+\sigma B_t}\mathbb{1}_{\{\mu^* t+\sigma B_t\geq \ln(K/x_0)\}}\right]-e^{-\alpha t}K\mathbb{E}\left[\mathbb{1}_{\{\mu^* t+\sigma B_t\geq \ln(K/x_0)\}}\right]. \tag{9}$$

Usando las propiedades de la distribución normal, tenemos que la probabilidad en el segundo término de la derecha corresponde a

$$\begin{aligned}\mathbb{P}\left[\sigma B_T\geq \ln(K/x_0)-\mu^* t\right]&=\mathbb{P}\left[\frac{B_T}{\sqrt{T}}\geq\frac{\ln(K/x_0)-\mu^* T}{\sigma\sqrt{T}}\right]=\mathbb{P}\left[-\frac{B_T}{\sqrt{T}}\leq-\frac{\ln(K/x_0)-\mu^* T}{\sigma\sqrt{T}}\right]\\&=\Phi\left(\frac{\ln(x_0/K)+\mu^* T}{\sigma\sqrt{T}}\right)\end{aligned} \tag{10}$$

Por otro lado, para calcular la probabilidad en el primer término de la derecha, notemos que $\mu^* T+\sigma B_T\sim\mathcal{N}(\mu^* T,\sigma^2 T)$, y por lo tanto

$$\begin{aligned}\mathbb{E}\left[e^{\mu^* T+\sigma B_T}\mathbb{1}_{\{\mu^* T+\sigma B_T\geq\ln(K/x_0)\}}\right]&=\int_{\ln(K/x_0)}^{\infty}e^y\frac{1}{\sqrt{2\pi\sigma^2 T}}e^{-\frac{(y-\mu^* T)^2}{2\sigma^2 T}}\,dy\\&=\int_{\ln(K/x_0)}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2 T}}e^{-\frac{y^2-2y\mu^* T+(\mu^* T)^2-2\sigma^2 Ty+T^2(\mu^*+\sigma^2)^2-T^2(\mu^*+\sigma^2)^2}{2\sigma^2 T}}\,dy\\&=\int_{\ln(K/x_0)}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2 T}}e^{-\frac{(y-(\sigma^2+\mu^*)T)^2}{2\sigma^2 T}}e^{\left(\mu^*+\frac{\sigma^2}{2}\right)T}\,dy\\&=e^{\left(\mu^*+\frac{\sigma^2}{2}\right)T}\Phi\left(\frac{\ln(x_0/K)+(\sigma^2+\mu^*)T}{\sigma\sqrt{T}}\right),\end{aligned} \tag{11}$$

donde usamos que si $N\sim\mathcal{N}((\sigma^2+\mu^*)T,\sigma^2 T)$ entonces

$$\mathbb{P}\left[N\geq\ln(K/x_0)\right]=\mathbb{P}\left[\frac{N-(\sigma^2+\mu^*)T}{\sigma\sqrt{T}}\geq\frac{\ln(K/x_0)-(\sigma^2+\mu^*)T}{\sigma\sqrt{T}}\right]=\Phi\left(\frac{\ln(x_0/K)+(\sigma^2+\mu^*)T}{\sigma\sqrt{T}}\right)$$

Si reemplazamos (10) y (11) en (9), tenemos que

$$\mathbb{E}\left[e^{-\alpha t}(X_T-K)\mathbb{1}_{X_T\geq K}\right]=e^{\left(\mu^*+\frac{\sigma^2}{2}-\alpha\right)T}x_0\Phi\left(\frac{\ln(x_0/K)+(\sigma^2+\mu^*)T}{\sigma\sqrt{T}}\right)-e^{-\alpha T}K\Phi\left(\frac{\ln(x_0/K)+\mu^* T}{\sigma\sqrt{T}}\right),$$

de donde deducimos lo deseado notando que $\mu^*=\alpha-\frac{\sigma^2}{2}$ y por lo tanto $\sigma^2+\mu^*=\alpha+\frac{\sigma^2}{2}$. □

Referencias

[1] Caballero, M. E., Rivero, V. M., Uribe, G. & Velarde, C. *Cadenas de Markov. Un enfoque elemental*, Número 29 en Textos Nivel Medio, 2004.

[2] Durret, R. *Essentials of Stochastic Processes*, New York: Springer-Verlag, 2012.

[3] Grimmett, G. R. & Stirzaker, D. *Probability and Random Processes*, Oxford University Press 3ed, 2001.

[4] Hoel, P. G.; Port, S. C. & Stone, C. J. *Introduction to stochastic processes*, Boston: Houghton Mifflin, 1972.

[5] Karlin, S. & Taylor, H. M. *An Introduction to Stochastic Modeling*, Academic Press 3Ed., 1998.

[6] Häggström, O. *Finite Markov chains and algorithmic applications*, Cambridge University Press, 2002.

[7] Rincón, L. *Introducción a los procesos estocásticos*, Las Prensas de Ciencias, UNAM, 2012.

[8] Ross, S. M. *Introduction to probability models*, Academic Press 7ed, 2007.