
PROCESOS DE MARKOV*

Esta versión: 7 de mayo de 2020

Daniel Remenik
Depto. de Ingeniería Matemática
Centro de Modelamiento Matemático
Universidad de Chile

*Esta es una versión en progreso de los apuntes del curso MA4401, Otoño 2020.

1. Cadenas de Markov a tiempo discreto

1.1. Definición, construcción y propiedades básicas

1.1.1. Breve recuerdo de probabilidades

Un *espacio de probabilidad* es un espacio de medida $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $\mathbb{P}(\Omega) = 1$. \mathbb{P} es la *medida de probabilidad*. Los $A \in \mathcal{F}$ son los *eventos*.

Espacio de Probabilidad
Medida de Probabilidad
Variable Aleatoria

Dado un espacio medible (S, \mathcal{S}) , una *variable aleatoria* X a valores en S es una función $X: \Omega \rightarrow S$ que es \mathcal{F} - \mathcal{S} medible. Usualmente nos olvidamos del espacio de probabilidad en la notación. Por ejemplo, cuando escribimos $\mathbb{P}(X \in A)$ a lo que nos referimos es a la medida de la preimagen de A por X , es decir

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega: X(\omega) \in A\}).$$

Rara vez incluimos los ω 's en la notación, pero es importante mantener la definición formal en mente.

Dos eventos A y B son *independientes* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. La *probabilidad condicional* de A dado B es

Independencia
Prob. Condicional

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

(asumiendo $\mathbb{P}(B) > 0$; a menudo omitiremos especificar esto cuando hablamos de probabilidades condicionales).

Notar que la aplicación $\mathbb{P}(\cdot|B): \mathcal{F} \rightarrow [0, 1]$ define una medida de probabilidad sobre $(\Omega, \mathcal{F}, \mathbb{P})$ soportada en B (o alternativamente una medida de probabilidad sobre $(B, \mathcal{F}, \mathbb{P})$). De esta forma, por ejemplo, podemos decir que dados $A, B, C \in \mathcal{F}$, A y B son *independientes condicionalmente a C* si $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$. Ejercicio: Asegúrese de que comprende el contenido de la última frase.

Med. de Prob. sobre un evento B

~~~~~ Repase la noción de independencia para variables aleatorias y familias de variables aleatorias.

~~~~~ Repase la definición y propiedades de esperanza condicional, dadas en el Apéndice A.

1.1.2. Procesos estocásticos y cadenas de Markov

Dados $(\Omega, \mathcal{F}, \mathbb{P})$ y (S, \mathcal{S}) como antes, decimos que una sucesión $(X_n)_{n \geq 0}$ de variables aleatorias es un *proceso estocástico*. Alternativamente uno puede pensar en el proceso estocástico $(X_n)_{n \geq 0}$ como una variable aleatoria a valores en el espacio producto $S^{\mathbb{N}}$, es decir $X: \Omega \rightarrow S^{\mathbb{N}}$ que es \mathcal{F} - $\mathcal{S}^{\mathbb{N}}$ medible, con $\mathcal{S}^{\mathbb{N}}$ dotado de la σ -álgebra producto.

Proceso Estocástico

(Más adelante consideraremos procesos estocásticos indexados por \mathbb{R}^+ en vez de \mathbb{N}).

El propósito del curso es estudiar un tipo especial pero muy importante de procesos estocásticos: los procesos de Markov.

En todo lo que sigue I (o a veces E) denotará un conjunto numerable, el “espacio de estados” donde nuestros procesos estocásticos tomarán valores. Sobre I consideramos la σ -álgebra de las partes $\mathcal{P}(I)$. Los elementos de I los denotaremos típicamente por i, j, k o x, y, z . Detrás de todo lo que haremos habrá un espacio de probabilidad fijo $(\Omega, \mathcal{F}, \mathbb{P})$, que solo invocaremos cuando sea necesario.

Espacio de Estados

Definición 1.1.1. (Cadena de Markov) Un proceso estocástico $X = (X_n)_{n \geq 0}$ a valores en I es un *proceso de Markov a tiempo discreto* o una *cadena de Markov* si satisface la *propiedad de Markov*:

Cadena de Markov

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

[1cm] para todo $i_0, \dots, i_{n+1} \in I$ (que satisfaga $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$).

Propiedad de Markov

Observación 1.1.2. La propiedad de Markov es equivalente a decir que X_{n+1} y (X_0, \dots, X_{n-1}) son independientes condicionales a X_n , es decir, “condicional al presente, el futuro del proceso es independiente del pasado”. Ejercicio: demuestre esto.

Interpretación de la CM

Proposición 1.1.3. Si X es una cadena de Markov entonces para todo $0 \leq t_0 \leq \dots \leq t_{n+1}$ en \mathbb{N} y todo $i_0, \dots, i_{n+1} \in I$,

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \dots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n).$$

[1cm]

CM a distintos tiempos

Demostración. Ejercicio. □

Ejemplo 1.1.4. (Marcha aleatoria simple en \mathbb{Z}) Una partícula o individuo camina por \mathbb{Z} de la siguiente forma. A tiempo 0 comienza en $X_0 = x \in \mathbb{Z}$. Luego en cada instante de tiempo decide dar un paso a la derecha con probabilidad $p \in [0, 1]$ y un paso a la izquierda con probabilidad $1 - p$, todo de manera independiente, es decir, para cada $n \geq 1$ tenemos

Marcha Aleatoria Simple en \mathbb{Z}

$$X_n = \begin{cases} X_{n-1} + 1 & \text{con probabilidad } p, \\ X_{n-1} - 1 & \text{con probabilidad } 1 - p. \end{cases}$$

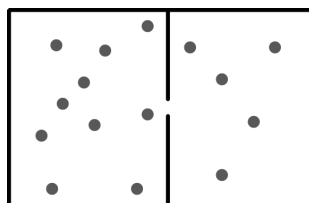
Entonces $(X_n)_{n \geq 0}$ es una cadena de Markov (ejercicio: chequee esto).

Por ahora esta descripción es informal, luego veremos como se puede construir esta cadena de manera rigurosa.

Ejemplo 1.1.5. (Ruina del jugador) Consideramos ahora un juego de apuestas, donde en cada turno el jugador gana 1 con probabilidad p y pierde 1 con probabilidad $1 - p$. El jugador para de jugar si se queda sin dinero. Esto es muy similar a la cadena anterior, excepto que ahora el espacio de estados es $\mathbb{Z}_{\geq 0}$ y tenemos

$$X_n = \begin{cases} X_{n-1} + 1 & \text{con probabilidad } p \text{ si } n > 0, \\ X_{n-1} - 1 & \text{con probabilidad } 1 - p \text{ si } n > 0, \\ 0 & \text{con probabilidad } 1 \text{ si } n = 0. \end{cases}$$

Ejemplo 1.1.6. (Cadena de Ehrenfest) En este ejemplo nos imaginamos una caja rectangular dividida en dos cámaras que están conectadas por un pequeño orificio, donde r moléculas de aire se mueven libremente al azar. P. Ehrenfest propuso a comienzos del



sigo XX el siguiente modelo para estudiar esta situación. Tenemos r bolas distribuidas en

2 dos urnas, una con k bolas (la cámara izquierda de la caja) y la otra con las restantes $r - k$ (la cámara derecha). Luego en cada instante de tiempo tomamos una de las r bolas elegida uniformemente al azar y la cambiamos de urna. La cadena de Markov detrás de este ejemplo es la siguiente: $I = \{0, \dots, r\}$ y, para $n \geq 1$, tenemos

$$X_n = \begin{cases} X_{n-1} + 1 & \text{con probabilidad } \frac{r - X_{n-1}}{r}, \\ X_{n-1} - 1 & \text{con probabilidad } \frac{X_{n-1}}{r}. \end{cases}$$

Definición 1.1.7. (Cadena de Markov homogénea, matriz de transición) Sea X una cadena de Markov. Decimos que X es *homogénea* si

$$\mathbb{P}(X_{m+1} = j | X_m = i) = \mathbb{P}(X_{n+1} = j | X_n = i) \quad \forall m, n \geq 0, i, j \in I.$$

En este caso a la matriz $P = (p_{ij})_{i,j \in I}$ dada por $p_{ij} = \mathbb{P}(X_1 = j | X_0 = i)$ le llamamos *matriz de transición* de la cadena de Markov X ; p_{ij} es la *probabilidad de transición de i a j* .

Ejemplo 1.1.8. La marcha aleatoria simple del Ejem. 1.1.4 es una cadena de Markov homogénea, y su matriz de transición P está dada por

$$p_{ij} = \begin{cases} p & \text{si } j = i + 1, \\ 1 - p & \text{si } j = i - 1, \\ 0 & \text{cualquier otro } j. \end{cases}$$

Ejercicio 1.1.9. Escriba la matriz de transición asociada a la cadena de Ehrenfest introducida en el Ejemplo 1.1.6.

Notar que P es una matriz indexada por un conjunto numerable, pero esto no nos traerá mayores dificultades (lo discutiremos en más detalle más adelante). P no puede ser cualquier matriz, si no que claramente debe satisfacer algunas propiedades especiales, precisamente:

Definición 1.1.10. (Matriz estocástica) Una matriz $P = (p_{ij})_{i,j \in I}$ se dice *matriz estocástica* si:

- $p_{ij} \geq 0$ para todo $i, j \in I$.
- $\sum_{j \in I} p_{ij} = 1$ para todo $i \in I$.

En otras palabras, P es una matriz estocástica si para cada $i \in I$, $(p_{ij})_{j \in I}$ es un vector de probabilidad.

Proposición 1.1.11. Si P es la matriz de transición de una cadena de Markov, entonces P es matriz estocástica.

Demostración. Ejercicio (fácil). □

Durante el curso estaremos interesados casi exclusivamente en el caso de cadenas de Markov homogéneas. Por lo tanto, en todo lo que sigue trabajaremos bajo la siguiente convención:

Salvo en los casos puntuales en que se indique lo contrario, todas las cadenas de Markov que aparezcan en estas notas serán consideradas implícitamente como cadenas homogéneas.

Ejercicio 1.1.12. Sea X una cadena de Markov no homogénea a valores en I . Encuentre una cadena de Markov \tilde{X} a valores en $I \times \mathbb{N}$ que es homogénea y que permite recuperar la distribución de la cadena original X .

CM Homogénea

Matriz de Transición
Prob. de Transición

Matriz Estocástica

Matr. de Trans. de CM
es Matr. Est.

Todas las CM son
homogéneas
(convención)

1.1.3. Caracterización y construcción

La matriz de transición gobierna la evolución de nuestra cadena de Markov X . Para determinar la distribución de la cadena hace falta un ingrediente más:

Definición 1.1.13. (Distribución inicial) Al vector de probabilidad μ sobre I definido por $\mu_i = \mathbb{P}(X_0 = i)$ le llamamos la *distribución inicial* de la cadena de Markov X .

Distribución Inicial

Usaremos indistintamente la notación μ para un vector de probabilidad como en la definición y para una medida de probabilidad sobre I , que en cualquier caso son la misma cosa.

Es relativamente fácil ver que toda cadena de Markov, y en particular sus distribuciones finito-dimensionales, queda determinada por su matriz de transición y su distribución inicial:

Proposición 1.1.14. X es una cadena de Markov con matriz de transición P y distribución inicial μ si y solo si

Caracterización de una Cadena de Markov

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$

Demostración. Para la primera implicancia escribimos

$$\begin{aligned} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &\quad \cdot \mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) \mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= p_{i_{n-1} i_n} \mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_0 = i_0), \end{aligned}$$

donde en la segunda igualdad usamos la propiedad de Markov, y luego repetimos inductivamente hasta

$$\begin{aligned} &= p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_1 i_0} \mathbb{P}(X_0 = i_0) \\ &= p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_1 i_0} \mu_{i_0}, \end{aligned}$$

que es lo que buscábamos.

Para la converso, probemos primero que X es cadena de Markov, es decir, que satisface la propiedad de Markov. Usando la hipótesis tenemos

$$\begin{aligned} \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) &= \frac{\mathbb{P}(X_{n+1} = i_{n+1}, \dots, X_0 = i_0)}{\mathbb{P}(X_n = i_n, \dots, X_0 = i_0)} \\ &= \frac{\mu_{i_0} p_{i_0 i_1} \cdots p_{i_n i_{n+1}}}{\mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}} = p_{i_n i_{n+1}}. \end{aligned}$$

Por otro lado tenemos

$$\begin{aligned} \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) &= \frac{\mathbb{P}(X_{n+1} = i_{n+1}, X_n = i_n)}{\mathbb{P}(X_n = i_n)} \\ &= \frac{1}{\mathbb{P}(X_n = i_n)} \sum_{i_0, \dots, i_{n-1} \in I} \mathbb{P}(X_{n+1} = i_{n+1}, \dots, X_0 = i_0) \\ &= \frac{1}{\mathbb{P}(X_n = i_n)} \sum_{i_0, \dots, i_{n-1} \in I} \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} p_{i_n i_{n+1}} \\ &= \frac{1}{\mathbb{P}(X_n = i_n)} \sum_{i_0, \dots, i_{n-1} \in I} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) p_{i_n i_{n+1}}, \end{aligned}$$

donde usamos probabilidades totales y luego la hipótesis nuevamente. Dejando el factor $p_{i_n i_{n+1}}$ afuera, la última suma vale $\mathbb{P}(X_n = i_n)$, y entonces deducimos que $\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}$. Juntando con lo anterior, esto nos da la propiedad de Markov, y además nos dice que P es la matriz de transición de la cadena. Falta derivar la condición inicial, pero esto es directo de la hipótesis, tomando $n = 1$ y sumando sobre $i_1 \in I$ (ejercicio: hágalo). \square

Es natural preguntarse ahora por la converso al resultado anterior: dados un vector de probabilidad μ y una matriz estocástica P , existe una cadena de Markov con condición inicial μ y matriz de transición P ?

Teorema 1.1.15. (Existencia de una cadena de Markov con μ y P dados) Sea μ una medida de probabilidad sobre I y P una matriz estocástica indexada por I . Entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y una sucesión $X = (X_n)_{n \geq 0}$ de variables aleatorias en este espacio tal que X es una cadena de Markov con distribución inicial μ y matriz de transición P .

Exist. de una CM con μ y P dados

La construcción de una cadena de Markov como en el teorema no es única. En la demostración usaremos la que se conoce como la *construcción canónica*, después veremos otra que también es importante.

~~~~~ Repase el Teorema de Consistencia de Kolmogorov (ver, por ej., [SM18]).

*Demostración.* Partimos eligiendo  $\Omega$  como  $I^{\mathbb{N}}$  y  $\mathcal{F} = \mathcal{P}(I)^{\mathbb{N}}$  (la  $\sigma$ -álgebra producto);  $(\Omega, \mathcal{F})$  elegidos así es el “espacio medible canónico” para la construcción de una cadena de Markov con  $\mu$  y  $P$  dados. Para definir  $\mathbb{P}$  especificamos su valor sobre cilindros mediante

$$\mathbb{P}(\{\omega \in I^{\mathbb{N}} : \omega_0 = i_0, \dots, \omega_n = i_n\}) = \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \quad (1.1)$$

Usando el Teorema de Consistencia de Kolmogorov se tiene que existe una única medida de probabilidad  $\mathbb{P}$  sobre todo  $I^{\mathbb{N}}$  que satisface (1.1).

**Ejercicio 1.1.16.** Verificar esto.

Ahora definimos la variable aleatoria  $X_n : \Omega \rightarrow I$  mediante

$$X_n(\omega) = \omega_n.$$

Por construcción se tiene entonces que

$$\begin{aligned} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) &= \mathbb{P}(\{\omega \in I^{\mathbb{N}} : X_0(\omega) = i_0, \dots, X_n(\omega) = i_n\}) \\ &= \mathbb{P}(\{\omega \in I^{\mathbb{N}} : \omega_0 = i_0, \dots, \omega_n = i_n\}) = \mu_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \end{aligned}$$

Gracias a la Prop. 1.1.14 deducimos que  $X = (X_n)_{n \geq 0}$  satisface lo requerido.  $\square$

**Notación.** Escribiremos  $X \sim \text{CM}(\mu, P)$  si  $X$  es cadena de Markov con distribución inicial  $\mu$  y matriz de transición  $P$ . Si no necesitamos especificar la distribución inicial, escribiremos  $X \sim \text{CM}(P)$ .

Notación para una dist. de una CM

En la notación anterior no hacemos mención al espacio de probabilidad donde está definida la cadena. En la mayoría de los resultados del curso esto no es relevante; los resultados valdrán para cadenas construidas en cualquier espacio de probabilidad si se asumen las hipótesis necesarias sobre su distribución (i.e. sobre  $\mu$  y  $P$ ). Si uno lo desea, puede considerar entonces que todas nuestras cadenas están construidas usando la construcción canónica, salvo que se indique lo contrario.

Hay una construcción alternativa que es importante y muy útil, pues permite identificar muchos procesos estocásticos como cadenas de Markov (y demostrar que lo son) fácilmente. La damos a continuación.

**Teorema 1.1.17. (Construcción directa de cadenas de Markov)** Sea  $(\Omega, \mathcal{F}, \mathbb{P})$  un espacio de probabilidad en el cual están definidas variables aleatorias independientes  $(U_n)_{n \geq 0}$  con distribución  $\text{Unif}[0, 1]$  y adicionalmente una variable aleatoria  $\xi_0$  a valores  $I$  con distribución  $\mu$ , independiente de todas las  $U_n$ . Sea  $\Phi: [0, 1] \times I \rightarrow I$  medible y definamos  $(X_n)_{n \geq 0}$  mediante

Constr. directa de CM

$$\begin{aligned} X_0 &= \xi_0, \\ X_{n+1} &= \Phi(U_n, X_n), \quad n \geq 0. \end{aligned} \tag{1.2}$$

Entonces  $X \sim \text{CM}(\mu, P)$  con  $P$  dada por  $p_{ij} = \mathbb{P}(\Phi(U_0, i) = j)$ .

*Demostración.* En clases.  $\triangle$

□

**Observación 1.1.18.**

1. Toda cadena de Markov puede construirse de esta manera, es decir, si  $\mu$  y  $P$  están dados entonces podemos darnos un espacio de probabilidad como en el teorema y encontrar un  $\Phi$  que hace lo requerido.

En efecto, supongamos que nos dan  $X \sim \text{CM}(\mu, P)$  y consideremos un espacio de probabilidad con  $\xi_0$  distribuido según  $\mu$  (a valores en  $I$ ) y una secuencia  $(U_n)_{n \geq 0}$  como en el teorema. Debemos construir una función  $\Phi$  que haga lo requerido. Para esto enumeramos  $I$  como  $I = \{i_\ell\}_{\ell \geq 0}$  y definimos, para cada  $k, \ell \in \mathbb{N}$ ,

$$a_\ell^k = \sum_{j=0}^{\ell} p_{i_k i_j} \quad (\text{que es } = \mathbb{P}(X_1 \in \{i_0, \dots, i_\ell\} | X_0 = i_k)),$$

y además  $a_{-1}^k = 0$ . Los intervalos  $\{[a_{\ell-1}^k, a_\ell^k], \ell \geq 0\}$ , que son fijos (y deterministas) en función de  $P$ , particionan  $[0, 1]$ , luego podemos definir  $\Phi(u, i_k) = i_\ell$  donde  $\ell$  es el único índice tal que  $u \in [a_{\ell-1}^k, a_\ell^k]$ . Con esto tenemos

$$\mathbb{P}(\Phi(U_0, i_k) = i_\ell) = \mathbb{P}(U_0 \in [a_{\ell-1}^k, a_\ell^k]) = a_\ell^k - a_{\ell-1}^k = p_{i_k i_\ell},$$

tal como requeríamos.

2. La conclusión del teorema sigue valiendo si cambiamos las  $U_n$  por cualquier otra sucesión de variables aleatorias i.i.d. (pero en ese caso no necesariamente vale el punto 1 anterior).
3. La construcción puede pensarse como una versión para cadenas de Markov del siguiente resultado clásico de probabilidades: si  $F$  es la función de distribución (acumulada) de una variable aleatoria real, y si definimos  $G: [0, 1] \rightarrow \mathbb{R}$  mediante la fórmula  $G(a) = \sup\{x \in \mathbb{R} : F(x) < a\}$ , entonces si  $U \sim \text{Unif}[0, 1]$ , se tiene que la variable aleatoria  $G(U)$  tiene función de distribución  $F$  ( $G$  en este caso puede pensarse como una especie de inversa de la función inyectiva, pero no necesariamente biyectiva,  $F$ ).

En particular, la construcción entrega una manera de extender este último resultado a variables aleatorias a valores en conjuntos numerables.

**Ejercicio 1.1.19.** Demuestre el resultado mencionado en (iii) de la observación anterior y, en particular, que  $G$  está bien definida (y coincide con  $F^{-1}$  si  $F$  es invertible).

Una de las gracias del teorema anterior es que nos permite chequear fácilmente que procesos estocásticos concretos satisfacen la propiedad de Markov.

**Ejemplo 1.1.20.** Volvemos a la marcha aleatoria simple del Ejem. 1.1.4. Aquí podemos tomar  $\xi_0 = x$  y definir

$$\Phi(u, i) = \begin{cases} i + 1 & \text{si } u \leq p, \\ i - 1 & \text{si } u > p. \end{cases}$$

El teorema nos asegura que el proceso estocástico construido de acuerdo a (1.2) es una cadena de Markov con matriz de transición requerida.

#### 1.1.4. Probabilidades de transición

**Notación.** Si  $X \sim \text{CM}(P)$  e  $i \in I$  usaremos denotaremos por  $\mathbb{P}_i$  a la medida de probabilidad correspondiente a la cadena  $X \sim \text{CM}(\delta_i, P)$  con  $\delta_i(j) = \mathbf{1}_{i=j}$  (es decir, a la cadena que parte en  $i$ ). Usaremos  $\mathbb{E}_i$  [2pt] para denotar la esperanza bajo  $\mathbb{P}_i$ .

Notación  $\mathbb{P}_i$

Notación  $\mathbb{E}_i$

Más generalmente, si  $\mu$  es una distribución de probabilidad sobre  $I$ , usamos  $\mathbb{P}_\mu$  (y  $\mathbb{E}_\mu$ ) para denotar la distribución de la cadena  $X \sim \text{CM}(\mu, P)$ .

Notación  $\mathbb{P}_\mu$  y  $\mathbb{E}_\mu$

**Ejercicio 1.1.21.** Convencerse de que  $\mathbb{P}_\mu = \sum_{i \in I} \mu_i \mathbb{P}_i$  y  $\mathbb{E}_\mu = \sum_{i \in I} \mu_i \mathbb{E}_i$ .

Como ya mencionamos,  $P$  es una matriz indexada por el conjunto numerable  $I$ . A menudo es útil interpretar a  $P$  como un operador lineal que actúa sobre funciones  $f: I \rightarrow I$  de acuerdo a

$$Pf(i) = \sum_{j \in I} p_{ij} f(j). \quad (1.3)$$

Interpretación de  $P$  como operador lineal

La suma podría ser o no convergente, pero en muchos casos de interés lo es; por ejemplo si  $f$  es acotada (por Teorema de Convergencia Dominada) o positiva (por Teorema de Convergencia Monótona, aunque en este caso debemos aceptar que la suma valga  $\infty$ ). Cuando  $I$  es finito, lo anterior corresponde exactamente a la acción de  $P$  como matriz indexada por  $I \times I$  al multiplicar a  $f$  como vector indexado por  $I$ , y esta perspectiva podemos extenderla naturalmente al caso en que  $I$  que infinito.

En particular, podemos definir productos de matrices estocásticas mediante la fórmula usual,  $(PQ)_{ij} = \sum_{k \in I} p_{ik} q_{kj}$ , que en este caso siempre es convergente. Naturalmente denotamos también por  $P^0$  a la matriz identidad, es decir  $(P^0)_{ij} = \mathbf{1}_{i=j}$ .

Producto de Matr. Est.

La fórmula (1.3) tiene una interpretación probabilística natural:

Interpr. de la ec. (1.3)

$$Pf(i) = \mathbb{E}_i(f(X_1))$$

De manera similar podemos multiplicar a un vector de probabilidad  $\mu$  sobre  $I$  por la derecha, y obtenemos

$$(\mu P)_j = \sum_{i \in I} \mu_i p_{ij} = \mathbb{P}_\mu(X_1 = j).$$

Ejercicio: Convénzase de las dos fórmulas anteriores.

**Definición 1.1.22. (Probabilidades de transición)** Dada  $X \sim \text{CM}(P)$  y  $n \in \mathbb{N}$ , denotamos por  $P^{(n)} = (p_{ij}^{(n)})_{i,j \in I}$  a la *matriz de transición en  $n$  pasos* dada por

Matriz de Transición en  $n$  pasos

$$p_{ij}^{(n)} = \mathbb{P}_i(X_n = j).$$

**Teorema 1.1.23. (Ecuaciones de Chapman-Kolmogorov)** Si  $X \sim \text{CM}(P)$  entonces para todo  $m, n \geq 0$  se tiene

Ec. de Chapman-Kolmogorov

$$p_{ij}^{(m+n)} = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)}.$$



En particular,

$$P^{(n)} = P^n$$

para todo  $n \geq 0$ .

*Demostración.*

$$\begin{aligned} p_{ij}^{(m+n)} &= \mathbb{P}_i(X_{m+n} = j) = \sum_{k \in I} \mathbb{P}_i(X_{m+n} = j | X_m = k) \mathbb{P}_i(X_m = k) \\ &= \sum_{k \in I} \mathbb{P}(X_{m+n} = j | X_0 = i, X_m = k) \mathbb{P}_i(X_m = k) \\ &= \sum_{k \in I} \mathbb{P}(X_{m+n} = j | X_m = k) \mathbb{P}_i(X_m = k) \\ &= \sum_{k \in I} \mathbb{P}_k(X_n = j) \mathbb{P}_i(X_m = k) = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)}. \end{aligned}$$

La segunda afirmación sale de la primera por inducción, y además claramente implica la primera por definición de potencia de matrices.  $\square$

### 1.1.5. Versión general de la propiedad de Markov

**Definición 1.1.24. (Filtración)** Sea  $(\Omega, \mathcal{F})$  un espacio medible. Una *filtración* en este espacio es una familia  $(\mathcal{F}_n)_{n \geq 0}$  de sub  $\sigma$ -álgebras de  $\mathcal{F}$  tal que  $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$  para cada  $n \geq 0$ .

Filtración

Si  $X$  es un proceso estocástico construido en un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ , la *filtración natural* asociada a  $X$  es aquella dada por

Filtración Natural

$$\mathcal{F}_n = \sigma(\{X_0, \dots, X_n\}).$$

La filtración natural es aquella filtración más pequeña que tal que  $X_0, \dots, X_n$  es  $\mathcal{F}_n$ -medible para cada  $n \geq 0$ , en el sentido de que para cualquier otra filtración  $(\mathcal{G}_n)_{n \geq 0}$  que satisfaga lo mismo necesariamente debe tenerse que  $\mathcal{G}_n \supseteq \mathcal{F}_n$  para cada  $n \geq 0$ .

Filtración natural es la más pequeña

Durante todo el curso,  $(\mathcal{F}_n)_{n \geq 0}$  denotará la filtración natural asociada al proceso estocástico que se esté considerando.

$\mathcal{F}_n$  nos permite representar la información contenida en la evolución del proceso hasta tiempo  $n$ .

$\mathcal{F}_n$  representa la información contenida

**Notación.** Si  $\mathcal{G}$  es sub  $\sigma$ -álgebra de  $\mathcal{F}$  y  $A \in \mathcal{F}$ , escribimos

Notación de  $\mathbb{P}(A|\mathcal{G})$

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(\mathbf{1}_A|\mathcal{G}).$$

En el siguiente resultado reinterpretamos la definición y caracterización de cadenas de Markov en el lenguaje de esperanzas condicionales:

**Proposición 1.1.25.** Sea  $X$  un proceso estocástico (y  $(\mathcal{F}_n)_{n \geq 0}$  su filtración natural asociada). Son equivalentes:

Caract. de CM usando filtración natural

- (i)  $X$  es cadena de Markov (homogénea).
- (ii)  $\mathbb{P}(X_{n+1} = j | \mathcal{F}_n) = \sum_{i \in I} \mathbb{P}(X_{n+1} = j | X_n = i) \mathbf{1}_{X_n = i}$ .
- (iii)  $\mathbb{P}(X_{n+1} = j | \mathcal{F}_n)$  es  $\sigma(\{X_n\})$ -medible.

En particular, si  $X \sim \text{CM}(P)$  se tiene que

$$\mathbb{P}(X_{n+1} = j | \mathcal{F}_n)(\omega) = p_{X_n(\omega), j}$$

(donde para mayor claridad hemos incluido los  $\omega$ 's, aunque casi nunca los escribiremos).

**Observación 1.1.26.** A veces abusaremos notación y escribiremos la última igualdad de la siguiente forma:

$$\mathbb{P}(X_{n+1} = j | \mathcal{F}_n)(\omega) = \mathbb{P}_{X_n(\omega)}(X_1 = j).$$

En estos casos, el lado derecho debe interpretarse como  $\mathbb{P}_{X_n(\omega)}(\tilde{X}_1 = j)$  donde  $\tilde{X}$  es una copia independiente de  $X$ .

*Demostración.* Tenemos que  $\mathcal{F}_n = \sigma(\{X_0, \dots, X_n\}) = \sigma(\{\bigcap_{j=0}^n X_j^{-1}(\{i_j\}), i_0, \dots, i_n \in I\})$ . La lista de conjuntos que aparece generando la  $\sigma$ -álgebra es una partición de  $\Omega$  y luego, gracias a la Prop. A.0.7,

$$\begin{aligned} \mathbb{P}(X_{n+1} = j | \mathcal{F}_n) &= \sum_{i_0, \dots, i_n \in I} \frac{\mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = j)}{\mathbb{P}(X_0 = i_0, \dots, X_n = i_n)} \mathbf{1}_{X_0=i_0, \dots, X_n=i_n} \\ &= \sum_{i_0, \dots, i_n \in I} \mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) \mathbf{1}_{X_0=i_0, \dots, X_n=i_n}. \end{aligned}$$

El lado derecho es igual a  $\sum_{i_n \in I} \mathbb{P}(X_{n+1} = j | X_n = i_n) \mathbf{1}_{X_n=i_n}$  si y sólo si  $\mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = j | X_n = i_n)$  (ejercicio: demuestre esta equivalencia), que es justamente la propiedad de Markov. Esto nos da la equivalencia de (i) y (ii).

Que (ii) implica (iii) es directo. Para la converso usamos la hipótesis de medibilidad para escribir

$$\begin{aligned} \mathbb{P}(X_{n+1} = j | \mathcal{F}_n) &= \mathbb{E}(\mathbb{P}(X_{n+1} = j | \mathcal{F}_n) | \sigma(\{X_n\})) = \mathbb{E}(\mathbb{E}(\mathbf{1}_{X_{n+1}=j} | \mathcal{F}_n) | \sigma(\{X_n\})) \\ &= \mathbb{E}(\mathbf{1}_{X_{n+1}=j} | \sigma(\{X_n\})). \end{aligned}$$

Desarrollando la última esperanza condicional como en la primera equivalencia deducimos que (ii) se tiene.  $\square$

La siguiente es una versión general de la propiedad de Markov.

**Teorema 1.1.27.** Sea  $X$  cadena de Markov (homogénea) y sea  $f: I^{\mathbb{N}} \rightarrow \mathbb{R}$  medible y acotada. Entonces

$$\mathbb{E}(f(X_{n+1}, X_{n+2}, \dots) | \mathcal{F}_n) = \mathbb{E}_{X_n}(f(X_1, X_2, \dots)).$$

En palabras simples, la propiedad de Markov dice que:

*Si reiniciamos la cadena a tiempo  $n$  y observamos su trayectoria a partir de ese tiempo, lo que vemos es una copia de la cadena original, partiendo desde el estado  $X_n$ , pero independiente de los estados  $X_0, \dots, X_{n-1}$ .*

**Observación 1.1.28.**

- (i) Lo anterior es una equivalencia (ejercicio: demuéstrela).
- (ii) Acá hemos abusado notación nuevamente, ver la Obs. 1.1.26, el significado de lo que escribimos es análogo.
- (iii) Se tiene también que  $\mathbb{E}(f(X_n, X_{n+1}, \dots) | \mathcal{F}_n) = \mathbb{E}_{X_n}(f(X_0, X_1, \dots))$ , la demostración es la misma (o puede obtenerse como corolario).

Abuso de notación de la igualdad anterior

Vers. general de la propiedad de Markov

*Demostración.* Partiremos probando primero por inducción que la propiedad se tiene para el conjunto  $\mathcal{R}$  de funciones de la forma  $f(i_1, i_2, \dots) = f_1(i_1)f_2(i_2) \cdots f_m(i_m)$ , donde  $m \geq 1$  y  $f_1, \dots, f_m: I \rightarrow \mathbb{R}$  son medibles y acotadas.

**Ejercicio 1.1.29.** Use la propiedad de Markov para probar el caso base,  $m = 1$ .

Ahora hacemos el paso inductivo, considerando el caso  $m + 1$ . Tenemos

$$\begin{aligned} \mathbb{E}(f(X_{n+1}, X_{n+2}, \dots) | \mathcal{F}_n) &= \mathbb{E}(f_1(X_{n+1})f_2(X_{n+2}) \cdots f_{m+1}(X_{n+m+1}) | \mathcal{F}_n) \\ &= \mathbb{E}(\mathbb{E}(f_1(X_{n+1})f_2(X_{n+2}) \cdots f_{m+1}(X_{n+m+1}) | \mathcal{F}_{n+m}) | \mathcal{F}_n) \\ &= \mathbb{E}(f_1(X_{n+1})f_2(X_{n+2}) \cdots f_m(X_{n+m}) \mathbb{E}(f_{m+1}(X_{m+1}) | \mathcal{F}_{n+m}) | \mathcal{F}_n) \\ &= \mathbb{E}(f_1(X_{n+1})f_2(X_{n+2}) \cdots f_m(X_{n+m}) \mathbb{E}_{X_{n+m}}(f_{m+1}(X_1)) | \mathcal{F}_n), \end{aligned}$$

donde usamos el caso base en la última igualdad. Notemos que el último factor dentro de la esperanza depende ahora solo de  $X_{n+m}$ , al igual que el que lo precede. Entonces si definimos  $\tilde{f}_m(j) = f_m(j)\mathbb{E}_j(f_1(X_1))$  y usamos la hipótesis inductiva, lo anterior es igual a

$$\begin{aligned} \mathbb{E}(f_1(X_{n+1})f_2(X_{n+2}) \cdots \tilde{f}_m(X_{n+m}) | \mathcal{F}_n) &= \mathbb{E}_{X_n}(f_1(X_1)f_2(X_2) \cdots \tilde{f}_m(X_m)) \\ &= \mathbb{E}_{X_n}(f_1(X_1)f_2(X_2) \cdots f_m(X_m)\mathbb{E}_{X_m}(f_{m+1}(X_1))) \\ &= \mathbb{E}_{X_n}(f_1(X_1)f_2(X_2) \cdots f_m(X_m)\mathbb{E}(f_{m+1}(X_{m+1}) | \mathcal{F}_{n+m})) \\ &= \mathbb{E}_{X_n}(\mathbb{E}(f_1(X_1)f_2(X_2) \cdots f_m(X_m)f_{m+1}(X_{m+1}) | \mathcal{F}_{n+m})) \\ &= \mathbb{E}_{X_n}(f_1(X_1)f_2(X_2) \cdots f_{m+1}(X_{m+1})), \end{aligned}$$

que es lo que queríamos.

Ahora concluimos. Sea  $\mathcal{A}$  el conjunto de funciones  $f: I^{\mathbb{N}} \rightarrow \mathbb{R}$  medibles y acotadas que satisfacen la igualdad que queremos probar.  $\mathcal{A}$  claramente es un espacio vectorial, contiene a las constantes, y por Teorema de la Clase Monótona, si  $(f_n)_{n \geq 0} \subset \mathcal{A}$  con  $f_n \geq 0$  satisface  $f_n \nearrow f$  entonces  $f \in \mathcal{A}$ . Arriba probamos que  $\mathcal{R} \subseteq \mathcal{A}$ , y es evidente que  $\mathcal{R}$  es cerrado bajo producto, y luego por la forma funcional del Teorema de la Clase Monótona,  $\sigma(\mathcal{R}) \subseteq \mathcal{A}$ . Pero  $\sigma(\mathcal{R})$  es claramente todo el espacio de funciones medibles y acotadas de  $I^{\mathbb{N}}$  en  $\mathbb{R}$ , lo que termina la demostración.  $\square$

## 1.2. Propiedad de Markov fuerte

Nuestro objetivo ahora es extender la propiedad de Markov a tiempos aleatorios. Lo primero es definir y estudiar una clase adecuada de tiempos aleatorios. intuitivamente, si queremos que la propiedad de Markov siga valiendo, lo que necesitamos es que estos tiempos solo dependan de la trayectoria del proceso hasta el presente, pero no de su futuro. A continuación veremos la manera de hacer esto preciso.

### 1.2.1. Tiempos de parada

**Definición 1.2.1. (Tiempos de parada)** Sea  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n \geq 0})$  un espacio de probabilidad filtrado, es decir, un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  con una filtración  $(\mathcal{F}_n)_{n \geq 0}$  definida sobre él. Decimos que una variable aleatoria  $\tau: \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  es un tiempo de parada si

$$\{\tau \leq n\} \in \mathcal{F}_n \quad \forall n \geq 0.$$

Abreviaremos tiempo de parada como t.d.p.

Tiempos de Parada

**Ejemplo 1.2.2.** Si  $X$  es una marcha aleatoria simple y  $a \in \mathbb{Z}$ ,  $\tau_a = \inf\{m \geq 0: X_m \geq a\}$  es un t.d.p.

**Ejercicio 1.2.3.** Demuestre que la propiedad que define t.d.p. es equivalente a pedir que  $\{\tau = n\} \in \mathcal{F}_n$  para todo  $n \geq 0$ .

**Proposición 1.2.4.**

Prop. de tiempos de parada

- (1) Si  $\tau_1, \tau_2$  son t.d.p. entonces  $\tau_1 \wedge \tau_2 := \min\{\tau_1, \tau_2\}$  y  $\tau_1 \vee \tau_2 := \max\{\tau_1, \tau_2\}$  son t.d.p.
- (2) Si  $(\tau_n)_{n \geq 0}$  son t.d.p. entonces  $\liminf_n \tau_n$ ,  $\limsup_n \tau_n$ ,  $\inf_n \tau_n$  y  $\sup_n \tau_n$  son t.d.p.

*Demostración.* Ejercicio. □

Dada una filtración  $(\mathcal{F}_n)_{n \geq 0}$ , escribiremos

Notación de  $\mathcal{F}_\infty$

$$\mathcal{F}_\infty = \lim_{n \rightarrow \infty} \mathcal{F}_n = \bigcup_{n \geq 0} \mathcal{F}_n.$$

En el caso de la filtración natural asociada a un proceso estocástico, esto corresponde intuitivamente a la información contenida en la trayectoria completa del proceso.

Intuición sobre  $\mathcal{F}_\infty$

**Definición 1.2.5.** Dado un espacio de probabilidad filtrado y un t.d.p.  $\tau$ , definimos

Def. de  $\mathcal{F}_\tau$

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty: A \cap \{\tau \leq n\} \in \mathcal{F}_n \forall n \geq 0\}.$$

**Ejemplo 1.2.6.** Probar que si  $\tau = n$ , entonces con la definición anterior se cumple  $\mathcal{F}_\tau = \mathcal{F}_n$ .

**Proposición 1.2.7. (Propiedades de  $\mathcal{F}_\tau$ )** Sean  $\tau, \tau_1, \tau_2$  t.d.p.

Propiedades de  $\mathcal{F}_\tau$

- (1)  $\mathcal{F}_\tau$  es  $\sigma$ -álgebra.
- (2)  $\tau$  es  $\mathcal{F}_\tau$ -medible.
- (3) Si  $\tau_1(\omega) \leq \tau_2(\omega)$  para cada  $\omega \in \Omega$  entonces  $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$ .
- (4)  $\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty: A \cap \{\tau = n\} \in \mathcal{F}_n \forall n \geq 0\}$ .
- (5)  $A \in \mathcal{F}_\tau$  si y sólo si existe  $(A_n)_{n \in \mathbb{N} \cup \{\infty\}}$  con  $A_n \in \mathcal{F}_n$  tal que  $A = \bigcup_{n \in \mathbb{N} \cup \{\infty\}} A_n \cap \{\tau = n\}$ .
- (6)  $\{\tau_1 \leq \tau_2\}, \{\tau_1 < \tau_2\} \in \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$ .

*Demostración.* Consideremos (1). Claramente  $\emptyset, \Omega \in \mathcal{F}_\tau$ . Ahora supongamos que  $A \in \mathcal{F}_\tau$ . Como  $A \cap \{\tau \leq n\} \in \mathcal{F}_n$  para cada  $n$ , tomando complemento tenemos  $A^c \cup \{\tau > n\} \in \mathcal{F}_n$  para cada  $n$ . Pero además  $\{\tau \leq n\} \in \mathcal{F}_n$ , y entonces  $\{\tau \leq n\} \cap (A^c \cup \{\tau > n\}) = A^c \cap \{\tau \leq n\} \in \mathcal{F}_n$ . Luego  $A^c \in \mathcal{F}_\tau$ . Ejercicio: verificar la propiedad de uniones numerables para terminar la demostración de (1). Ejercicio: Demostrar (2), (3) y (4).

Para demostrar (5) tomamos  $A \in \mathcal{F}_\tau$  y escribimos

$$A = \bigcup_{n \in \mathbb{N} \cup \{\infty\}} A \cap \{\tau = n\} = \bigcup_{n \in \mathbb{N} \cup \{\infty\}} A_n \cap \{\tau = n\}$$

con  $A_n = A \cap \{\tau = n\}$ . La otra implicancia es aún más directa. □

Terminamos esta parte con un resultado que será útil a continuación, y que es muy similar a la propiedad (5) anterior y se demuestra de manera análoga (ejercicio: hágalo).

**Lema 1.2.8.**  $Y: \Omega \rightarrow \mathbb{R}$  es  $\mathcal{F}_\tau$ -medible si y sólo si existe una sucesión  $(Y_n)_{n \in \mathbb{N} \cup \{\infty\}}$  de variables aleatorias, con  $Y_n$   $\mathcal{F}_n$ -medible, tal que  $Y = \sum_{n \in \mathbb{N} \cup \{\infty\}} Y_n \mathbf{1}_{\tau=n}$ .

Caract. de que una VA sea  $\mathcal{F}_\tau$ -med.

### 1.2.2. Propiedad de Markov fuerte

Con las definiciones y resultados anteriores ya estamos en condiciones de establecer la versión esencialmente más fuerte de la propiedad de Markov disponible para cadenas de Markov.

**Teorema 1.2.9. (Propiedad de Markov fuerte)** *Sea  $X$  cadena de Markov (homogénea) y sea  $f: I^{\mathbb{N}} \rightarrow \mathbb{R}$  medible y acotada. Sea  $\tau$  un tiempo de parada. Entonces*

[Prop. de Markov Fuerte](#)

$$\mathbb{E}(f(X_\tau, X_{\tau+1} \dots) | \mathcal{F}_\tau) \mathbf{1}_{\tau < \infty} = \mathbb{E}_{X_\tau}(f(X_0, X_1, \dots)) \mathbf{1}_{\tau < \infty}. \quad (1.4)$$

Esto es, la misma propiedad demostrada en el Teo. 1.1.27 (y todas las que le preceden, que son casos particulares de ella) valen también en el caso de tiempos de parada  $\tau$ , en el evento en que  $\tau$  es finito (la multiplicación por  $\mathbf{1}_{\tau < \infty}$  significa que la igualdad de las dos esperanzas se establece para  $\omega \in \Omega$  tal que  $\tau(\omega) < \infty$ ; en caso contrario las indicatrices hacen que la igualdad sea trivial).

*Demostración.* La idea es reducirnos a la propiedad Markov ya conocida. Sea  $\varphi: I \rightarrow \mathbb{R}$  dada por  $\varphi(i) = \mathbb{E}_i(f(X_0, X_1, \dots))$ . Como  $f$  es  $\mathcal{P}(I^{\mathbb{N}})$ -medible y  $X_n$  es  $\mathcal{F}_n$ -medible,  $\varphi(X_n)$  es  $\mathcal{F}_n$ -medible. Luego gracias al Lem. 1.2.8 tenemos que  $\mathbf{1}_{\tau < \infty} \varphi(X_\tau)$ , que es igual a  $\sum_{n \geq 0} \mathbf{1}_{\tau = n} \varphi(X_n)$ , es  $\mathcal{F}_\tau$ -medible.

En la expresión del lado izquierdo de (1.4) podemos mover la indicatriz  $\mathbf{1}_{\tau < \infty}$  adentro de la esperanza condicional (pues  $\{\tau < \infty\} \in \mathcal{F}_\tau$ ). Luego la propiedad que queremos probar es equivalente, por definición de la esperanza condicional, a pedir que para todo  $A \in \mathcal{F}_\tau$

$$\mathbb{E}(f(X_\tau, X_{\tau+1} \dots) \mathbf{1}_A \mathbf{1}_{\tau < \infty}) = \mathbb{E}(\mathbb{E}_{X_\tau}(f(X_0, X_1, \dots)) \mathbf{1}_A \mathbf{1}_{\tau < \infty}).$$

El lado derecho es igual a  $\mathbb{E}(\varphi(X_\tau) \mathbf{1}_A \mathbf{1}_{\tau < \infty})$ , y luego descomponiendo de acuerdo al valor de  $\tau$ , lo anterior es lo mismo que probar que

$$\mathbb{E}(f(X_n, X_{n+1} \dots) \mathbf{1}_A \mathbf{1}_{\tau = n}) = \mathbb{E}(\mathbb{E}_{X_n}(f(X_0, X_1, \dots)) \mathbf{1}_A \mathbf{1}_{\tau = n})$$

para cada  $A \in \mathcal{F}_\tau$  y cada  $n \geq 0$ , gracias a que  $f$  es acotada y el Teorema de Convergencia Dominada. Pero  $A \cap \{\tau = n\} \in \mathcal{F}_n$ , y entonces por definición de esperanza condicional nuevamente y la propiedad de Markov usual (Teo. 1.1.27) la igualdad anterior se tiene, lo que termina la demostración.  $\square$

**Observación 1.2.10.** Hemos escrito la propiedad de Markov para  $f$  acotada por simplicidad, pero la demostración puede adaptarse a otros casos, por ejemplo  $f \geq 0$  usando Teorema de Convergencia Monótona.

**Ejemplo 1.2.11.** La hipótesis de que  $\tau$  sea t.d.p. es necesaria para que una propiedad como la del teorema anterior sea válida. Por ejemplo podemos considerar la marcha aleatoria simple en  $\mathbb{Z}$  del Ej. 1.1.4 partiendo desde 0 y definir el tiempo  $\sigma = \sup\{k \geq 0: X_k \leq 0\}$ , que claramente depende del futuro de la trayectoria (pues necesita saber que la marcha nunca más será negativa), por lo que no es tiempo de parada. Si asumimos  $\sigma < \infty$  y observamos la trayectoria de marcha a partir de tiempo  $\sigma$ , donde tenemos necesariamente  $X_\sigma = 0$ , lo que vemos no es una copia independiente de la cadena comenzada en 0, pues si a tiempo  $\sigma$  es la última vez que estuvimos en  $\mathbb{Z}_{\leq 0}$  entonces el resto de la trayectoria debe mantenerse sobre 0 (por ejemplo,  $X_{\sigma+1} = 1$ ).

### 1.2.3. Ejemplo: probabilidades de pasada

La propiedad de Markov fuerte es una de las herramientas más importantes en el estudio de cadenas de Markov. Para ilustrar esto veremos un ejemplo. Consideremos de nuevo la marcha aleatoria simple en  $\mathbb{Z}$  del Ej. 1.1.4 partiendo desde 0 y definamos el siguiente tiempo de parada (ya introducido en el Ej. 1.2.2): para  $a > 0$  dado,

$$\tau_a = \inf\{n \geq 0: X_n \geq a\},$$

es decir, el primer tiempo en que la marcha alcanza el nivel  $a$ . La pregunta que queremos responder es la siguiente: ¿cuál es la probabilidad de que la cadena llegue a  $a$  alguna vez, es decir, del evento  $\{\tau_a < \infty\}$ ?

El truco es considerar la función generadora de momentos de  $\tau_a$ : para  $s \in (0, 1)$  definimos

$$\phi_a(s) = \mathbb{E}_0(s^{\tau_a}) = \sum_{k \geq 0} s^k \mathbb{P}_0(\tau_a = k). \quad (1.5)$$

Partimos escribiendo

$$\phi_a(s) = \mathbb{E}_0(s^{\tau_a} \mathbf{1}_{\tau_1 < \infty}) + \mathbb{E}_0(s^{\tau_a} \mathbf{1}_{\tau_1 = \infty}) = \mathbb{E}_0(s^{\tau_a} \mathbf{1}_{\tau_1 < \infty}),$$

pues si la marcha nunca alcanza el nivel 1 (que es lo que indica el evento  $\{\tau_1 = \infty\}$ ) entonces tampoco podrá alcanzar el nivel  $a \geq 1$ , lo que implica  $\tau_a = \infty$  y luego  $s^{\tau_a} = 0$ . Ahora usando propiedades de la esperanza condicional tenemos

$$\phi_a(s) = \mathbb{E}_0(\mathbb{E}(s^{\tau_a} \mathbf{1}_{\tau_1 < \infty} | \mathcal{F}_{\tau_1})) = \mathbb{E}_0(\mathbb{E}(s^{\tau_a} | \mathcal{F}_{\tau_1}) \mathbf{1}_{\tau_1 < \infty}) = \mathbb{E}_0(\mathbb{E}(s^{\tau_a - \tau_1} | \mathcal{F}_{\tau_1}) s^{\tau_1} \mathbf{1}_{\tau_1 < \infty}).$$

Ahora notamos dos cosas: 1.  $s^{\tau_a - \tau_1}$  es acotado (por 1, porque  $\tau_a \geq \tau_1$ ), y 2. la variable aleatoria  $\tau_a - \tau_1$  solo depende de la trayectoria de la marcha aleatoria a partir de tiempo  $\tau_1$  (pues es el tiempo que le toma llegar al nivel  $a$  luego de haber alcanzado el nivel 1, que sucede precisamente a tiempo  $\tau_1$ ). Luego podemos aplicar la propiedad de Markov fuerte al lado derecho anterior para deducir que es igual a

$$\mathbb{E}_0(\mathbb{E}_{X_{\tau_1}}(s^{\tau_a}) s^{\tau_1} \mathbf{1}_{\tau_1 < \infty}).$$

**Ejercicio 1.2.12.** Justifique esto cuidadosamente. Para ello escriba  $s^{\tau_a - \tau_1}$  como  $f(X_{\tau_1} + 1, X_{\tau_1} + 2, \dots)$  para alguna  $f: I^{\mathbb{N}} \rightarrow \mathbb{R}$  adecuada y use (1.4).

Lo anterior es igual a

$$\mathbb{E}_0(\mathbb{E}_1(s^{\tau_a}) s^{\tau_1} \mathbf{1}_{\tau_1 < \infty}) = \mathbb{E}_1(s^{\tau_a}) \mathbb{E}_0(s^{\tau_1} \mathbf{1}_{\tau_1 < \infty}) = \mathbb{E}_0(s^{\tau_{a-1}}) \mathbb{E}_0(s^{\tau_1}) = \phi_{a-1}(s) \phi_1(s),$$

donde usamos que la marcha aleatoria es invariante bajo traslaciones, por lo que el tiempo de pasada a  $a$  partiendo desde 1 tiene la misma distribución que el tiempo de pasada a  $a - 1$  partiendo desde 0. Si repetimos esto inductivamente obtenemos

$$\phi_a(s) = \phi_1(s)^a.$$

Ahora descomponiendo de acuerdo al primer paso, tenemos

$$\phi_1(s) = p \mathbb{E}_0(s^{\tau_1} | X_1 = 1) + q \mathbb{E}_0(s^{\tau_1} | X_1 = -1) = ps + q \mathbb{E}_{-1}(s^{\tau_1 + 1}),$$

donde aplicamos la propiedad de Markov (débil); ejercicio: justifique este último paso; ¿por qué sale un término  $+1$  en el último exponente? Para la última esperanza usamos de nuevo la invarianza bajo traslación de la marcha, de forma que es igual

Ej. de Prop. de Markov  
Fuerte en Marcha Aleat.

a  $\mathbb{E}_0(s^{\tau_2+1}) = s\phi_2(s) = s\phi_1(s)^2$  (lo que habíamos probado recién) y obtenemos  $\phi_1(s) = ps + qs\phi_1(s)^2$ , lo que lleva a

$$\phi_1(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2qs}.$$

Un análisis sencillo muestra que el lado derecho anterior está en  $[0, 1]$  para todo  $s \in [0, 1]$  (como debe ser por de  $\phi_1(s)$ ) para la raíz con signo menos, y de acá concluimos que

$$\phi_a(s) = \left( \frac{1 - \sqrt{1 - 4pqs^2}}{2qs} \right)^a.$$

De aquí podemos sacar por ejemplo la distribución de  $\tau_a$ , mediante

$$\mathbb{P}_0(\tau_a = k) = \frac{1}{k!} \frac{d^k}{ds^k} \phi_a(s) \Big|_{s=0}$$

(comparando con el lado derecho de (1.5) y usando unicidad de la expansión en serie de potencias en  $s$ ).

Para el problema que nos interesaba tenemos, usando (1.5) nuevamente,

$$\mathbb{P}_0(\tau_a < \infty) = \sum_{k \geq 0} \mathbb{P}_0(\tau_a = k) = \lim_{s \rightarrow 1^-} \phi_a(s) = \left( \frac{1 - \sqrt{1 - 4pq}}{2q} \right).$$


El argumento de la raíz es  $(1 - 2p)^2$ , y de esto concluimos que

$$\mathbb{P}_0(\tau_a < \infty) = \begin{cases} \left( \frac{p}{q} \right)^a & \text{si } p < \frac{1}{2}, \\ 1 & \text{si } p \geq \frac{1}{2}. \end{cases}$$

Notar que, como debiera esperarse, cuando la probabilidad de salto a la derecha es mayor que  $\frac{1}{2}$ , la marcha alcanza cualquier nivel  $a > 0$  eventualmente con probabilidad 1. Lo que también se concluye, y que es quizás menos claro intuitivamente, es que para el caso  $p < \frac{1}{2}$  siempre hay una probabilidad positiva de nunca alcanzar  $a$ .

**Ejemplo 1.2.13.** Para la cadena de Markov dada por la ruina del jugador, Ej. 1.1.5, demuestre que la probabilidad de que el jugador se quede eventualmente sin dinero si parte en 1 es  $(1 - p)/p$  si  $p > \frac{1}{2}$  y 1 si  $p \leq \frac{1}{2}$ . : si bien puede repetir el análisis de arriba, esto no es necesario y puede comparar directamente con el ejemplo anterior. ¿Puede calcular el tiempo esperado de llegada a 0 para el caso  $p > \frac{1}{2}$ ?

### 1.3. Probabilidades de absorción

Ver P2 aux. 3 .

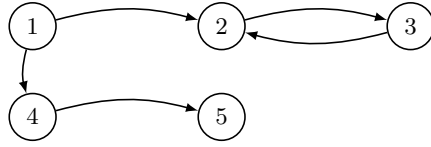
### 1.4. Estructura de clases

#### 1.4.1. Clases e irreducibilidad

**Definición 1.4.1.** Sea  $X$  una cadena de Markov y sean  $i, j \in I$ . Decimos que  $i$  lleva a  $j$ , y escribimos  $i \rightarrow j$ , si  $\mathbb{P}_i(\exists n \geq 0: X_n = j) > 0$ . Decimos que  $i$  y  $j$  se comunican, y escribimos  $i \leftrightarrow j$ , si  $i \rightarrow j$  y  $j \rightarrow i$ .

Def. “ $i$  lleva a  $j$ ” ( $i \rightarrow j$ )  
Def. “ $i$  y  $j$  se comunican” ( $i \leftrightarrow j$ )

**Ejemplo 1.4.2.** Consideremos una cadena de Markov a valores en  $I = \{1, 2, 3, 4, 5\}$ . En el siguiente grafo dirigido representamos por aristas entre nodos las transiciones que suceden con probabilidad estrictamente positiva (es decir, ponemos una arista de  $i$  a  $j$  si y solo si  $p_{ij} > 0$ ); usaremos esta convención varias veces a lo largo de las notas.



Para esta cadena tenemos  $1 \rightarrow 2$ ,  $2 \leftrightarrow 3$  y  $1 \rightarrow 5$ , pero  $4 \nrightarrow 2$ .

**Proposición 1.4.3.** Las siguientes son equivalentes:

Caract. de  $i \rightarrow j$

- (1)  $i \rightarrow j$ .
- (2) Existen  $n \in \mathbb{N}$  e  $i_1, \dots, i_n$  tales que  $p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n} > 0$ .
- (3) Existe  $n \geq 0$  tal que  $\mathbb{P}_i(X_n = j) > 0$ .

*Demostración.* Que (2) implica (1) es directo. Para ver que (1) implica (3) probamos la contrarecíproca: asumiendo que  $\mathbb{P}_i(X_n = j) = 0$  para todo  $n$ , tenemos que

$$\mathbb{P}_i(\exists n \geq 0: X_n = j) \leq \sum_{n \geq 0} \mathbb{P}_i(X_n = j) = 0.$$

Veamos ahora que (3) implica (2): tenemos que

$$\begin{aligned} 0 < \mathbb{P}_i(X_n = j) &= \sum_{i_1, \dots, i_{n-1}} \mathbb{P}_i(X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_1, \dots, i_{n-1}} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} j}, \end{aligned}$$

y luego alguno de los sumandos debe ser positivo.  $\square$

**Proposición 1.4.4.** La relación  $\rightarrow$  es reflexiva y transitiva y la relación  $\leftrightarrow$  es de equivalencia.

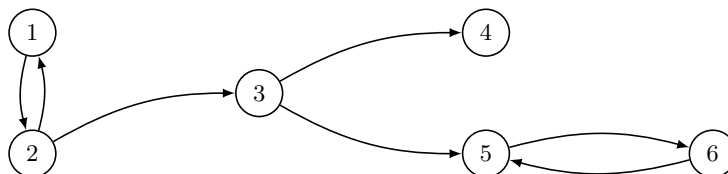
$\rightarrow$  es refl. y trans. y  $\leftrightarrow$  es de equiv.

*Demostración.* Demostremos que  $\rightarrow$  es transitiva, el resto queda de ejercicio. Supongamos que  $i \rightarrow j$  y  $j \rightarrow k$ , de manera que por la proposición anterior existen  $m, n \in \mathbb{N}$  tales que  $p_{ij}^{(m)} > 0$  y  $p_{jk}^{(n)} > 0$ . Por Chapman-Kolmogorov tenemos  $p_{ik}^{(m+n)} = \sum_{\ell \in \mathbb{N}} p_{i\ell}^{(m)} p_{\ell k}^{(n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0$ , y entonces  $i \rightarrow k$ .  $\square$

**Definición 1.4.5. (Clase y clase cerrada)** La clase de  $i \in I$ , que denotaremos por  $C(i)$ , es la clase de equivalencia de  $i$  en  $I$  con respecto a la relación  $\leftrightarrow$ . Decimos que una clase  $C$  es *cerrada* si cada vez que  $i \in C$  y  $i \rightarrow j$  se tiene necesariamente que  $j \in C$ .

Clase y clase cerrada

**Ejemplo 1.4.6.**



Para esta cadena hay tres clases:  $\{1, 2, 3\}$ ,  $\{4\}$ ,  $\{5, 6\}$ . Las últimas dos son cerradas.

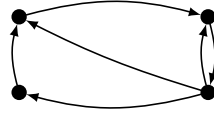


**Definición 1.4.7. (Irreducibilidad)** Decimos que una cadena de Markov  $X$  es *irreducible*, o alternativamente que su matriz de transición  $P$  lo es, si  $X$  tiene una única clase.

Irreducibilidad

Notar que si  $X$  es irreducible, su única clase es necesariamente cerrada.

**Ejemplo 1.4.8.**



Esta cadena es irreducible. La marcha aleatoria simple en  $\mathbb{Z}$  también lo es. Pero



no es irreducible, pues los dos estados de más a la izquierda forman una clase separada del resto de los estados de la cadena.

La noción de irreducibilidad será clave en nuestro estudio de cadenas de Markov: típicamente asumiremos que nuestras cadenas son irreducibles. La idea es que cuando hay más de una clase, la cadena puede estudiarse por separado en cada una de ellas.

### 1.4.2. Periodo

Otra noción importante, que asumiremos a menudo de nuestras cadenas, es la de aperiodicidad. Partimos por la siguiente definición.

**Definición 1.4.9. (Periodo)** El *periodo* de  $i \in I$  es

Periodo

$$d(i) = \text{mcd}\{n > 0: p_{ii}^{(n)} > 0\},$$

donde mcd denota al máximo común divisor (y tomamos  $\text{mcd}(\emptyset) = 0$ ).

**Ejemplo 1.4.10.** La marcha aleatoria en  $\mathbb{Z}$  tiene período 2 (ejercicio: verifíquelo); esto implica en este caso, por ejemplo, que si la marcha comienza en un estado par, entonces estará en un estado par en cada instante de tiempo par y en uno impar en cada instante de tiempo impar. La primera cadena del Ej. 1.4.8 tiene periodo 1.

**Definición 1.4.11.** Decimos que una propiedad definida para estados de una cadena de Markov es *propiedad de clase* si se tiene que cada vez que  $i$  satisface la propiedad y  $j \in C(i)$  entonces  $j$  satisface la propiedad.

Propiedad de Clase

**Proposición 1.4.12.** El periodo es propiedad de clase (es decir,  $i \leftrightarrow j$  implica que  $d(i) = d(j)$ ).

*Demostración.* Tomemos  $i \neq j$  en la misma clase, de forma que existen  $m$  y  $n$  tales que  $p_{ij}^{(m)} > 0$  y  $p_{ji}^{(n)} > 0$ . Además sabemos que existe  $r > 0$  tal que  $p_{ii}^{(r)} > 0$  (por ejemplo podemos tomar  $r = m + n$  por Chapman-Kolmogorov). Luego para todo  $k \geq 0$  Chapman-Kolmogorov da

$$p_{jj}^{(n+m+kr)} \geq p_{ji}^{(m)} (p_{ii}^{(r)})^k p_{ij}^{(n)}$$

(ejercicio: verifique esto), y concluimos que  $d(j)$  es divisor de  $m + n + kr$  para todo  $k \geq 0$ . En particular,  $d(j)$  es divisor de  $(m + n + 2r) - (m + n + r) = r$ . Obtenemos como conclusión que para cada  $r > 0$  tal que  $p_{ii}^{(r)} > 0$  se tiene que  $d(j)$  divide a  $r$  y luego, en particular,  $d(j)$  divide a  $d(i)$ . Intercambiando los roles de  $i$  y  $j$  concluimos que  $d(i)$  divide a  $d(j)$ , y luego  $d(i) = d(j)$ .  $\square$

**Definición 1.4.13. (Aperiodicidad)** Decimos que un estado  $i \in I$  es *aperiódico* si  $d(i) = 1$ . Decimos que una cadena de Markov es *aperiódica* si todos sus estados lo son.

Aperiodicidad

Ciertos aspectos del estudio de cadenas de Markov se simplifican al considerar cadenas aperiódicas, que evitan comportamientos como los descritos en el Ex. 1.4.10 para la marcha aleatoria simple en  $\mathbb{Z}$ . Sin embargo, y tal como el mismo ejemplo muestra, muchas cadenas naturales no son aperiódicas, por lo que estas sí tendrán que ser consideradas en nuestro estudio.

El siguiente resultado se demuestra de manera similar a la Prop. 1.4.12:

**Lema 1.4.14.** Sea  $X$  irreducible y sea  $i \in I$ . Si  $i$  es aperiódico entonces  $X$  lo es. Más aún, en este caso se tiene que para todo  $i, j \in I$  existe un  $n_0 \geq 0$  tal que  $p_{ij}^{(n)} > 0$  para todo  $n \geq n_0$ .

**Ejercicio 1.4.15.** Demuestre el lema anterior.

## 1.5. Recurrencia y transiencia

Ahora examinaremos uno de los aspectos claves en el estudio de cadenas de Markov: la distinción entre recurrencia y transiencia.

### 1.5.1. Definición y propiedades básicas

**Definición 1.5.1. (Tiempo de retorno)** Dada una cadena de Markov  $X$ , definimos su *tiempo de retorno* (o *tiempo de llegada*) a  $i \in I$  como

Tiempo de retorno

$$\tau_i = \inf\{n \geq 1: X_n = i\}.$$

Más generalmente, si  $A \subseteq I$ , escribimos

$$\tau_i = \inf\{n \geq 1: X_n \in A\}.$$

Los anteriores son claramente tiempos de parada (ejercicio: demuéstrello). Para efectos del estudio de recurrencia de cadenas de Markov que haremos a continuación, típicamente estaremos interesados en el caso en que la cadena comienza en  $i$  y nos preguntamos por el primer tiempo en que vuelve a  $i$ , lo que explica llamar a  $\tau_i$  el tiempo de retorno a  $i$  (pero, por supuesto, la definición tiene sentido para la cadena comenzando en cualquier estado  $j$ ).

**Definición 1.5.2.** El *número de visitas* a  $i$  es

$$N_i = \sum_{n \geq 0} \mathbf{1}_{X_n = i}.$$

**Definición 1.5.3. (Recurrencia y transiencia)** Decimos que  $i \in I$  es *recurrente* si

Recurrencia

$$\mathbb{P}_i(\tau_i < \infty) = 1.$$

En caso contrario, decimos que  $i$  es *transiente*.

Transiencia

En palabras, un estado es recurrente si partiendo de  $i$  la cadena vuelve a  $i$  en el futuro con probabilidad 1. Nuestro objetivo es demostrar la siguiente dicotomía entre recurrencia y transiencia. Enunciaremos el resultado por separado en los casos recurrente y transiente, aunque son esencialmente equivalentes.

**Teorema 1.5.4.** Sea  $i \in I$ .

Teo: Dicotomía entre  
recurrencia y transiencia

1. ( $i$  recurrente) Las siguientes son equivalentes:

- (i)  $\mathbb{P}_i(\tau_i < \infty) = 1$ .
- (ii)  $\mathbb{P}_i(N_i = \infty) = 1$ .
- (iii)  $\mathbb{E}_i(N_i) = \infty$ .

2. ( $i$  transiente) Las siguientes son equivalentes:

- (i)  $\mathbb{P}_i(\tau_i < \infty) < 1$ .
- (ii)  $\mathbb{P}_i(N_i = \infty) = 0$ .
- (iii)  $\mathbb{E}_i(N_i) < \infty$ .

El contenido esencial de este teorema es el hecho de que la recurrencia de un estado  $i \in I$  equivale a que, partiendo de  $i$ , la cadena visita  $i$  una infinidad de veces.

Partimos con un lema sencillo:

**Lema 1.5.5.**  $\mathbb{E}_i(N_i) = \sum_{n \geq 0} p_{ii}^{(n)}$ .

$\mathbb{E}_i(N_i) = \sum_{n \geq 0} p_{ii}^{(n)}$ .

*Demostración.*

$$\mathbb{E}_i(N_i) = \mathbb{E}_i(\sum_{n \geq 0} \mathbf{1}_{X_n=i}) = \sum_{n \geq 0} \mathbb{E}_i(\mathbf{1}_{X_n=i}) = \sum_{n \geq 0} \mathbb{P}_i(X_n = i) = \sum_{n \geq 0} p_{ii}^{(n)},$$

donde en la segunda igualdad usamos el Teorema de Convergencia Monótona.  $\square$

Como consecuencia de este lema y el teorema obtenemos un criterio explícito para chequear si un estado es recurrente o transiente:

**Corolario 1.5.6.** Un estado  $i \in I$  es recurrente si y solo si  $\sum_{n \geq 0} p_{ii}^{(n)} = \infty$ .

Caract. para det. si un  
est. es recurrente

**Ejemplo 1.5.7.** Veremos más adelante ( $\triangle$ ) que para la marcha aleatoria simple en  $\mathbb{Z}$  se tiene  $p_{00}^{(2n)} \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$  (mientras que  $p_{00}^{(2n)} = 0$  para todo  $n$ ). En el caso simétrico,  $p = \frac{1}{2}$ , tenemos  $4p(1-p) = 1$ , y por lo tanto  $\sum_{n \geq 0} p_{ii}^{(n)} = \infty$ , de donde concluimos que la marcha aleatoria es recurrente. En el caso asimétrico,  $p \neq \frac{1}{2}$ , como la función  $p \mapsto p(1-p)$  alcanza su máximo en  $[0, 1]$  en  $p = \frac{1}{2}$ , tenemos  $4p(1-p) < 1$  y por lo tanto  $\sum_{n \geq 0} p_{ii}^{(n)} < \infty$ , de donde se sigue que la marcha aleatoria es transiente. Ejercicio: ¿Es este resultado consistente con lo que obtuvimos en el ejemplo de la Sección 1.2.3?

Para la demostración del Teorema 1.5.4 utilizaremos una idea que se conoce a veces como descomposición por tiempo de llegada. Esto lo hacemos en el resultado siguiente. Sea

Descomposición por  
tiempo de llegada

$$u_{ij}^{(n)} = \mathbb{P}_i(\tau_j = n). \quad (1.6)$$

Notemos que  $u_{ii}^{(0)} = 0$ , por definición de  $\tau_i$ .

**Lema 1.5.8.** Para todo  $n \geq 0$  se tiene que

Fórm. para  $p_{ij}^{(n)}$

$$p_{ij}^{(n)} = \sum_{k=1}^n u_{ij}^{(k)} p_{jj}^{n-k} \mathbf{1}_{n>0} + \mathbf{1}_{n=0, i=j}.$$

*Demostración.* El caso  $n = 0$  es trivial, así que asumamos que  $n > 0$ . Tenemos

$$\begin{aligned} p_{ij}^{(n)} &= \mathbb{P}_i(X_n = j) = \mathbb{E}_i(\mathbf{1}_{X_n=j} \mathbf{1}_{\tau_j \leq n}) = \mathbb{E}_i(\mathbb{E}(\mathbf{1}_{X_n=j} | \mathcal{F}_{\tau_j}) \mathbf{1}_{\tau_j \leq n}) \\ &= \sum_{k=1}^n \mathbb{E}_i(\mathbb{E}(\mathbf{1}_{X_n=j} | \mathcal{F}_{\tau_j}) \mathbf{1}_{\tau_j=k}), \end{aligned}$$

donde usamos el hecho que  $X_n = j$  implica que  $\tau_j \leq n$ . Ahora usamos la propiedad de Markov fuerte en la esperanza condicional. Para ello reescribimos el término  $k$ -ésimo de la suma al lado derecho como  $\mathbb{E}_i(\mathbb{E}(\mathbf{1}_{X_{n-k+\tau_j=j}}|\mathcal{F}_{\tau_j})\mathbf{1}_{\tau_j=k})$  (ejercicio: ¿por qué podemos hacer esto?), de manera que la esperanza condicional que aparece adentro es igual a  $\mathbb{E}_{X_{\tau_j}}(\mathbf{1}_{X_{n-k=j}})$ . Luego

$$p_{ij}^{(n)} = \sum_{k=1}^n \mathbb{E}_i(\mathbb{P}_j(X_{n-k} = j)\mathbf{1}_{\tau_j=k}) = \sum_{k=1}^n p_{jj}^{(n-k)} \mathbb{P}_i(\tau_j = k) = \sum_{k=1}^n p_{jj}^{(n-k)} u_{ij}^{(k)},$$

que es lo queríamos probar.  $\square$

Estamos listos para demostrar la dicotomía entre recurrencia y transiencia. Antes de hacerlo introducimos una notación que usaremos en la demostración.

**Definición 1.5.9.** Definimos en *tiempo del  $k$ -ésimo retorno a  $i$* ,  $\tau_i^{(k)}$ , inductivamente como

$$\tau_i^{(1)} = \tau_i, \quad \tau_i^{(k)} = \inf\{n > \tau_i^{(k-1)} : X_n = i\} \quad \text{para } k > 1.$$

Tiempo del  $k$ -ésimo retorno

*Demostración del Teorema 1.5.4.* Comenzamos por el caso recurrente, probando la equivalencia entre (i) y (iii). Sean

$$p_{ii}(z) = \sum_{n \geq 0} p_{ii}^{(n)} z^n, \quad u_{ii}(z) = \sum_{n \geq 0} u_{ii}^{(n)} z^n.$$

Como  $p_{ii}^{(n)}$  y  $u_{ii}^{(n)}$  están en  $[0, 1]$  para todo  $n$ , ambas series son absolutamente convergentes si  $|z| < 1$ . Para tales  $z$  tenemos entonces

$$u_{ii}(z)p_{ii}(z) = \sum_{n \geq 0} \sum_{k=0}^n u_{ii}^{(k)} z^k p_{ii}^{(n-k)} z^{n-k} = \sum_{n \geq 0} z^n \sum_{k=0}^n u_{ii}^{(k)} p_{ii}^{(n-k)}$$

(simplemente por regla de multiplicación de series de potencias). Sea  $S_n = \sum_{k=0}^n u_{ii}^{(k)} p_{ii}^{(n-k)}$ , que es el coeficiente  $n$ -ésimo de la serie de potencias del lado derecho arriba. Si  $n = 0$  tenemos  $S_n = u_{ii}^{(0)} p_{ii}^{(0)} = 0$ . En caso contrario, gracias al lema anterior tenemos  $S_n = u_{ii}^{(0)} p_{ii}^{(0)} + p_{ii}^{(n)}$ . Concluimos entonces que para todo  $n \geq 0$  se tiene  $S_n = p_{ii}^{(n)} \mathbf{1}_{n>0} = p_{ii}^{(n)} - \mathbf{1}_{n=0}$  (ejercicio: ¿por qué se tiene esto último?). Usando esto arriba obtenemos

$$u_{ii}(z)p_{ii}(z) = \sum_{n \geq 0} (p_{ii}^{(n)} - \mathbf{1}_{n=0}) z^n = \sum_{n \geq 0} p_{ii}^{(n)} z^n - 1 = p_{ii}(z) - 1$$

o, equivalentemente,

$$p_{ii}(z)(1 - u_{ii}(z)) = 1$$

para todo  $|z| < 1$ . Ahora notamos que

$$\begin{aligned} \lim_{z \rightarrow 1^-} p_{ii}(z) &= \sum_{n \geq 0} p_{ii}^{(n)} = \mathbb{E}_i(N_i), \\ \lim_{z \rightarrow 1^-} u_{ii}(z) &= \sum_{n \geq 0} u_{ii}^{(n)} = \sum_{n \geq 0} \mathbb{P}_i(\tau_i = n) = \mathbb{P}_i(\tau_i < \infty). \end{aligned}$$

Pero, por lo anterior,

$$\lim_{z \rightarrow 1^-} p_{ii}(z)(1 - u_{ii}(z)) = 1.$$

Luego  $\mathbb{E}_i(N_i) = \infty$  implica que necesariamente  $\mathbb{P}_i(\tau_i < \infty) = 1$  y viceversa, lo que nos dice que (i) y (iii) son equivalentes.

Que (ii) implica (iii) es directo. Veamos entonces que (i) implica (ii). Tenemos que

$$\{N_i = \infty\} = \{\tau_i^{(k)} < \infty \mid \forall k \geq 1\}.$$

Ejercicio: Pruebe lo anterior. Luego

$$\mathbb{P}_i(N_i = \infty) = \mathbb{P}_i\left(\bigcap_{k \geq 1} \{\tau_i^{(k)} < \infty\}\right) = \lim_{k \rightarrow \infty} \mathbb{P}_i(\tau_i^{(k)} < \infty),$$

donde en la segunda igualdad usamos que los eventos  $\{\tau_i^{(k)} < \infty\}$  son decrecientes. De esta forma la implicancia que queremos probar equivale a probar que

$$\mathbb{P}_i(\tau_i < \infty) = 1 \implies \lim_{k \rightarrow \infty} \mathbb{P}_i(\tau_i^{(k)} < \infty) = 1.$$

Y el lado derecho equivale a pedir que  $\mathbb{P}_i(\tau_i^{(k)} < \infty) = 1$  para todo  $k \geq 1$  (ejercicio: ¿por qué?). Probaremos esto por inducción. El caso  $k = 1$  es la hipótesis. Para el paso inductivo usamos el siguiente:

**Ejercicio 1.5.10.** Demuestre que para todo  $k \geq 1$ ,

$$\mathbb{P}_i(\tau_i^{(k+1)} < \infty) = \mathbb{P}_i(\tau_i < \infty) \mathbb{P}_i(\tau_i^{(k)} < \infty).$$

La conclusión ahora es directa.

Resta probar el caso transiente. Del caso recurrente obtenemos directamente que

$$\mathbb{P}_i(\tau_i < \infty) < 1 \iff \mathbb{P}_i(N_i = \infty) < 1 \iff \mathbb{E}_i(N_i) < \infty.$$

Luego lo que falta demostrar es que las propiedades anteriores equivalen a  $\mathbb{P}_i(N_i = \infty) = 0$ , pero esto es directo: esta propiedad implica trivialmente que  $\mathbb{P}_i(N_i = \infty) < 1$ , mientras que la propiedad  $\mathbb{E}_i(N_i) < \infty$  implica que  $\mathbb{P}_i(N_i = \infty) = 0$ .  $\square$

La siguiente es una consecuencia más o menos directa del argumento usado en la demostración para probar que (ii) implica (iii):

**Proposición 1.5.11.** Sea  $\theta_i = \mathbb{P}_i(\tau_i < \infty)$ . Entonces para todo  $k \geq 1$ ,

$$\mathbb{P}_i(\tau_i^{(k)} < \infty) = \theta_i^k.$$

En particular, en el caso  $i$  transiente ( $\theta_i < 1$ ) tenemos

$$\mathbb{P}_i(N_i = k) = \theta_i^k (1 - \theta_i) \quad y \quad \mathbb{E}_i(N_i) = \frac{\theta_i}{1 - \theta_i},$$

es decir,  $N_i$  es una variable  $\text{Geom}[\theta_i]$ .

**Ejercicio 1.5.12.** Demuestre la Proposición 1.5.11 y asegúrese de comprender su contenido (¿por qué es natural esperar que  $N_i$  sea una variable geométrica?).

### 1.5.2. Más propiedades

En el resto de esta sección profundizaremos nuestro estudio de recurrencia y transiencia.

**Proposición 1.5.13.** La recurrencia es propiedad de clase, i.e. si  $i \leftrightarrow j$  entonces  $i$  es recurrente si y solo si  $j$  es recurrente.

Recurrencia es prop. de clase

*Demostración.* Supongamos que  $i \leftrightarrow j$ , de manera que existen  $m, n \geq 0$  tal que  $p_{ij}^{(m)} > 0$  y  $p_{ji}^{(n)} > 0$ , y asumamos que  $i$  es transiente. Por Chapman-Kolmogorov tenemos  $p_{ii}^{(m+n+r)} \geq p_{ij}^{(m)} p_{jj}^{(r)} p_{ji}^{(n)}$  para cada  $r \geq 0$ , y luego

$$\mathbb{E}_j(N_j) = \sum_{r \geq 0} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(m)} p_{ji}^{(n)}} \sum_{r \geq 0} p_{ii}^{(m+n+r)} \leq \frac{1}{p_{ij}^{(m)} p_{ji}^{(n)}} \mathbb{E}_i(N_i) < \infty,$$

pues  $i$  es transiente. Deducimos entonces que  $j$  es transiente. La otra implicancia sale igual, intercambiando  $i$  y  $j$ .  $\square$

En vista del resultado anterior hacemos la siguiente:

**Definición 1.5.14.** Decimos que una clase  $C$  es *recurrente* si algún (o equivalentemente todo)  $i \in C$  lo es. Decimos que la cadena  $X$  es *recurrente*, o que su matriz de transición  $P$  lo es, si todo  $i \in I$  es recurrente.

**Proposición 1.5.15.**

- (i) Para  $i \neq j$  se tiene  $i \rightarrow j$  si y solo si  $\mathbb{P}_i(\tau_j < \infty) > 0$ .
- (ii)  $\mathbb{P}_i(N_j = \infty) = \mathbb{P}_i(\tau_j < \infty) \mathbb{P}_j(N_j = \infty)$ .
- (iii) Si  $i \rightarrow j$  e  $i$  es recurrente entonces  $\mathbb{P}_i(\tau_j < \infty) = 1$ .

*Demostración.* La parte (i) es trivial.

**Ejercicio 1.5.16.** Use la propiedad de Markov para demostrar (ii).

La parte (iii) es la más interesante. Partimos con la descomposición

$$\mathbb{P}_i(\tau_j < \infty) = \mathbb{P}_i(\tau_j < \tau_i) + \mathbb{P}_i(\tau_i \leq \tau_j, \tau_j < \infty).$$

Para el segundo término del lado derecho usamos la propiedad de Markov fuerte a tiempo  $\tau_i$ :

$$\begin{aligned} \mathbb{P}_i(\tau_i \leq \tau_j, \tau_j < \infty) &= \mathbb{E}_i(\mathbb{E}(\mathbf{1}_{\tau_j < \infty} | \mathcal{F}_{\tau_i}) \mathbf{1}_{\tau_i \leq \tau_j}) = \mathbb{E}_i(\mathbb{E}_{X_{\tau_i}}(\mathbf{1}_{\tau_j < \infty}) \mathbf{1}_{\tau_i \leq \tau_j}) \\ &= \mathbb{E}_i(\mathbb{P}_i(\tau_j < \infty) \mathbf{1}_{\tau_i \leq \tau_j}) = \mathbb{P}_i(\tau_j < \infty) \mathbb{P}_i(\tau_i \leq \tau_j) \end{aligned}$$

donde, en la primera igualdad, usamos que  $\{\tau_i \leq \tau_j\} \in \mathcal{F}_{\tau_i}$ . Deducimos que  $\mathbb{P}_i(\tau_j < \infty) = \mathbb{P}_i(\tau_j < \tau_i) + \mathbb{P}_i(\tau_j < \infty) \mathbb{P}_i(\tau_i \leq \tau_j)$ , que podemos reescribir como

$$\mathbb{P}_i(\tau_j < \infty) \mathbb{P}_i(\tau_j < \tau_i) = \mathbb{P}_i(\tau_j < \tau_i).$$

Luego basta probar que  $\mathbb{P}_i(\tau_j < \tau_i) > 0$ , que es consecuencia de la hipótesis  $i \rightarrow j$  y se deja como ejercicio.  $\square$

**Ejercicio 1.5.17.** Termine la demostración del resultado anterior.

**Corolario 1.5.18.** Sea  $X$  una cadena de Markov irreducible. Entonces  $X$  visita todo  $i \in I$  una infinidad de veces con probabilidad 1 (es decir,  $\mathbb{P}(N_i = \infty \forall i \in I) = 1$ ).

**Ejercicio 1.5.19.** Demuestre lo anterior.

**Proposición 1.5.20.** *Toda clase recurrente es cerrada.*

Toda clase Recurrente es cerrada

*Demostración.* Sea  $C$  una clase recurrente y sean  $i, j$  tales que  $i \in C$  e  $i \rightarrow j$ . Debemos demostrar entonces que  $j \in C$ , es decir que  $j \rightarrow i$ . Supongamos por contradicción que esto no se tiene, es decir, que  $j \nrightarrow i$ . Sabemos que hay un  $n \geq 0$  tal que  $p_{ij}^{(n)} > 0$ . Tenemos que  $\mathbb{P}_i(X_n = j, N_i < \infty) = \mathbb{P}_i(X_n = j)$ , pues  $j \nrightarrow i$ . Luego

$$\mathbb{P}_i(N_i < \infty) \geq \mathbb{P}_i(X_n = j, N_i < \infty) = \mathbb{P}_i(X_n = j) = p_{ij}^{(n)} > 0,$$

de donde se sigue que  $i$  es transiente, lo que nos da la contradicción buscada.  $\square$

**Proposición 1.5.21.** *Si  $j$  es transiente entonces  $\mathbb{E}_i(N_j) = \sum_{n \geq 0} p_{ij}^{(n)} < \infty$  para todo  $i \in I$ .*

*Demostración.* Esto lo sabemos ya en el caso  $i = j$ . Para el caso  $i \neq j$  usaremos la misma idea que en la demostración de ese caso (Teorema 1.5.4).

**Ejercicio 1.5.22.** Sean  $p_{ij}(z) = \sum_{n \geq 0} p_{ij}^{(n)} z^n$  y  $u_{ij}(z) = \sum_{n \geq 0} u_{ij}^{(n)} z^n$ , con  $u_{ij}^{(n)}$  como en (1.6). Demuestre que para todo  $|z| < 1$  se tiene

$$p_{ij}(z) = u_{ij}(z)p_{jj}(z).$$

Usando el ejercicio y tomando  $z \rightarrow 1^-$  obtenemos (usando el Teorema de Convergencia Monótona)

$$\sum_{n \geq 0} p_{ij}^{(n)} = \left( \sum_{n \geq 0} u_{ij}^{(n)} \right) \left( \sum_{n \geq 0} p_{jj}^{(n)} \right) = \mathbb{P}_i(\tau_j < \infty) \mathbb{E}_j(N_j),$$

y luego el lado derecho es finito pues  $j$  es transiente.  $\square$

**Corolario 1.5.23.** *Si  $I$  es finito entonces existe un estado recurrente. En particular, si  $I$  es finito y  $X$  es irreducible entonces  $X$  es recurrente.*

*Demostración.* Como  $I$  es finito tenemos

$$\sum_{j \in I} \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \lim_{n \rightarrow \infty} \sum_{j \in I} p_{ij}^{(n)} = 1.$$

Pero si todo  $j$  fuera transiente, la suma del lado izquierdo necesariamente sería nula (ejercicio: ¿por qué?). Deducimos entonces que debe existir algún  $j$  recurrente.  $\square$

### 1.5.3. Caracterización

Concluimos esta sección con una caracterización algebraica de recurrencia en base a propiedades de la matriz de transición  $P$ .

**Definición 1.5.24. (Funciones armónicas)** Sea  $i \in I$ . Decimos que una función  $f: I \rightarrow \mathbb{R}$  es *armónica* en  $i$  (con respecto a  $P$ ) si  $Pf(i) = \sum_{j \in I} p_{ij} f(j) = f(i)$ . Decimos que  $f$  es *superarmónica* (o *subarmónica*) en  $i$  si  $Pf(i) \leq f(i)$  (o  $Pf(i) \geq f(i)$ ).

**Teorema 1.5.25. (Caracterización de recurrencia)** *Sea  $X \sim \text{CM}(P)$  irreducible. Entonces  $X$  es recurrente si y solo si existe un  $i_0 \in I$  (equivalentemente, para todo  $i_0 \in I$ ) que satisface lo siguiente: si  $f$  es armónica en  $I \setminus \{i_0\}$  (i.e.  $\sum_j p_{ij} f(j) = f(i)$  para todo  $i \neq i_0$ ) y acotada, entonces  $f$  es constante.*

*Demostración.* Supongamos primero que la propiedad se cumple. Sea  $\tilde{\tau}_{i_0} = \inf\{k \geq 0: X_k = i_0\}$  (este es el primer tiempo de llegada a  $i_0$ , notar que permitimos que este tiempo sea 0). Definimos  $f: I \rightarrow \mathbb{R}$  mediante  $f(i_0) = 1$  y  $f(i) = \mathbb{P}_i(\tilde{\tau}_{i_0} < \infty)$  para  $i \neq i_0$ . Para  $i \neq i_0$  tenemos

$$f(i) = \sum_{j \in I} \mathbb{P}_i(\tilde{\tau}_{i_0} < \infty, X_1 = j) = p_{ii_0} + \sum_{j \neq i_0} p_{ij} \mathbb{P}_j(\tilde{\tau}_{i_0} < \infty) = p_{ii_0} + \sum_{j \neq i_0} p_{ij} f(j).$$

Ejercicio: Justifique este cálculo. Como  $f(i_0) = 1$ , lo anterior equivale a afirmar que  $f(i) = Pf(i)$  para todo  $i \neq i_0$  (esto es el resultado de la Sección 1.3  $\blacktriangle$ ). Luego  $f$  es armónica en  $I \setminus \{i_0\}$ , y es claramente acotada, de forma que debe ser constante. Pero entonces para todo  $i \neq i_0$  tenemos

$$\mathbb{P}_i(\tau_{i_0} < \infty) = \mathbb{P}_i(\tilde{\tau}_{i_0} < \infty) = f(i) = p_{ii_0} + \sum_{j \neq i_0} p_{ij} = 1.$$

Argumentando como en la demostración de Prop. 1.5.15(iii) tenemos, para  $i \neq i_0$ ,

$$\mathbb{P}_{i_0}(\tau_{i_0} < \infty) = \mathbb{P}_{i_0}(\tau_{i_0} < \tau_i) + \mathbb{P}_{i_0}(\tau_i \leq \tau_{i_0}) \mathbb{P}_i(\tau_{i_0} < \infty).$$

Acabamos de probar que el factor de más a la derecha es igual a 1, de donde obtenemos  $\mathbb{P}_{i_0}(\tau_{i_0} < \infty) = 1$ . Luego  $i_0$  es recurrente, y por irreducibilidad la cadena completa lo es.

Ahora asumamos que  $X$  es recurrente y, para  $i_0 \in I$  dado, tomemos  $f$  acotada y armónica en  $I \setminus \{i_0\}$ . Definamos una matriz de transición auxiliar  $\hat{P}$  dada por  $\hat{p}_{ij} = p_{ij}$  para  $i \neq i_0$  y  $\hat{p}_{i_0j} = \mathbf{1}_{j=i_0}$ . Afirmamos que entonces  $f$  es armónica con respecto a  $\hat{P}$  en todo  $I$ . En efecto,  $\hat{P}f(i) = Pf(i) = f(i)$  para  $i \neq i_0$ , mientras que trivialmente  $\hat{P}f(i_0) = f(i_0)$ . En otras palabras, tenemos  $\hat{P}f = f$ , e iterando obtenemos  $\hat{P}^{(n)}f = f$  para todo  $n \geq 0$ . Luego escribiendo  $f(i) = \hat{p}_{ii_0}^{(n)}f(i_0) + \sum_{j \neq i_0} \hat{p}_{ij}^{(n)}f(j)$  y aplicando desigualdad triangular deducimos que

$$|f(i) - \hat{p}_{ii_0}^{(n)}f(i_0)| \leq \sum_{j \neq i_0} \hat{p}_{ij}^{(n)}|f(j)| \leq \|f\|_\infty(1 - \hat{p}_{ii_0}^{(n)}).$$

Como consecuencia, es suficiente demostrar que  $\lim_{n \rightarrow \infty} \hat{p}_{ii_0}^{(n)} = 1$  pues, en ese caso, la desigualdad anterior entrega (en el límite  $n \rightarrow \infty$ )  $|f(i) - f(i_0)| \leq 0$ , que implica lo requerido.

Demostremos entonces que  $\hat{p}_{ii_0}^{(n)}$  converge a 1 cuando  $n \rightarrow \infty$ . El caso  $i = i_0$  es trivial (y de hecho no se necesita en el argumento de arriba), así que tomemos  $i \neq i_0$ . Sea  $\hat{\mathbb{P}}$  la medida de probabilidad asociada a la cadena  $\hat{X} \sim \text{CM}(\hat{P})$  y  $\hat{\tau}_{i_0}$  el tiempo de retorno a  $i_0$  de  $\hat{X}$ .

**Ejercicio 1.5.26.** Demuestre que  $\mathbb{P}_i(\tau_{i_0} < \infty) = \hat{\mathbb{P}}_i(\hat{\tau}_{i_0} < \infty)$ .

Luego, por hipótesis de recurrencia e irreducibilidad, tenemos  $\hat{\mathbb{P}}_i(\hat{\tau}_{i_0} < \infty) = 1$ . Pero como  $i_0$  es absorbente para  $\hat{X}$ , es decir, una vez que  $\hat{X}$  llega a  $i_0$ , se queda ahí para siempre, deducimos que

$$\begin{aligned} \hat{\mathbb{P}}_i(\hat{\tau}_{i_0} < \infty) &= \mathbb{P}(\cup_{N \geq 0} \{\exists n \leq N: \hat{X}_n = i_0\}) \\ &= \lim_{N \rightarrow \infty} \mathbb{P}(\exists n \leq N: \hat{X}_n = i_0) = \lim_{N \rightarrow \infty} \mathbb{P}(\hat{X}_N = i_0). \end{aligned}$$

Ejercicio: Justifique las igualdades anteriores. Como el lado izquierdo es igual a 1, el límite también lo es, y obtenemos lo deseado.  $\square$



**Ejemplo 1.5.27.** Estudiemos nuevamente la marcha aleatoria simple en  $\mathbb{Z}$ . Consideremos primero el caso simétrico,  $p = \frac{1}{2}$ . Elegimos  $i_0 = 0$  y suponemos que  $f$  es acotada y armónica en  $\mathbb{Z} \setminus \{0\}$ , es decir,  $Pf(i) = f(i)$  para todo  $i \neq 0$ , que se escribe como

$$\frac{1}{2}f(i+1) + \frac{1}{2}f(i-1) = f(i).$$

La recurrencia anterior puede entonces para  $i \geq 2$  (ejercicio: hágalo), y obtenemos

$$f(i) = (f(1) - f(0))i + f(0) \quad \forall i \geq 0.$$

Pero  $f$  es acotada, por lo que necesariamente  $f(1) = f(0)$  y entonces  $f(i) = f(0)$  para todo  $i \geq 0$ . El mismo argumento puede aplicarse para  $i \leq 0$ , y deducimos que  $f \equiv f(0)$ . Luego, gracias al teorema, concluimos que  $X$  es recurrente, como ya sabíamos del Ejemplo 1.5.7.

Consideremos ahora el caso asimétrico. Por simetría basta considerar  $p > \frac{1}{2}$ . Del Ejemplo 1.5.7 sabemos que la cadena es ahora transiente, por lo que debiera existir una función  $f$  que es acotada y que es armónica en  $\mathbb{Z} \setminus \{0\}$ , pero que no es constante. Veamos que esto es cierto. Para ellos definimos

$$f(n) = \left(\frac{1-p}{p}\right)^n \mathbf{1}_{n>0} + \mathbf{1}_{n \leq 0},$$

que es claramente es acotada y no constante. Para  $n > 0$  tenemos

$$\begin{aligned} Pf(n) &= (1-p)f(n-1) + pf(n+1) = (1-p) \left(\frac{1-p}{p}\right)^{n-1} + p \left(\frac{1-p}{p}\right)^{n+1} \\ &= \left(\frac{1-p}{p}\right)^n (p + (1-p)) = f(n). \end{aligned}$$

Por otro lado, la igualdad  $Pf(n) = f(n)$  para  $n < 0$  se satisface trivialmente. Deducimos entonces que  $f$  es armónica en  $\mathbb{Z} \setminus \{0\}$ , como queríamos.

Un criterio alternativo al anterior, cuya demostración dejamos de ejercicio, es el siguiente:

**Teorema 1.5.28. (Caracterización de recurrencia, versión alternativa)** Sea  $X \sim \text{CM}(P)$  irreducible. Entonces  $X$  es recurrente si y solo si las únicas funciones superarmónicas y no-negativas en  $I$  son las constantes.

**Ejercicio 1.5.29.** Utilice el criterio anterior para escribir una demostración alternativa del resultado del Corolario 1.5.23 que afirma que toda cadena irreducible a valores en  $I$  finito es recurrente. Por simplicidad, si lo desea restrínjase al caso  $|I| = 3$ .

## 1.6. Recurrencia positiva y distribuciones invariantes

**Definición 1.6.1. (Medida invariante)** Sea  $P$  una matriz estocástica y  $\mu$  una medida sobre  $I$  (que interpretamos como un vector fila indexado por  $I$ ). Decimos que  $\mu$  es una medida invariante para  $P$  (o medida  $P$ -invariante) si

$$\mu P = \mu$$

(es decir,  $\sum_i \mu_i p_{ij} = \mu_j$  para todo  $j \in I$ ). Si  $\mu$  es además una medida de probabilidad, decimos que es una *distribución invariante*. Diremos también que  $\mu$  es medida o distribución invariante para una cadena  $X$  si lo es para su matriz de transición.

Es claro que si  $\mu$  es medida invariante, entonces cualquier múltiplo de ella lo es.

Como veremos en esta y la próxima sección, la noción de medida invariante juega un rol muy importante en la teoría de cadenas de Markov. La relevancia de la definición (y la explicación de la terminología usada) es evidente, al menos para distribuciones invariantes: si  $\mu$  es una distribución para  $X \sim \text{CM}(P)$ , entonces  $\mu P^n = \mu$  para todo  $n \geq 0$ , y entonces

$$\mathbb{P}_\mu(X_n = i) = \mu(i) \quad \forall i \in I, \forall n \geq 0.$$

Las distribuciones invariantes se conocen a veces también como *distribuciones estacionarias* o *distribuciones de equilibrio*. El siguiente ejercicio explica en parte este último nombre:

**Ejercicio 1.6.2.** Sea  $I$  finito. Supongamos que existe un  $i \in I$  tal que los límites

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

existen para todo  $j$ . Entonces  $\pi$  es una distribución invariante. ¿Qué puede decir del caso  $I$  infinito? (Para esta última parte piense en una marcha aleatoria asimétrica y recuerde el Ejemplo 1.5.7).

**Observación 1.6.3.** Como  $P$  es matriz estocástica, necesariamente tiene al 1 como valor propio, asociado al vector propio por la derecha  $v \equiv 1$  (esto es  $Pv = v$ ). Como 1 es valor propio, se sigue que debe tener asociado también un vector propio por la izquierda  $\hat{v}$ ,  $\hat{v}^\top P = \hat{v}^\top$  (al menos si  $P$  fuera una matriz finita, pero lo mismo vale en el caso  $I$  numerable). Esto nos dice que  $P$  siempre tiene una medida con signo invariante; buscar una medida invariante corresponde a buscar un vector propio no-negativo de  $P$  por la izquierda, mientras que buscar una distribución invariante corresponde a pedir que este vector propio esté además en  $\ell^1(I)$  (ejercicio: ¿por qué?).

En vista de la observación anterior, no tenemos por qué esperar que toda cadena de Markov tenga una medida invariante asociada, ni mucho menos una distribución invariante. Este es uno de los puntos centrales que queremos dilucidar en esta sección. Antes, veamos dos ejemplos donde estas medidas existen.

**Ejemplo 1.6.4.** Para la marcha aleatoria simple con  $p \in (0, 1)$ , la medida uniforme en  $\mathbb{Z}$ ,  $\mu \equiv 1$ , es invariante. Sin embargo, esta medida no es finita, por lo que no puede normalizarse a una distribución invariante (más adelante estudiaremos la pregunta de si esta cadena posee o no una distribución invariante).

**Ejemplo 1.6.5.** Consideremos ahora la cadena de Ehrenfest del Ejemplo 1.1.6. Aquí  $I = \{0, \dots, k\}$  y  $P$  está dada por  $p_{k,k+1} = (r - k)/r$  y  $p_{k,k-1} = k/r$ , todo el resto de la matriz es cero. En este caso hay una distribución invariante  $\mu$ , y es la que corresponde a una variable aleatoria Binomial $[r, \frac{1}{2}]$ , es decir

$$\mu(k) = \frac{1}{2^k} \binom{r}{k}.$$

Esto puede verificarse directamente, pero también puede entenderse de manera intuitiva:  $\mu$  corresponde a elegir la urna en que se coloca cada bola de acuerdo a lanzamientos independientes de  $r$  monedas, y cambiar una bola de urna corresponde a elegir una moneda al azar y darla vuelta. Ejercicio: Demuestre directamente que  $\mu$  es distribución invariante y asegúrese de comprender la explicación intuitiva.

**Ejercicio 1.6.6.** Considere una *marcha aleatoria simple con reflexión* en  $\mathbb{Z}_{\geq 0}$ , definida de igual forma que la marcha aleatoria usual (saltando a la derecha con probabilidad  $p$  y

a la izquierda con probabilidad  $1 - p$ ), salvo que cuando la cadena intenta pasar a  $-1$ , se le fuerza a quedarse en cero en vez. Determine condiciones para la existencia de una distribución invariante y calcule dicha distribución (el Ejemplo 1.5.27 podría servirle de inspiración).

**Ejercicio 1.6.7.** Analice la existencia de medidas y distribuciones invariantes para la ruina del jugador (Ejemplo 1.1.5).

El principal resultado de esta sección mostrará que la existencia de distribuciones invariantes está íntimamente relacionada con una noción más fuerte de recurrencia, que definimos a continuación:

**Definición 1.6.8. (Recurrencia positiva)** Sea  $i$  un estado recurrente. Decimos que  $i$  es *recurrente positivo* si su tiempo de retorno esperado es finito, es decir, si  $\mathbb{E}_i(\tau_i) < \infty$ . En caso contrario, decimos que  $i$  es *recurrente nulo*.

**Teorema 1.6.9. (Existencia de distribuciones invariantes)** Sea  $X$  irreducible. Entonces las siguientes son equivalentes:

- (i) Existe un estado recurrente positivo.
- (ii) Todos los estados son recurrentes positivos.
- (iii) Existe una distribución invariante  $\pi$ .

Más aún, en este caso se tiene que la distribución invariante está dada por

$$\pi(i) = \frac{1}{\mathbb{E}_i(\tau_i)}.$$

La demostración de este teorema nos tomará bastante trabajo. Antes de pasar a ella enunciamos un resultado de unicidad en el contexto anterior.

**Teorema 1.6.10.** Sea  $X$  irreducible. Si  $X$  tiene una distribución invariante  $\pi$ , entonces toda medida invariante es un múltiplo de  $\pi$ . En particular, la distribución invariante es única.

De estos dos teoremas se sigue directamente el siguiente:

**Corolario 1.6.11.** Si  $X$  es irreducible y tiene una medida invariante con masa infinita, entonces no existe distribución invariante (y en particular la cadena no es recurrente positiva).

De esta forma tenemos, para  $X$  irreducible:

1. La existencia de una distribución invariante equivale a que la cadena sea recurrente positiva.
2. Si existe una medida invariante infinita, entonces la cadena es recurrente nula o transiente.
3. Se puede demostrar también ( $\triangleleft$ ) que si  $X$  es recurrente nula entonces existe una medida invariante infinita.

*Demostración Teorema 1.6.10.* Sea  $\pi$  una distribución invariante. Gracias al Teorema 1.6.9 tenemos  $\pi(i) > 0$  para todo  $i \in I$ . Fijemos ahora una función  $\varphi: [0, \infty) \rightarrow \mathbb{R}$  estrictamente convexa (la elección particular no es importante) y acotada. Dada una medida  $\mu$  sobre  $I$ , definimos

$$\mathcal{E}(\mu) = \sum_{j \in I} \varphi\left(\frac{\mu(j)}{\pi(j)}\right) \pi(j).$$

La demostración se basa en estudiar  $\mathcal{E}(\mu P)$ . Partimos escribiendo

$$\mathcal{E}(\mu P) = \sum_j \varphi\left(\frac{\sum_i \mu(i) p_{ij}}{\pi(j)}\right) \pi(j) = \sum_j \varphi\left(\sum_i \frac{p_{ij} \pi(i)}{\pi(j)} \frac{\mu(i)}{\pi(i)}\right) \pi(j).$$

Ahora  $\frac{\sum_i p_{ij} \pi(i)}{\pi(j)} = \frac{(\pi P)_j}{\pi(j)} = 1$ , pues  $\pi$  es invariante. Luego, como  $\varphi$  es convexa (y usando Teorema de Convergencia Dominada), tenemos

$$\mathcal{E}(\mu P) \leq \sum_j \sum_i \frac{p_{ij} \pi(i)}{\pi(j)} \varphi\left(\frac{\mu(i)}{\pi(i)}\right) \pi(j) = \sum_j \sum_i p_{ij} \pi(i) \varphi\left(\frac{\mu(i)}{\pi(i)}\right) = \sum_i \pi(i) \varphi\left(\frac{\mu(i)}{\pi(i)}\right) = \mathcal{E}(\mu).$$

Supongamos ahora que  $p_{ij} > 0$  para todo  $i, j$ . Entonces como  $\varphi$  es estrictamente convexa, la desigualdad anterior es estricta a menos que  $\frac{\mu(i)}{\pi(i)}$  sea constante. Como  $\mu$  es medida invariante, obviamente tenemos  $\mathcal{E}(\mu P) = \mathcal{E}(\mu)$ , y entonces deducimos que necesariamente  $\frac{\mu(i)}{\pi(i)}$  es constante. El paso al caso general se deja como el siguiente:

**Ejercicio 1.6.12.** Concluya la demostración. Para ello, considere la matriz de transición  $\bar{p}_{ij} = \sum_{n \geq 0} 2^{-n} p_{ij}^{(n)}$ .  $\square$

**Observación 1.6.13.** La cantidad  $\mathcal{E}(\mu)$  que introducimos en la demostración anterior se conoce usualmente como la *entropía relativa* o la *divergencia de Kullback-Leibler* de  $\mu$  con respecto a  $\pi$  (típicamente con la elección de  $\varphi(x) = -\log(x)$ , aunque en este caso levantamos el requerimiento de que  $\varphi$  sea acotada, y extendemos la definición a 0 mediante considerar  $\lim_{x \rightarrow 0^+} x \log(x) = 0$ ). Esta noción tiene muchas aplicaciones en probabilidades y otras áreas de las matemáticas, así como también a la teoría de la información, la estadística y el aprendizaje de máquinas.

La idea es que  $\mathcal{E}(\mu)$  mide la distancia, o más precisamente el desorden, de  $\mu$  con respecto a  $\pi$  (note por ejemplo que  $\mathcal{E}(\pi) = 0$  para el caso  $\varphi(x) = -\log(x)$ ). Una consecuencia de la demostración anterior es la siguiente: la entropía relativa de cualquier distribución  $\mu$  (no necesariamente invariante) con respecto a  $\pi$  decrece luego de aplicar  $P$ . Esto sugiere que las iteraciones sucesivas de una cadena de Markov acercan su distribución a la distribución de equilibrio  $\pi$ ; esto será el tema de la siguiente sección.

Comenzamos ahora con la demostración del Teorema 1.6.9. El primer paso es la siguiente:

**Proposición 1.6.14.** Sea  $P$  irreducible y recurrente. Sea  $N_n(i) = \sum_{k=0}^n \mathbf{1}_{X_k=i}$  el número de visitas a  $i$  hasta tiempo  $n$ . Entonces

$$\lim_{n \rightarrow \infty} \frac{N_n(i)}{n} = \frac{1}{\mathbb{E}_i(\tau_i)} \quad \text{c.s.}$$

*Demostración.* Sea  $\tau_i^{(k)}$  el tiempo del  $k$ -ésimo retorno a  $i$ , introducido en la Definición 1.5.9. Supondremos que  $X_0 = i$ , el caso general queda de ejercicio. Por propiedad de Markov fuerte, las variables aleatorias  $(\tau_i^{(k+1)} - \tau_i^{(k)})_{k \geq 0}$  son i.i.d., y todas con distribución igual a la de  $\tau_i$ . Luego

$$\frac{\tau_i^{(n)}}{n} = \frac{1}{n} \sum_{k=0}^{n-1} (\tau_i^{(k+1)} - \tau_i^{(k)}) \xrightarrow[n \rightarrow \infty]{\text{c.s.}} \mathbb{E}_i(\tau_i) \quad (1.7)$$

por Ley de los Grandes Números. Pero necesariamente tenemos  $\tau_i^{(N_n(i))} \leq n < \tau_i^{(N_n(i)+1)}$ , y luego

$$\frac{\tau_i^{(N_n(i))}}{N_n(i)} \leq \frac{n}{N_n(i)} < \frac{\tau_i^{(N_n(i)+1)}}{N_n(i)+1} \frac{N_n(i)+1}{N_n(i)}.$$

Como  $X$  es recurrente,  $N_n(i) \rightarrow \infty$ , y luego de (1.7) obtenemos

$$\frac{\tau_i^{(N_n(i))}}{N_n(i)} \longrightarrow \mathbb{E}_i(\tau_i) \quad \text{c.s.},$$

y lo mismo para el lado derecho. Concluimos entonces  $\frac{n}{N_n(i)} \longrightarrow \mathbb{E}_i(\tau_i)$  casi seguramente, lo que nos da lo requerido.

**Ejercicio 1.6.15.** Concluya la demostración, adaptando el argumento anterior al caso  $X_0 \neq i$ .  $\square$

Para el próximo resultado introducimos, para cada  $k \in I$ ,

$$\gamma_i^k = \mathbb{E}_k \left( \sum_{n=0}^{\tau_k-1} \mathbf{1}_{X_n=i} \right) = \sum_{n \geq 0} \mathbb{P}_k(X_n = i, \tau_k > n),$$

el número de visitas a  $i$  antes de retornar a  $k$ .

**Proposición 1.6.16.** Sea  $P$  irreducible y recurrente.

- (a)  $\gamma_k^k = 1$ .
- (b)  $\gamma^k$  es medida invariante (es decir,  $\gamma^k P = \gamma^k$ ).
- (c)  $\gamma_i^k \in (0, \infty)$  para todo  $i$ .

Más aún,  $\mu = \gamma^k$  es la única medida invariante tal que  $\mu_k = 1$ .

*Demostración.* La parte (a) es directa. Para la parte (b) definimos  $\bar{p}_{ki}^{(n)} = \mathbb{P}_k(X_n = i, \tau_k > n)$  y notamos que

$$(\gamma^k P)_\ell = \sum_{j \in I} \gamma_j^k p_{j\ell} = \sum_{j \in I} \sum_{n \geq 0} \bar{p}_{kj}^{(n)} p_{j\ell}.$$

Necesitamos considerar dos casos. Primero, si  $\ell \neq k$ , lo anterior y la propiedad de Markov nos da

$$\begin{aligned} (\gamma^k P)_\ell &= \sum_{j \in I} \sum_{n \geq 0} \mathbb{P}_k(X_n = j, \tau_k > n, X_{n+1} = \ell) = \sum_{n \geq 0} \mathbb{P}_k(\tau_k > n+1, X_{n+1} = \ell) \\ &= \sum_{n \geq 0} \bar{p}_{k\ell}^{(n+1)} = \sum_{n \geq 1} \bar{p}_{k\ell}^{(n)} \end{aligned}$$

(ejercicio: justifique la primera igualdad). Pero  $\bar{p}_{k\ell}^{(0)} = \mathbb{P}_k(X_0 = \ell, \tau_k > 0) = 0$  para  $\ell \neq k$  y por lo tanto la última suma puede extenderse hasta  $n = 0$ , de modo que la suma es igual a  $\gamma_\ell^k$  como queríamos. Supongamos ahora que  $\ell = k$ . Argumentando como antes tenemos

$$(\gamma^k P)_k = \sum_{j \in I} \sum_{n \geq 0} \mathbb{P}_k(X_n = j, \tau_k > n, X_{n+1} = k) = \sum_{n \geq 0} \mathbb{P}_k(\tau_k > n, X_{n+1} = k).$$

El término  $n$ -ésimo del lado derecho es igual a  $\mathbb{P}_k(\tau_k = n+1)$ , y entonces la suma completa es igual a  $\mathbb{P}_k(1 \leq \tau_k < \infty) = 1 = \gamma_k^k$  por (a) y gracias a que  $P$  es recurrente.

Para (iii), como la cadena es irreducible podemos elegir un  $n > 0$  tal que  $p_{ik}^{(n)} > 0$ , y entonces  $1 = \gamma_k^k = (\gamma^k P^n)_k \leq \gamma_i^k p_{ik}^{(n)}$  y deducimos que  $\gamma_i^k > 0$ .

**Ejercicio 1.6.17.** Use un argumento similar para probar que  $\gamma_i^k < \infty$ .

Veamos ahora la unicidad. Supongamos que  $\mu$  es una medida invariante con  $\mu_k = 1$ . Iterando la relación  $\mu P = \mu$  obtenemos, para  $j \neq k$ ,

$$\mu_j = \sum_{i_0 \in I} \mu_{i_0} p_{i_0 j} = \sum_{i_0 \neq k} \mu_{i_0} p_{i_0 j} + p_{kj} = \sum_{i_0, i_1 \neq k} \mu_{i_1} p_{i_1 i_0} p_{i_0 j} + p_{kj} + \sum_{i_0 \neq k} p_{ki_0} p_{i_0 j}$$

e, inductivamente,

$$= \sum_{i_0, \dots, i_n \neq k} \mu_{i_n} p_{i_n i_{n-1}} \cdots p_{i_0 j} + p_{kj} + \sum_{i_0 \neq k} p_{ki_0} p_{i_0 j} + \cdots + \sum_{i_0, \dots, i_{n-1} \neq k} p_{ki_{n-1}} \cdots p_{i_0 j}.$$

Omitiendo el primer término, el lado derecho es (recordando  $j \neq k$ )

$$\geq \mathbb{P}_k(X_1 = j, \tau_k > 1) + \mathbb{P}_k(X_2 = j, \tau_k > 2) + \cdots + \mathbb{P}_k(X_n = j, \tau_k > n) \xrightarrow{n \rightarrow \infty} \gamma_j^k.$$

Deducimos entonces que  $\mu_j \geq \gamma_j^k$  (notar que hasta acá no hemos usado recurrencia, y luego esta desigualdad es cierta en general). Como  $P$  es recurrente,  $\gamma^k$  es medida invariante por (b), y luego la medida  $\lambda = \mu - \gamma^k$  (notar que  $\lambda \geq 0$ ) también lo es. Luego para  $i \in I$  y cualquier  $n > 0$  tenemos  $\lambda_i p_{ik}^{(n)} \leq (\lambda P)_k = \lambda_k = 0$  (por hipótesis y (a)), y entonces usando la irreducibilidad para elegir  $n > 0$  tal que  $p_{ik}^{(n)} > 0$  obtenemos  $\lambda_i = 0$ .  $\square$

*Demostración Teorema 1.6.9.* Veamos primero que (i) implica (iii). Sea  $k$  un estado recurrente positivo. Gracias a la proposición,  $\gamma^k$  es una medida invariante. Basta probar entonces que  $\gamma^k(I) < \infty$ . Tenemos

$$\gamma^k(I) = \sum_{\ell \in I} \gamma_\ell^k = \sum_{\ell \in I} \sum_{n \geq 0} \mathbb{P}_k(X_n = \ell, \tau_k > n) = \sum_{n \geq 0} \mathbb{P}_k(\tau_k > n) = \mathbb{E}_k(\tau_k) < \infty,$$

gracias a la recurrencia positiva de  $k$ .

Que (ii) implica (i) es trivial. Para ver que (iii) implica (ii), sea  $\pi$  una distribución invariante y fijemos  $k \in I$ . Gracias al Teorema 1.6.10,  $\gamma^k = \alpha \pi$  para algún  $\alpha \geq 0$ . Tal como en la parte anterior, tenemos  $\mathbb{E}_k(\tau_k) = \gamma^k(I)$ , y luego  $\mathbb{E}_k(\tau_k) = \alpha \pi(I) = \alpha < \infty$ , y luego  $k$  es recurrente positivo.  $\square$

Antes de terminar esta sección, enunciamos un resultado que puede demostrarse usando los métodos de esta sección ( $\blacktriangle$ ):

**Teorema 1.6.18. (Teorema Ergódico para Cadenas de Markov)** *Sea  $P$  irreducible con distribución invariante  $\pi$ . Sea  $f: I \rightarrow \mathbb{R}$  tal que  $\sum_i |f(i)|\pi(i) < \infty$  (es decir  $f \in L^1(I, \pi)$ ). Entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n f(X_k) = \sum_i f(i)\pi(i).$$

Este tipo de resultados se conocen también como leyes de grandes números para funcionales aditivos (en este caso, de una cadena de Markov recurrente positiva). Bajo hipótesis adicionales sobre  $f$  y  $X$ , es posible también demostrar una versión del Teorema Central del Límite en este contexto.

## A. Repaso sobre esperanzas condicionales

En lo que sigue nos damos un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  y asumimos que  $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ , es decir  $X$  es una variable aleatoria a valores reales en este espacio, tal que  $\mathbb{E}(|X|) < \infty$ .

**Definición A.0.1. (Esperanza condicional)** Sea  $\mathcal{G}$  una sub  $\sigma$ -álgebra de  $\mathcal{F}$ . La *esperanza condicional* de  $X$  dada  $\mathcal{G}$  es la única variable aleatoria  $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  tal que:

Esperanza condicional

- $Z$  es  $\mathcal{G}$ -medible (o sea  $Z \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ ).
- Para todo  $A \in \mathcal{G}$ ,  $\mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(Z \mathbf{1}_A)$ .

A esta variable aleatoria la denotamos por  $Z = \mathbb{E}(X|\mathcal{G})$ .

La existencia de la esperanza condicional es una consecuencia del Teorema de Radon-Nikodym, ver [SM18] o [Dur19, Cap. 5].

¿Qué significa condicionar con respecto a una  $\sigma$ -álgebra  $\mathcal{G}$ ? La idea es que  $\mathcal{G}$  describe información que tenemos a disposición: para cada  $A \in \mathcal{G}$  sabemos si el evento  $A$  ha ocurrido o no (en otras palabras, dado  $\omega \in \Omega$  sabemos si  $\omega \in A$  o  $\omega \notin A$  para cada evento  $A \in \mathcal{G}$ ). En particular, mientras más fina sea la  $\sigma$ -álgebra  $\mathcal{G}$ , más información contiene. Veamos un par de ejemplos que ilustran esto:

**Ejemplo A.0.2.** Supongamos que  $X$  es  $\mathcal{G}$ -medible. Esto significa que  $X^{-1}(U) \in \mathcal{G}$  para todo  $U \in \mathcal{B}(\mathbb{R})$  y por lo tanto, de acuerdo a lo anterior, la información contenida en  $\mathcal{G}$  permitiría determinar si  $X(\omega) \in U$  para cualquiera de estos  $U$ , lo que obviamente permitiría determinar el valor de  $X(\omega)$ . Luego debiéramos esperar que  $\mathbb{E}(X|\mathcal{G}) = X$  (en otras palabras, si ya conocemos  $X$  entonces nuestra mejor aproximación a  $X$  es simplemente  $X$ ). Esto es cierto, y puede chequearse muy fácilmente: la segunda condición de la definición es trivial, y la primera es la hipótesis (ver también la Prop. A.0.6(4)).

**Ejemplo A.0.3.** En el caso anterior  $\mathcal{G}$  contenía información perfecta sobre  $X$ . La situación opuesta es cuando  $\mathcal{G}$  no contiene ninguna información sobre  $X$ , que corresponde al caso en que  $X$  es independiente de  $\mathcal{G}$ , lo que significa que  $\mathbb{P}(\{X \in U\} \cap B) = \mathbb{P}(X \in U)\mathbb{P}(B)$  para cada  $U \in \mathcal{B}(\mathbb{R})$  y  $B \in \mathcal{G}$ . En este caso tenemos  $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$  (en otras palabras, como no tenemos ninguna información sobre  $X$ , nuestra mejor aproximación a  $X$  es su media  $\mathbb{E}(X)$ ).

**Ejercicio A.0.4.** Demuestre el resultado descrito en el ejemplo anterior.

**Ejemplo A.0.5.** Por último consideremos un ejemplo concreto. Tomemos  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,  $\mathcal{F} = \mathcal{P}(\Omega)$ ,  $\mathbb{P}$  la medida de probabilidad uniforme sobre  $\Omega$  y  $X(\omega) = \omega$  (en otras palabras,  $X$  modela el lanzamiento de un dado). Supongamos que  $\mathcal{G} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ . Entonces  $\mathcal{G}$  contiene la respuesta a la pregunta ¿fue el resultado del lanzamiento un número par o impar?. En base a esto, tenemos por ejemplo que  $\mathbb{E}((-1)^X|\mathcal{G})(\omega) = \mathbf{1}_{\omega \text{ es par}} - \mathbf{1}_{\omega \text{ es impar}}$  y  $\mathbb{E}(X|\mathcal{G})(\omega) = 4 \cdot \mathbf{1}_{\omega \text{ es par}} - 3 \cdot \mathbf{1}_{\omega \text{ es impar}}$  (Ejercicio: chequear estas fórmulas usando la Prop. A.0.6).

Enumeraremos a continuación las propiedades básicas de la esperanza condicional (para las demostraciones, ver [SM18; Dur19]):

**Proposición A.0.6.** Sean  $X$  e  $Y$  variables aleatorias integrables y  $\alpha \in \mathbb{R}$ .

- (1)  $\mathbb{E}(1|\mathcal{G}) = 1$ .
- (2) Si  $X \geq 0$  c.s. entonces  $\mathbb{E}(X|\mathcal{G}) \geq 0$  c.s. En particular, si  $X \geq Y$  c.s. entonces  $\mathbb{E}(X|\mathcal{G}) \geq \mathbb{E}(Y|\mathcal{G})$  c.s.

- (3)  $\mathbb{E}(\alpha X + Y|\mathcal{G}) = \alpha\mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Y|\mathcal{G})$ .
- (4) Si  $Y$  es  $\mathcal{G}$ -medible y  $\mathbb{E}(|XY|) < \infty$  entonces  $\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$ . En particular,  $\mathbb{E}(Y|\mathcal{G}) = Y$  y  $\mathbb{E}(X|\mathcal{F}) = X$ .
- (5) Si  $\varphi$  es convexa y  $\mathbb{E}(\varphi(X)) < \infty$ , entonces  $\varphi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\varphi(X)|\mathcal{G})$ . En particular,  $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$ .
- (6) Si  $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$  entonces

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(X|\mathcal{H}).$$

Las propiedades (4) y (6) son clave, y se usan de manera repetida a lo largo de las notas. (6) es consistente con la interpretación de la esperanza condicional como una proyección (la  $\sigma$ -álgebra, o proyección, más chica siempre gana), que puede hacerse rigurosa en el caso de variables en  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  (visto en auxiliar, o ver [Dur19, Teo. 5.1.8]). En términos de la interpretación de  $\sigma$ -álgebras como información, una posible manera de entender la propiedad es la siguiente: si pensamos en  $\mathcal{G}$  como la información que posee un genio y  $\mathcal{H}$  la que posee un necio, entonces el genio no tiene cómo mejorar la estimación del necio, pero el necio siempre puede echar a perder la estimación del genio. Por otro lado, (4) dice que si la variable aleatoria  $Y$  es conocida dada la información contenida en  $\mathcal{G}$ , entonces puede ser considerada como una cantidad determinista, y puede salir como un factor fuera de la esperanza condicional.

La siguiente propiedad se conoce a veces como la versión estocástica de la Ley de Probabilidades Totales (fue vista en clase auxiliar):

**Proposición A.0.7.** Sea  $\Lambda \subseteq \mathbb{N}$  y  $(A_n)_{n \in \Lambda} \subseteq \mathcal{F}$  una partición de  $\Omega$  con  $\mathbb{P}(A_n) > 0$  para cada  $n \in \Lambda$ . Sea  $\mathcal{G} = \sigma(\{A_n\}_{n \in \Lambda})$ . Entonces si  $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ ,

$$\mathbb{E}(X|\mathcal{G})(\omega) = \sum_{n \in \Lambda} \frac{\mathbb{E}(X\mathbf{1}_{A_n})}{\mathbb{P}(A_n)} \mathbf{1}_{A_n}(\omega).$$

Sea  $A \in \mathcal{F}$  y  $\mathcal{G}$  sub  $\sigma$ -álgebra de  $\mathcal{F}$ . Definimos la *probabilidad de  $A$  condicional a  $\mathcal{G}$*  como


$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(\mathbf{1}_A|\mathcal{G}).$$

**Ejercicio A.0.8.** Demuestre que si  $\mathcal{G} = \sigma(\{B, B^c\})$  entonces

$$\mathbb{P}(A|\mathcal{G})(\omega) = \mathbb{P}(A|B)\mathbf{1}_{\omega \in B} + \mathbb{P}(A|B^c)\mathbf{1}_{\omega \in B^c}.$$

~~~~~ Asegúrese de comprender el contenido del ejercicio anterior en función de la interpretación de  $\sigma$ -álgebras como información.

Notación. Si X es una variable aleatoria real como arriba e Y_1, \dots, Y_n es una familia de variables aleatorias en el mismo espacio (no necesariamente reales), escribiremos $\mathbb{E}(X|Y_1, \dots, Y_n)$ para denotar $\mathbb{E}(X|\sigma(\{Y_1, \dots, Y_n\}))$. Lo mismo vale para probabilidades condicionales: $\mathbb{P}(A|Y_1, \dots, Y_n)$ denota $\mathbb{P}(A|\sigma(\{Y_1, \dots, Y_n\}))$.

Observación A.0.9. La noción de esperanza y probabilidad condicional con respecto a variables aleatorias que se ve en cursos básicos de probabilidades es consistente con la que se ha introducido acá, lo que justifica la notación anterior. Para más detalles, ver [Dur19, Cap. 5] .

Referencias

- [Dur19] R. Durrett. *Probability—theory and examples*. Vol. 49. Cambridge Series in Statistical and Probabilistic Mathematics. Fifth edition of [MR1068527]. Cambridge University Press, Cambridge, 2019, págs. xii+419.
- [SM18] J. San Martín. *Teoría de la Medida*. 1.^a ed. Editorial Universitaria, 2018.