
PROCESOS DE MARKOV*

Esta versión: 12 de abril 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*.

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

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

(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.

~~~~> 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.

(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.

Definición 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*:

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

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

Observación 1.2. La propiedad de Markov es equivalente a decir que X_{n+1} y (X_0, \dots, X_n) son independientes condicionales a X_n , es decir, “condicional al presente, el futuro del proceso es independiente del pasado”. Ejercicio: demuestre esto.

Proposición 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 $i_0, \dots, i_{n+1} \in I$,

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

Demostración. Ejercicio. □

Ejemplo 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

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

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.

Definición 1.5. (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 1.6. La marcha aleatoria simple del Ejem. 1.4 es una cadena de Markov homogénea, y su matriz de transición P está dada por

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

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 1.7. (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 1.8. *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 1.9. 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 .

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 1.10. (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 .

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

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

Demostración. Para la primera implicancia escribimos

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

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

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

que es lo que buscábamos.

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

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

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 conversa 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 1.12. (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 .*

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 1.13.** 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. 1.11 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  $X \sim \text{CM}(P)$ .

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 1.14. (Construcción directa de cadenas de Markov)** Sea  $(\Omega, \mathcal{F}, \mathbb{P})$  una 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

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

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 1.15.**

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 1.16.** 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 1.17.** Volvemos a la marcha aleatoria simple del Ejem. 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  $\mathbb{E}_i$  para denotar la esperanza bajo  $\mathbb{P}_i$ .

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  $X \sim \text{CM}(\mu, P)$ .

**Ejercicio 1.18.** 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

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

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  $(P^0)_{ij} = \mathbf{1}_{i=j}$ .

La fórmula (1.3) tiene una interpretación probabilística natural:

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

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 1.19. (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

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

**Teorema 1.20. (Ecuaciones de Chapman-Kolmogorov)** Si  $X \sim \text{CM}(P)$  entonces para todo  $m, n \geq 0$  se tiene

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

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 1.21. (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  $n \geq 0$ .

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

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

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  $n \geq 0$ .

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  $n$ .

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

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

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



**Proposición 1.22.** Sea  $X$  un proceso estocástico (y  $(\mathcal{F}_n)_{n \geq 0}$  su filtración natural asociada). Son equivalentes:

- (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 1.23.** 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.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 conversamos 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 1.24.** Sea  $X$  cadena de Markov (homogénea) y sea  $f: I^{\mathbb{N}} \rightarrow \mathbb{R}$  medible y acotada. Entonces

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

**Observación 1.25.**

- (i) Lo anterior es una equivalencia (ejercicio: demuéstrela).
- (ii) Acá hemos abusado notación nuevamente, ver la Obs. 1.23, 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).

*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 1.26.** 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 1.27. (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.

**Ejemplo 1.28.** 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 1.29.** 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 1.30.**

- (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

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

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.

**Definición 1.31.** Dado un espacio de probabilidad filtrado y un t.d.p.  $\tau$ , definimos

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

**Ejemplo 1.32.** Probar que si  $\tau = n$ , entonces con la definición anterior se cumple  $\mathcal{F}_\tau = \mathcal{F}_n$ .

**Proposición 1.33. (Propiedades de  $\mathcal{F}_\tau$ )** Sean  $\tau, \tau_1, \tau_2$  t.d.p.

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

*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 1.34.**  $Y : \Omega \longrightarrow \mathbb{R}$  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  $Y = \sum_{n \in \mathbb{N} \cup \{\infty\}} Y_n \mathbf{1}_{\tau=n}$ .

### 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 1.35. (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*

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

Esto es, la misma propiedad demostrada en el Teo. 1.24 (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. 1.34 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. 1.24) la igualdad anterior se tiene, lo que termina la demostración.  $\square$

**Observación 1.36.** 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 1.37.** La hipótesis de que  $\tau$  sea t.d.p. es necesaria en el teorema anterior: en general la propiedad no vale para tiempos aleatorios que no sean tiempos de parada. Por ejemplo podemos considerar la marcha aleatoria simple en  $\mathbb{Z}$  del Ej. 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. Como la marcha parte en 0 se tiene que si  $\sigma < \infty$  entonces  $X_\sigma = 0$  (recordando que la marcha solo se mueve con pasos de largo  $\pm 1$ ). Si la propiedad de Markov se tuviera, entonces por ejemplo tendríamos  $\mathbb{E}(X_{\sigma+1} | \mathcal{F}_\sigma) \mathbf{1}_{\sigma < \infty} = \mathbb{E}_{X_\sigma}(X_1) \mathbf{1}_{\sigma < \infty} = \mathbb{E}_0(X_1) \mathbf{1}_{\sigma < \infty} = (p - q) \mathbf{1}_{\sigma < \infty}$ . Pero claramente  $\mathbb{E}(X_{\sigma+1} | \mathcal{F}_\sigma) \mathbf{1}_{\sigma < \infty} = \mathbf{1}_{\sigma < \infty}$ , porque si a tiempo  $\sigma$  es la última vez que estuvimos en  $\mathbb{Z}_{\leq 0}$ , entonces un instante de tiempo después necesariamente estaremos en 1. Como  $p - q < 1$  si  $p < 1$ , vemos que esto no es posible.

### 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. 1.4 partiendo desde 0 y definamos el siguiente tiempo de parada (ya introducido en el Ej. 1.28): 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  $\tau_1 + 1$ ). 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 1.38.** Justifique esto cuidadosamente.

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 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,


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

Usando ahora  $q = 1 - p$ , 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  $a$ .

### 1.3. Probabilidades de absorción

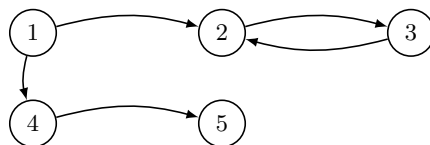
Ver P2 aux. 3 .

### 1.4. Clasificación de estados

#### 1.4.1. Clases e irreducibilidad

**Definición 1.39.** 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  $j \rightarrow i$ .

**Ejemplo 1.40.** 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 1.41.** *Las siguientes son equivalentes:*

- (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

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

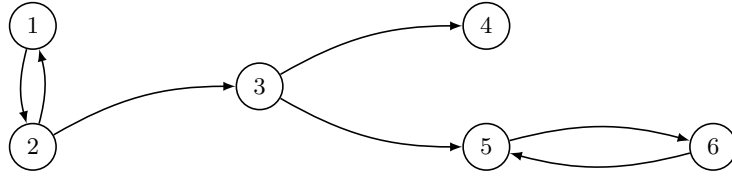
y luego alguno de los sumandos debe ser positivo.  $\square$

**Proposición 1.42.** *La relación  $\rightarrow$  es reflexiva y transitiva y la relación  $\leftrightarrow$  es de equivalencia.*

*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 1.43. (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  $j \in C$ .

**Ejemplo 1.44.**

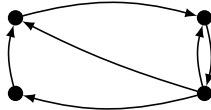


Para esta cadena hay tres clases:  $\{1, 2, 3\}$ ,  $\{4\}$ ,  $\{5, 6\}$ . Las últimas dos son cerradas.

**Definición 1.45. (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.

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

**Ejemplo 1.46.**



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 aperioidicidad. Partimos por la siguiente definición.

**Definición 1.47. (Periodo)** El *periodo* de  $i \in I$  es

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

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

**Ejemplo 1.48.** La marcha aleatoria en  $\mathbb{Z}$  tiene período 2 (ejercicio: verifíquelo). La primera cadena del Ej. 1.46 tiene periodo 1.

**Definición 1.49.** 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.

**Proposición 1.50.** El periodo es propiedad de clase (es decir,  $i \leftrightarrow j$  implica que  $d(i) = d(j)$ ).

*Demostración.* Pospuesta  $\triangleleft$ . □

**Definición 1.51. (Aperioidicidad)** 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.

**Lema 1.52.** 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  $n \geq n_0$ .

*Demostración.* Ejercicio. □



## 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.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:

- $Z$  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.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 debíamos 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.6(4)).

**Ejemplo A.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.4.** Demuestre el resultado descrito en el ejemplo anterior.

**Ejemplo A.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.6).

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

**Proposición A.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.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.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.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.