

**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO**

Facultad de Estudios Superiores Acatlán

Procesos Estocásticos I

Semestre 2026-2

Sonny Medina (Draft)

22 de marzo de 2026

Índice general

1. Modelos Markovianos en Tiempo Discreto	5
1.1. Definición de Proceso Estocástico	5
1.2. Espacio de Estados	6
1.3. Caminata Aleatoria	7
1.4. Distribución Inicial y Matriz de Transición	7
1.5. Cadenas de Markov	12
1.6. Probabilidades de Transición en n Pasos	16
1.7. Comunicación entre Estados	23
1.8. Distribuciones Periódicas: Período de un Estado	25
1.9. Primeras Visitas	27
1.10. Estados Recurrentes y Transitorios	30
1.10.1. Visitas a un estado: el caso general N_{ij}	35
1.11. Tiempo medio de retorno	37
1.11.1. Análisis del primer paso	39
1.12. Teorema ergódico	40
1.13. Recurrencia positiva y nula	44
1.14. Evolución de la distribución de X_n	47

Capítulo 1

Modelos Markovianos en Tiempo Discreto

La teoría de procesos estocásticos tuvo sus orígenes en los trabajos de Andréi Márkov a principios del siglo XX. Márkov, discípulo de Pafnuty Chebyshev al igual que Andréi Kolmogorov, estudió sucesiones de variables aleatorias que satisfacían una propiedad especial: la dependencia entre observaciones sucesivas. Aunque las contribuciones de Márkov no fueron tan formales como las posteriores de Kolmogorov, quien desarrolló la teoría axiomática de la probabilidad, Márkov es considerado el precursor de la teoría moderna de procesos estocásticos. En sus trabajos iniciales, Márkov se enfocó principalmente en procesos a tiempo discreto con un enfoque probabilístico. Con el desarrollo de la teoría, especialmente tras las contribuciones de Kolmogorov, la disciplina se formalizó considerablemente e incorporó tanto procesos a tiempo discreto como continuo, ampliando su alcance hacia aplicaciones contemporáneas en física estadística, finanzas matemáticas, biología, ingeniería y ciencias de la computación. En la actualidad, los procesos estocásticos constituyen una herramienta fundamental para modelar sistemas dinámicos bajo incertidumbre, con aplicaciones que van desde la modelación de mercados financieros y sistemas de colas, hasta algoritmos de aprendizaje automático, métodos de Monte Carlo vía cadenas de Markov y el análisis de datos temporales en contextos científicos e industriales.

1.1. Definición de Proceso Estocástico

Definición 1.1. Un **proceso estocástico** es una colección de variables aleatorias $\{X_t : t \in T\}$ definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde:

- T es el **índice temporal**, que puede ser discreto ($T = \mathbb{N}$) o continuo ($T = \mathbb{R}^+$). Asumiremos que ambos conjuntos contienen al cero.
- X_t representa el estado del sistema en el tiempo t .

Un proceso estocástico es **discreto** si T es un conjunto discreto, y **continuo** si T es un conjunto continuo.

Interpretación como función

Un proceso estocástico puede interpretarse como una función

$$X : T \times \Omega \rightarrow S$$

tal que a cada pareja $(t, \omega) \in T \times \Omega$ se le asocia el estado $X(t, \omega) \in S$. Para cada tiempo $t \in T$ fijo, la función $\omega \mapsto X(t, \omega)$ es una variable aleatoria. Por el contrario, para cada resultado $\omega \in \Omega$ fijo, la función $t \mapsto X(t, \omega)$ se llama una **trayectoria** o **realización** del proceso, y representa la evolución temporal del sistema bajo el escenario particular ω .

EJEMPLOS

1.2. Espacio de Estados

El **espacio de estados** de un proceso estocástico es el conjunto S de todos los valores posibles que las variables aleatorias X_t pueden tomar. Se requiere que S sea equipado con una σ -álgebra apropiada para poder hablar de eventos relacionados con los valores del proceso. Consideraremos principalmente espacios de estados discretos (subconjuntos de $\mathbb{Z}$) o continuos (subconjuntos de $\mathbb{R}$, $\mathbb{R}^d$, etc.). Por ejemplo:

- Si $S = \{0, 1\}$, el proceso es binario.
- Si $S = \mathbb{R}$, el proceso tiene un espacio de estados continuo.
- Si $S = \mathbb{Z}$, el proceso tiene un espacio de estados discreto infinito.
- Si $S = \mathbb{R}^d$ para $d \geq 2$, el espacio de estados es multidimensional.

Ejemplo 1.1 (1). Sucesión de resultados de un dado. Considere un dado justo. El espacio parametral es $T = \mathbb{N}$, ya que observamos los resultados en pasos discretos (cada lanzamiento). El espacio de estados es $S = \{1, 2, 3, 4, 5, 6\}$ y supondremos que s.p.g. $P(X_0 = 6) = 1$. Ejemplos de probabilidades de interés incluyen:

- $\mathbb{P}(X_3 = 6)$: Probabilidad de obtener un 6 en el tercer lanzamiento.
- $\mathbb{P}(\text{al menos un 6 en 5 lanzamientos})$.
- $\mathbb{P}(\text{El primer 6 sale en el lanzamiento } k)$.
- Probabilidades condicionales como $\mathbb{P}(X_4 = 5 \mid X_3 = 2)$.

Un proceso estocástico se llama cadena cuando es un proceso indexado en tiempo discreto, con espacio de estados numerable. Si dibujamos los posibles estados como nodos en un grafo, y las transiciones entre estados como aristas dirigidas obtenemos una representación visual útil para analizar el comportamiento del proceso.

1.3. Caminata Aleatoria

Un ejemplo fundamental en la teoría de procesos estocásticos es la **caminata aleatoria**. Sea $\{Y_i : i \in \mathbb{N}\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.), todas definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, que toman valores en $S_Y \subseteq \mathbb{R}$. Denotamos por μ_Y la distribución común de las variables Y_i . Consideremos la sucesión de sumas parciales:

$$X_n = \sum_{i=1}^n Y_i, \quad n \in \mathbb{N}, \quad \text{con } X_0 = 0.$$

Este proceso se llama **caminata aleatoria** o **proceso de suma acumulada**, y es un proceso a tiempo discreto con espacio de estados $S \subseteq \mathbb{R}$ (típicamente $S = \mathbb{Z}$ o $S = \mathbb{R}$ dependiendo de la distribución de Y_i).

Proposición 1.1. *Para una caminata aleatoria $X_n = \sum_{i=1}^n Y_i$ donde los Y_i son i.i.d.:*

- *La distribución de X_n es la convolución de n copias de μ_Y .*
- *Para $n \geq 1$, se tiene $X_n = X_{n-1} + Y_n$, por lo que la probabilidad de transición depende únicamente de la distribución de Y_i .*

Ejemplo 1.2 (2). **Caminata aleatoria simétrica en $\mathbb{Z}$.** Sea Y_i con distribución $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2}$. El proceso $X_n = \sum_{i=1}^n Y_i$ es una caminata aleatoria simétrica donde en cada paso el proceso se mueve una unidad hacia la derecha o hacia la izquierda con igual probabilidad. Algunas probabilidades de interés:

- $\mathbb{P}(X_1 = 1) = \frac{1}{2}$ y $\mathbb{P}(X_1 = -1) = \frac{1}{2}$.
- $\mathbb{P}(X_2 = 0) = \mathbb{P}(Y_1 = 1, Y_2 = -1) + \mathbb{P}(Y_1 = -1, Y_2 = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.
- $\mathbb{P}(X_2 = 2) = \mathbb{P}(Y_1 = 1, Y_2 = 1) = \frac{1}{4}$.
- $\mathbb{P}(X_3 = 0)$: Es necesario un número par de pasos $+1$ y un número par de pasos -1 . Esto es imposible en 3 pasos, por lo que esta probabilidad es 0.

Ejemplo 1.3 (3). **Caminata aleatoria de Pólya.** Consideremos Y_i con distribución $\mathbb{P}(Y_i = k) = p^k(1-p)^{1-k}$ para $k \in \{0, 1\}$, donde $0 < p < 1$. El proceso $X_n = \sum_{i=1}^n Y_i$ cuenta el número de “éxitos” hasta el tiempo n , y representa el número de caras en n lanzamientos de una moneda sesgada. Este tipo de caminata aleatoria tiene aplicaciones en economía, biología y otras áreas.

1.4. Distribución Inicial y Matriz de Transición

Distribución inicial. La distribución inicial (o distribución de probabilidad inicial) de un proceso estocástico describe la ley de probabilidad del estado del sistema en el tiempo inicial. Para un proceso con espacio de estados S , se define como:

$$\pi = (\pi_i)_{i \in S}, \quad \text{donde } \pi_i = \mathbb{P}(X_0 = i) \geq 0, \quad \sum_{i \in S} \pi_i = 1.$$

En el caso de espacios de estados infinitos, la distribución inicial es una medida de probabilidad sobre S .

Probabilidades de transición. Para procesos a tiempo discreto, la **probabilidad de transición** de un paso describe la probabilidad de pasar de un estado i al estado j entre tiempo n y $n + 1$:

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad \forall i, j \in S.$$

Cuando el espacio de estados es finito, estas probabilidades pueden organizarse en una **matriz de transición**:

$$P = (p_{ij})_{i,j \in S},$$

que es una matriz cuadrada cuyas entradas satisfacen propiedades específicas.

Proposición 1.2. Sea $P = (p_{ij})$ la matriz de probabilidades de transición de un paso para un proceso estocástico con espacio de estados finito $S = \{1, 2, \dots, m\}$. Entonces:

1. $p_{ij} \geq 0$ para todos $i, j \in S$.
2. $\sum_{j=1}^m p_{ij} = 1$ para todo $i \in S$ (cada fila suma 1).
3. P es una **matriz estocástica**.

Ejemplo 1.4 (4). Movilidad social y clases socioeconómicas. Considere una población clasificada en tres clases: baja (B), media (M) y alta (A). Sea X_n la clase socioeconómica del primogénito de una familia en la generación n . El espacio de estados es $S = \{B, M, A\}$ y el espacio parametral es $T = \mathbb{N}$.

Suponga que en el momento actual ($n = 0$), la distribución inicial es:

$$\pi_B = 0.5, \quad \pi_M = 0.3, \quad \pi_A = 0.2.$$

Esto significa que el 50 % de la población pertenece a la clase baja, el 30 % a la clase media y el 20 % a la clase alta. Una pregunta natural es: ¿Cuál es la probabilidad de que el nieto de una persona de clase baja pertenezca a la clase alta? Para responder esto necesitamos conocer las probabilidades de transición $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ para cada par de clases (i, j) .

Ejemplo 1.5 (5). Dinero en la cartera. Sea X_n la cantidad de dinero (en pesos) en nuestra cartera al final del día n . El espacio de estados es $S = \mathbb{R}_0^+$ (números reales no negativos). El espacio parametral es $T = \mathbb{N}$ (días). La distribución inicial X_0 depende de cuánto dinero teníamos al inicio, y las probabilidades de transición modelan cómo cambia nuestro saldo diariamente en función de nuestros gastos e ingresos, que pueden ser aleatorios.

Ejemplo 1.6 (6). Precio de una acción en la bolsa. Sea X_n el precio de una acción particular al cierre del día n . El espacio parametral es $T = \mathbb{N}$ (días de operación bursátil) y el espacio de estados es $S = \mathbb{R}^+$ (precios positivos). La distribución inicial depende del precio de cierre inicial de la acción. Las probabilidades de transición $\mathbb{P}(X_{n+1} \in [a, b] \mid X_n = x)$ modelan cómo evoluciona el precio día a día en respuesta a condiciones de mercado, y son un objeto central de estudio en matemática financiera.

Ejemplo 1.7 (7). Temperatura ambiente cada hora. Sea X_t la temperatura ambiente (en grados Celsius) registrada cada hora t durante un periodo de tiempo. Aquí, $T = \mathbb{N}$ (horas) y $S = \mathbb{R}$ (cualquier temperatura es posible en principio). La distribución inicial X_0 depende de la temperatura al momento inicial, y las probabilidades de transición reflejan patrones de cambio de temperatura entre horas consecutivas, que dependen de factores como la hora del día, la estación del año, y otras condiciones meteorológicas.

Ejemplo 1.8 (8). Deciles de ingresos. Sea X_n el decil de ingresos al que pertenece el primogénito de una familia en la generación n , donde los deciles son $S = \{D_1, D_2, \dots, D_{10}\}$ con D_1 siendo el 10 % más pobre y D_{10} el 10 % más rico. El espacio parametral es $T = \mathbb{N}$ (generaciones). La distribución inicial $(\pi_{D_1}, \dots, \pi_{D_{10}})$ representa la proporción de la población en cada decil. Las probabilidades de transición $p_{ij} = \mathbb{P}(X_{n+1} = D_j \mid X_n = D_i)$ describen la movilidad socioeconómica entre generaciones.

Ejemplo 1.9 (9). Estatura de una persona en el tiempo. Sea X_n la estatura de una persona (en centímetros) medida al final del año n . Aquí, $T = \mathbb{N}$ (años de vida) y $S = \mathbb{R}^+$ (estaturas positivas). La distribución inicial depende de la estatura al nacimiento. Las probabilidades de transición son determinísticas en el caso del crecimiento infantil (la estatura tiende a aumentar de manera predecible años a años durante la infancia), pero pueden incluir variabilidad en períodos posteriores.

Ejemplo 1.10 (10). Población de una especie. Sea X_n la cantidad total de individuos de una especie particular en una región al tiempo n . El espacio de estados es $S = \mathbb{Z}_{\geq 0}$ (números enteros no negativos), el espacio parametral es $T = \mathbb{N}$ (periodos de tiempo, p. ej., años), y la distribución inicial es el tamaño de la población inicial. Las probabilidades de transición dependen de tasas de natalidad, mortalidad, inmigración y emigración, que pueden ser determinísticas o estocásticas.

Ejemplo 1.11 (11). Movimiento de una partícula en un fluido. Sea $\mathbf{X}_t = (X_t, Y_t, Z_t) \in \mathbb{R}^3$ la posición de una partícula suspendida en un fluido al tiempo t (movimiento browniano). El espacio de estados es $S = \mathbb{R}^3$. Esta es una descripción clásica del movimiento aleatorio de partículas coloidales, estudiado por primera vez por Robert Brown. El espacio parametral puede ser discreto (si observamos la posición en intervalos regulares) o continuo (si consideramos la trayectoria continua de la partícula).

Observación 1.1. La distribución inicial y las probabilidades de transición determinan completamente la evolución probabilística de un proceso estocástico a tiempo discreto. Para un proceso con espacio de estados finito, la matriz de transición es una herramienta algebraica fundamental para estudiar el comportamiento a largo plazo del proceso.

La ruina del jugador

Consideremos uno de los problemas más antiguos e importantes de la teoría de procesos estocásticos: *el problema de la ruina del jugador*. Este problema ilustra cómo las probabilidades de transición pueden usarse para resolver preguntas fundamentales sobre la evolución de un proceso.

Planteamiento del problema. Un jugador inicia con una fortuna de k dólares. En cada ronda de juego:

- Con probabilidad p , el jugador gana \$1 (su fortuna aumenta a $k + 1$).
- Con probabilidad $q = 1 - p$, el jugador pierde \$1 (su fortuna disminuye a $k - 1$).

El juego continúa hasta que el jugador se arruina (fortuna = \$0) o alcanza una fortuna objetivo de \$N. Sea X_n la fortuna del jugador después de n rondas, donde $X_0 = k$. El espacio de estados es $S = \{0, 1, 2, \dots, N\}$, siendo 0 y N *estados absorbentes* (el juego termina cuando se alcanzan).

Las probabilidades de transición de un paso son:

$$p_{i,i+1} = p, \quad p_{i,i-1} = q, \quad \text{para } i \in \{1, 2, \dots, N-1\},$$

y $p_{0,0} = 1, p_{N,N} = 1$ (los estados absorbentes permanecen absorbentes).

Objetivo. Calcular $P_k := \mathbb{P}(\text{el jugador alcanza } \$N \mid X_0 = k)$, es decir, la probabilidad de que el jugador gane alcanzando la fortuna objetivo antes de arruinarse.

Solución mediante ecuaciones de recurrencia.

Razonemos de manera intuitiva. Si el jugador se encuentra en la fortuna k (con $0 < k < N$), en la siguiente ronda ocurre uno de dos eventos:

1. Con probabilidad p , el jugador gana \$1 y su fortuna pasa a $k + 1$.
2. Con probabilidad q , el jugador pierde \$1 y su fortuna pasa a $k - 1$.

Por lo tanto, la probabilidad de que el jugador eventualmente alcance \$N partiendo de k depende de lo que suceda en la próxima ronda. Podemos escribir:

$$P_k = p \cdot P_{k+1} + q \cdot P_{k-1}, \quad \text{para } 0 < k < N.$$

Con condiciones de frontera:

$$P_0 = 0 \quad (\text{si el jugador está arruinado, no puede ganar}),$$

$$P_N = 1 \quad (\text{si el jugador alcanza la fortuna objetivo, ha ganado}).$$

Esta es una ecuación de recurrencia lineal homogénea de segundo orden. Su solución depende del valor de p .

Caso 1: Juego justo ($p = q = \frac{1}{2}$).

Si $p = \frac{1}{2}$, la ecuación se convierte en:

$$P_k = \frac{1}{2}P_{k+1} + \frac{1}{2}P_{k-1},$$

que puede reescribirse como:

$$P_{k+1} - P_k = P_k - P_{k-1}.$$

Esto significa que las diferencias consecutivas son constantes. Por lo tanto, P_k es una función lineal de k :

$$P_k = a + bk$$

para algunas constantes a y b . Aplicando las condiciones de frontera:

$$P_0 = a = 0 \implies a = 0,$$

$$P_N = b \cdot N = 1 \implies b = \frac{1}{N}.$$

Por lo tanto:

$$P_k = \frac{k}{N} \quad (\text{para } p = \frac{1}{2}).$$

Esta es una conclusión importante: *en un juego justo, la probabilidad de ganar es proporcional a la fortuna inicial.*

Caso 2: Juego sesgado ($p \neq \frac{1}{2}$).

Cuando $p \neq q$, buscamos soluciones de la forma $P_k = r^k$. Sustituyendo en la ecuación de recurrencia:

$$r^k = p \cdot r^{k+1} + q \cdot r^{k-1}.$$

Dividiendo por r^{k-1} :

$$r = p \cdot r^2 + q,$$

o equivalentemente:

$$pr^2 - r + q = 0.$$

Las soluciones son:

$$r = \frac{1 \pm \sqrt{1 - 4pq}}{2p} = \frac{1 \pm \sqrt{(1 - 2p)^2}}{2p} = \frac{1 \pm |1 - 2p|}{2p}.$$

Como $p + q = 1$, tenemos $1 - 4pq = (p - q)^2 = (1 - 2p)^2$, por lo que:

$$r_1 = 1 \quad \text{y} \quad r_2 = \frac{q}{p}.$$

La solución general es:

$$P_k = A + B \left(\frac{q}{p} \right)^k$$

para constantes A y B . Aplicando las condiciones de frontera:

$$P_0 = A + B = 0 \implies B = -A,$$

$$P_N = A + B \left(\frac{q}{p} \right)^N = A \left(1 - \left(\frac{q}{p} \right)^N \right) = 1.$$

Despejando A :

$$A = \frac{1}{1 - (q/p)^N}.$$

Por lo tanto:

$$P_k = \frac{1 - (q/p)^k}{1 - (q/p)^N} \quad (\text{para } p \neq \frac{1}{2}).$$

Nótese que esta fórmula recupera el resultado del caso justo cuando $p = \frac{1}{2}$ (tomando el límite adecuado).

Interpretación de los resultados.

1. **Cuando $p > \frac{1}{2}$ (juego favorable).** Tenemos $q/p < 1$, por lo que $(q/p)^k \rightarrow 0$ cuando $k \rightarrow \infty$. Para un jugador con fortuna inicial k pequeña comparada con N , la probabilidad $P_k \approx 1 - (q/p)^k$ es cercana a 1. Esto indica que un jugador con ventaja tiene una alta probabilidad de ganar, especialmente si su objetivo es modesto.
2. **Cuando $p < \frac{1}{2}$ (juego desfavorable).** Tenemos $q/p > 1$, por lo que $(q/p)^k$ crece exponencialmente con k . Incluso si el jugador comienza con una buena fortuna inicial, la probabilidad de ganar disminuye exponencialmente. Para un jugador con $p < \frac{1}{2}$ y en contra de un oponente con una fortuna mucho mayor (N muy grande), la ruina es prácticamente segura.
3. **El caso justo.** Cuando $p = \frac{1}{2}$, la probabilidad de ganar es simplemente k/N , una relación lineal. Un jugador con fortuna k y objetivo N tiene una probabilidad de éxito igual a la proporción de su fortuna inicial respecto al capital total en juego.

Observación fundamental. Este problema demuestra la potencia de analizar procesos estocásticos a través de sus probabilidades de transición. Sin invocar teorías sofisticadas, simplemente planteando ecuaciones de recurrencia basadas en cómo el proceso evoluciona en un paso, hemos obtenido resultados exactos y muy interpretables. Esta es la esencia del enfoque mediante cadenas de Markov: comprender el comportamiento global del sistema a partir del conocimiento local de cómo transiciona entre estados adyacentes.

1.5. Cadenas de Markov

Un **proceso de Markov** es un proceso estocástico que satisface una propiedad especial de “falta de memoria”: el futuro del proceso, dado el estado presente, es independiente del pasado. Más formalmente:

Definición 1.2. Decimos que un proceso estocástico $\{X_n : n \in \mathbb{N}\}$ es un **proceso de Markov** si para todos $n \in \mathbb{N}$ y todos los estados $i_0, i_1, \dots, i_n, j \in S$:

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

siempre que las probabilidades condicionales estén bien definidas.

Una **cadena de Markov** es un proceso de Markov a tiempo discreto con espacio de estados discreto.

Definición 1.3. Una **cadena de Markov homogénea** es una cadena de Markov donde las probabilidades de transición no dependen del tiempo n , es decir, $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ es la misma para todo $n \geq 0$.

Observación 1.2. A menos que se indique lo contrario, en este curso supondremos que todas nuestras cadenas de Markov son homogéneas. Esta hipótesis simplifica significativamente el análisis matemático y es apropiada para muchas aplicaciones prácticas.

Observación 1.3. La distribución de una cadena de Markov homogénea depende únicamente de la distribución inicial y de las probabilidades de transición, es decir, de la pareja (π, P) .

Demostremos que la caminata aleatoria $\{X_n : n \in \mathbb{N}\}$ definida por $X_n = \sum_{i=1}^n Y_i$ (donde los Y_i son i.i.d. con distribución en $S_Y \subseteq \mathbb{R}$) es una cadena de Markov homogénea.

Proposición 1.3. *La caminata aleatoria $X_n = \sum_{i=1}^n Y_i$ con $X_0 = 0$ es una cadena de Markov homogénea cuando los Y_i son independientes e idénticamente distribuidas.*

Demostración. Demostramos dos cosas: primero que es una cadena de Markov, y luego que es homogénea.

Parte 1: Propiedad de Markov.

Sea $n \in \mathbb{N}$ y sean $i_0, i_1, \dots, i_n, j$ valores específicos en el espacio de estados. Queremos mostrar que:

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

Por definición, $X_{n+1} = X_n + Y_{n+1}$. Dado que $X_n = i_n$, tenemos:

$$X_{n+1} = j \iff Y_{n+1} = j - i_n.$$

Por lo tanto:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(Y_{n+1} = j - i_n \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0).$$

Ahora bien, la variable aleatoria Y_{n+1} es independiente de todas las variables $Y_1, Y_2, \dots, Y_n$ por la hipótesis de independencia. Puesto que $X_0, X_1, \dots, X_n$ son funciones únicamente de $Y_1, Y_2, \dots, Y_n$, podemos concluir que Y_{n+1} es independiente de $(X_0, X_1, \dots, X_n)$. En particular:

$$\mathbb{P}(Y_{n+1} = j - i_n \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(Y_{n+1} = j - i_n).$$

Asimismo:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i_n) = \mathbb{P}(Y_{n+1} = j - i_n \mid X_n = i_n) = \mathbb{P}(Y_{n+1} = j - i_n).$$

Por lo tanto, se cumple la igualdad requerida, y $\{X_n\}$ es una cadena de Markov.

Parte 2: Homogeneidad.

Para demostrar que la cadena es homogénea, debemos verificar que la probabilidad de transición $p_{ij}^{[n]} := \mathbb{P}(X_{n+1} = j \mid X_n = i)$ no depende de n .

Para cualquier n , condicionando en $X_n = i$:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_n + Y_{n+1} = j \mid X_n = i) = \mathbb{P}(Y_{n+1} = j - i \mid X_n = i).$$

Como Y_{n+1} es independiente de X_n :

$$\mathbb{P}(Y_{n+1} = j - i \mid X_n = i) = \mathbb{P}(Y_{n+1} = j - i).$$

Dado que todos los Y_i tienen la misma distribución, la probabilidad $\mathbb{P}(Y_{n+1} = j - i)$ es idéntica para cualquier n . Por lo tanto:

$$p_{ij} := \mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(Y_1 = j - i)$$

no depende de n , lo que demuestra que la cadena es homogénea. □

Observación 1.4. Si considerásemos un proceso estocástico $\{Z_n : n \in \mathbb{N}\}$ donde cada Z_n es una variable aleatoria independiente, entonces el proceso sería trivialmente una cadena de Markov homogénea: la probabilidad de transición sería:

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i) = \mathbb{P}(Z_{n+1} = j),$$

que no depende del estado actual i . De hecho, para procesos con variables completamente independientes, la propiedad de Markov es automática porque el futuro es independiente de todo el pasado, incluyendo el presente. La especificidad de las cadenas de Markov reside en permitir dependencia entre variables aleatorias sucesivas (a través de las probabilidades de transición) mientras se mantiene la ausencia de memoria más allá del estado presente.

Ejemplos de cadenas de Markov con matriz de transición explícita

Ejemplo 1.12 (12). *Cadena de Markov con matriz de transición finita.* Consideremos una cadena de Markov con espacio de estados $S = \{1, 2, 3\}$ y matriz de transición:

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{pmatrix},$$

donde P_{ij} representa la probabilidad $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$. Por ejemplo:

- $p_{1,1} = 0.5$: Si el proceso está en estado 1, permanece en estado 1 con probabilidad 0.5.
- $p_{1,2} = 0.3$: Si el proceso está en estado 1, transiciona a estado 2 con probabilidad 0.3.
- $p_{2,3} = 0.3$: Si el proceso está en estado 2, transiciona a estado 3 con probabilidad 0.3.

Cada fila de la matriz suma 1, como corresponde a una matriz estocástica. Supongamos una distribución inicial $\pi = (0.4, 0.3, 0.3)$, lo que significa:

$$\mathbb{P}(X_0 = 1) = 0.4, \quad \mathbb{P}(X_0 = 2) = 0.3, \quad \mathbb{P}(X_0 = 3) = 0.3.$$

A partir de esta distribución inicial y la matriz de transición, podemos calcular la distribución de probabilidades en tiempos posteriores. Por ejemplo, la distribución en tiempo 1 es:

$$\pi^{(1)} = \pi \cdot P = (0.4, 0.3, 0.3) \cdot P.$$

Calculando:

$$\pi_1^{(1)} = 0.4 \cdot 0.5 + 0.3 \cdot 0.1 + 0.3 \cdot 0.2 = 0.2 + 0.03 + 0.06 = 0.29,$$

$$\pi_2^{(1)} = 0.4 \cdot 0.3 + 0.3 \cdot 0.6 + 0.3 \cdot 0.2 = 0.12 + 0.18 + 0.06 = 0.36,$$

$$\pi_3^{(1)} = 0.4 \cdot 0.2 + 0.3 \cdot 0.3 + 0.3 \cdot 0.6 = 0.08 + 0.09 + 0.18 = 0.35.$$

Por lo tanto, $\pi^{(1)} = (0.29, 0.36, 0.35)$.

Ejemplo 1.13 (13). *La Urna de Ehrenfest es una cadena de Markov homogénea que representa una descripción matemática simplificada del proceso de difusión de gases o líquidos a través de una membrana. El modelo consiste de dos cajas A y B que contienen un total de N bolas. Seleccionamos al azar una de las N bolas y la colocamos en la otra caja. Sea X_n el número de bolas en la caja A después de n pasos. El espacio de estados es $S = \{0, 1, 2, \dots, N\}$, y las probabilidades de transición son:*

$$p_{i,i+1} = \frac{N-i}{N} \quad (\text{si seleccionamos una bola de la caja B}),$$

$$p_{i,i-1} = \frac{i}{N} \quad (\text{si seleccionamos una bola de la caja A}),$$

y $p_{ij} = 0$ para cualquier otro par (i, j) .

Ejemplo 1.14 (14). ***La cadena de las horas del reloj.** Sea X_n la hora que marca un reloj después de n horas, donde las horas se representan como $S = \{0, 1, 2, \dots, 11\}$ (con 0 representando las 12 en punto). El proceso transiciona según la regla determinística:*

$$X_{n+1} = (X_n + 1) \text{ mód } 12.$$

Esto significa que de cualquier hora i , el reloj avanza a la hora $i + 1$ (módulo 12) con probabilidad 1. Formalmente, las probabilidades de transición son:

$$p_{i,(i+1) \text{ mód } 12} = 1 \quad \text{para todo } i \in S,$$

y $p_{ij} = 0$ para cualquier otro par (i, j) .

La matriz de transición es:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

donde cada fila contiene exactamente un 1 (en la posición correspondiente a la siguiente hora) y ceros en todas partes.

Este es un ejemplo de cadena de Markov con estructura cíclica. Si el reloj comienza en las 3 en punto ($X_0 = 3$), después de 1 hora marca las 4 ($X_1 = 4$), después de 2 horas marca las 5 ($X_2 = 5$), y después de 12 horas regresa a las 3 ($X_{12} = 3$). De manera general, $X_n = (X_0 + n) \text{ mód } 12$.

Aunque el proceso es determinístico (no hay aleatoriedad en las transiciones), sigue siendo una cadena de Markov válida: el futuro depende únicamente del estado presente, no del pasado.

Para aumentar nuestro entendimiento de la propiedad de Markov, probaremos la siguiente proposición, que ayuda en la interpretación del futuro independiente del pasado dado el presente.

Proposición 1.4. *Sea X una cadena de Markov homogénea (π, P) , con espacio de estados S , $i, j, k \in S$ y $0 \leq m \leq n - 1$. Se tiene que*

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_m = k) = \mathbb{P}(X_{n+1} = j | X_n = i).$$

1.6. Probabilidades de Transición en n Pasos

Hasta ahora hemos estudiado las probabilidades de transición de un paso, definidas como $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$. Sin embargo, en muchas aplicaciones es fundamental conocer la probabilidad de que la cadena pase del estado i al estado j transcurridos exactamente n pasos, independientemente del instante inicial.

Definición 1.4. Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov homogénea con espacio de estados S . Definimos la **probabilidad de transición en n pasos** como:

$$p_{ij}^{(n)} = \mathbb{P}(X_{m+n} = j \mid X_m = i),$$

para cualesquiera $i, j \in S$ y $m, n \in \mathbb{N}$. La homogeneidad de la cadena garantiza que esta probabilidad es independiente del instante inicial m .

Observación 1.5. Por supuesto, para $n = 1$, recuperamos las probabilidades de transición de un paso:

$$p_{ij}^{(1)} = p_{ij}.$$

Asimismo, para $n = 0$, se define:

$$p_{ij}^{(0)} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Esta es la matriz identidad, lo que tiene sentido: sin dar ningún paso, la cadena permanece en su estado actual con certeza.

Cuando el espacio de estados es finito, estas probabilidades pueden organizarse en una **matriz de transición en n pasos**:

$$P^{(n)} = (p_{ij}^{(n)})_{i,j \in S},$$

donde $P^{(1)} = P$ es la matriz de transición estándar y $P^{(0)} = I$ es la matriz identidad.

Ejemplo 1.15 (15). Consideremos la cadena de Markov del Ejemplo 12 con matriz de transición:

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}.$$

¿Cuál es la probabilidad de que la cadena transite del estado 1 al estado 2 en exactamente 2 pasos? Esta es la cantidad $p_{1,2}^{(2)}$. Para calcularla, podemos identificar todos los caminos de longitud 2 que comienzan en el estado 1 y terminan en el estado 2:

- Camino $1 \rightarrow 1 \rightarrow 2$: Probabilidad $p_{1,1} \cdot p_{1,2} = 0.5 \times 0.3 = 0.15$.
- Camino $1 \rightarrow 2 \rightarrow 2$: Probabilidad $p_{1,2} \cdot p_{2,2} = 0.3 \times 0.6 = 0.18$.
- Camino $1 \rightarrow 3 \rightarrow 2$: Probabilidad $p_{1,3} \cdot p_{3,2} = 0.2 \times 0.2 = 0.04$.

Por lo tanto:

$$p_{1,2}^{(2)} = 0.15 + 0.18 + 0.04 = 0.37.$$

Alternativamente, podemos calcular $P^{(2)} = P \cdot P = P^2$:

$$P^2 = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{pmatrix} \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}.$$

El elemento $(1, 2)$ de P^2 es:

$$(P^2)_{1,2} = 0.5 \times 0.3 + 0.3 \times 0.6 + 0.2 \times 0.2 = 0.15 + 0.18 + 0.04 = 0.37,$$

que coincide con nuestro cálculo anterior.

La Ecuación de Chapman-Kolmogorov

La ecuación de Chapman-Kolmogorov es una relación fundamental que conecta probabilidades de transición en diferentes números de pasos. Esta ecuación expresa una intuición natural: para ir del estado i al estado j en $n + m$ pasos, necesariamente pasamos por algún estado intermedio k después de los primeros n pasos.

Teorema 1.1 (Ecuación de Chapman-Kolmogorov). *Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov homogénea con espacio de estados S . Entonces, para cualesquiera enteros $n, m \geq 0$ y estados $i, j \in S$:*

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

En notación matricial:

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)},$$

donde el producto es el producto usual de matrices.

Demostración. Demostraremos la ecuación utilizando la ley de probabilidad total y la propiedad de Markov.

Sean $n, m \geq 0$ y consideremos los estados $i, j \in S$. Por definición:

$$p_{ij}^{(n+m)} = \mathbb{P}(X_{n+m} = j \mid X_0 = i).$$

Condicionamos sobre el estado de la cadena después de n pasos, X_n . Por la ley de probabilidad total:

$$\mathbb{P}(X_{n+m} = j \mid X_0 = i) = \sum_{k \in S} \mathbb{P}(X_{n+m} = j \mid X_n = k, X_0 = i) \cdot \mathbb{P}(X_n = k \mid X_0 = i).$$

Ahora aplicamos la propiedad de Markov (Definición de cadena de Markov). Dado que la cadena es Markoviana y el estamos en el estado $X_n = k$, el futuro (desde el tiempo n en adelante) es independiente del pasado (antes del tiempo n). Específicamente:

$$\mathbb{P}(X_{n+m} = j \mid X_n = k, X_0 = i) = \mathbb{P}(X_{n+m} = j \mid X_n = k).$$

Por lo tanto:

$$p_{ij}^{(n+m)} = \sum_{k \in S} \mathbb{P}(X_{n+m} = j \mid X_n = k) \cdot \mathbb{P}(X_n = k \mid X_0 = i).$$

Reconocemos en esta expresión:

$$\mathbb{P}(X_{n+m} = j \mid X_n = k) = \mathbb{P}(X_m = j \mid X_0 = k) = p_{kj}^{(m)},$$

y

$$\mathbb{P}(X_n = k \mid X_0 = i) = p_{ik}^{(n)},$$

por la homogeneidad de la cadena. Por lo tanto:

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} \cdot p_{kj}^{(m)},$$

que es exactamente la ecuación de Chapman-Kolmogorov.

En forma matricial, esta ecuación correspond a la multiplicación de matrices, es decir, $P^{(n+m)} = P^{(n)} \cdot P^{(m)}$. En particular, la siguiente desigualdad será utilizada más adelante: para cualesquiera estados i, j, k y para $0 \leq r \leq n$, se tiene que:

$$p_{ij}^{(n)} \geq p_{ik}^{(r)} p_{kj}^{(n-r)}.$$

□

Observación 1.6. La ecuación de Chapman-Kolmogorov es fundamental porque implica que:

$$P^{(n)} = \underbrace{P \times P \times \cdots \times P}_{n \text{ veces}} = P^n,$$

es decir, la matriz de transición en n pasos es simplemente la n -ésima potencia de la matriz de transición de un paso P . Esta es una consecuencia inmediata de la ecuación de Chapman-Kolmogorov aplicada recursivamente:

$$P^{(n)} = P^{(n-1+1)} = P^{(n-1)} \cdot P^{(1)} = P^{(n-1)} \cdot P.$$

Por inducción, obtenemos $P^{(n)} = P^n$.

Ejemplo 1.16 (16). Retomemos la cadena de Markov del Ejemplo 12. Queremos calcular $p_{1,2}^{(3)}$, la probabilidad de ir del estado 1 al estado 2 en exactamente 3 pasos.

Por la ecuación de Chapman-Kolmogorov con $n = 2$ y $m = 1$:

$$p_{1,2}^{(3)} = \sum_{k=1}^3 p_{1,k}^{(2)} \cdot p_{k,2}^{(1)}.$$

Primero necesitamos $P^{(2)} = P^2$. Ya calculamos este en el Ejemplo 15:

$$P^2 = \begin{pmatrix} 0.29 + 0.11 & 0.36 & 0.35 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \quad (\text{mostramos solo la fila relevante con detalle}).$$

Calculemos P^2 completamente. El elemento (i, j) de P^2 es $\sum_k p_{ik}p_{kj}$. Después del cálculo:

$$\begin{aligned} P^2 &= \begin{pmatrix} 0.5 \cdot 0.5 + 0.3 \cdot 0.1 + 0.2 \cdot 0.2 & 0.5 \cdot 0.3 + 0.3 \cdot 0.6 + 0.2 \cdot 0.2 & 0.5 \cdot 0.2 + 0.3 \cdot 0.3 + 0.2 \cdot 0.6 \\ 0.1 \cdot 0.5 + 0.6 \cdot 0.1 + 0.3 \cdot 0.2 & 0.1 \cdot 0.3 + 0.6 \cdot 0.6 + 0.3 \cdot 0.2 & 0.1 \cdot 0.2 + 0.6 \cdot 0.3 + 0.3 \cdot 0.6 \\ 0.2 \cdot 0.5 + 0.2 \cdot 0.1 + 0.6 \cdot 0.2 & 0.2 \cdot 0.3 + 0.2 \cdot 0.6 + 0.6 \cdot 0.2 & 0.2 \cdot 0.2 + 0.2 \cdot 0.3 + 0.6 \cdot 0.6 \end{pmatrix} \\ &= \begin{pmatrix} 0.25 + 0.03 + 0.04 & 0.15 + 0.18 + 0.04 & 0.10 + 0.09 + 0.12 \\ 0.05 + 0.06 + 0.06 & 0.03 + 0.36 + 0.06 & 0.02 + 0.18 + 0.18 \\ 0.10 + 0.02 + 0.12 & 0.06 + 0.12 + 0.12 & 0.04 + 0.06 + 0.36 \end{pmatrix} \\ &= \begin{pmatrix} 0.32 & 0.37 & 0.31 \\ 0.17 & 0.45 & 0.38 \\ 0.24 & 0.30 & 0.46 \end{pmatrix}. \end{aligned}$$

Ahora aplicamos Chapman-Kolmogorov:

$$\begin{aligned} p_{1,2}^{(3)} &= p_{1,1}^{(2)} \cdot p_{1,2} + p_{1,2}^{(2)} \cdot p_{2,2} + p_{1,3}^{(2)} \cdot p_{3,2} \\ &= 0.32 \times 0.3 + 0.37 \times 0.6 + 0.31 \times 0.2 = 0.096 + 0.222 + 0.062 = 0.380. \end{aligned}$$

Alternativamente, podemos calcular $P^3 = P^2 \times P$ y leer el elemento $(1, 2)$.

Observación 1.7. La ecuación de Chapman-Kolmogorov no es únicamente un resultado teórico interesante; tiene profundas implicaciones prácticas. Al expresar $P^{(n)} = P^n$, nos proporciona un algoritmo eficiente (usando exponenciación rápida de matrices) para calcular probabilidades de transición en un gran número de pasos sin necesidad de enumerar todos los caminos posibles. Esto es especialmente relevante en aplicaciones computacionales donde frecuentemente nos interesa el comportamiento a largo plazo de cadenas de Markov.

Diagonalización de la Matriz de Transición

Para calcular P^n de manera eficiente, especialmente cuando n es grande, podemos utilizar la diagonalización de la matriz P . El siguiente teorema es fundamental en el álgebra lineal y tiene aplicaciones directas al estudio de cadenas de Markov.

Teorema 1.2. Sea P una matriz cuadrada $m \times m$ con espectro (conjunto de valores propios) $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$, donde λ_i son valores propios (no necesariamente distintos) y sean $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ los correspondientes vectores propios por la derecha.¹

Si P es diagonalizable², entonces existe una matriz V cuyas columnas son los vectores propios $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$, y una matriz $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ diagonal de valores propios, tales que:

$$P = VDV^{-1}.$$

En consecuencia, para el cálculo de potencias:

$$P^n = VD^nV^{-1},$$

donde $D^n = \text{diag}(\lambda_1^n, \lambda_2^n, \dots, \lambda_m^n)$ es trivial de calcular.

¹Un vector propio derecho satisface $P\mathbf{v}_i = \lambda_i\mathbf{v}_i$.

²Una condición suficiente para que una matriz sea diagonalizable es que todos sus valores propios sean distintos.

Observación 1.8. Para una matriz cuadrada P (en particular, una matriz de transición), existen tres propiedades fundamentales relacionadas con los valores propios:

1. Si la matriz es estocástica, tiene al menos un valor propio igual a 1, lo cual está relacionado con la existencia de distribuciones estacionaria (por verse en las secciones siguientes).
2. La **traza** de P (suma de los elementos diagonales) es igual a la suma de los valores propios:

$$\text{tr}(P) = \sum_{i=1}^m p_{ii} = \sum_{i=1}^m \lambda_i.$$

3. El **determinante** de P es el producto de los valores propios:

$$\det(P) = \prod_{i=1}^m \lambda_i.$$

En los casos 2×2 o 3×3 , es posible obtener los valores propios resolviendo el sistema de ecuaciones que resulta de estas tres propiedades, sin necesidad de calcular explícitamente el polinomio característico.

Ejemplo 1.17 (17). *Cadena de Markov con dos estados: caso general.* Consideremos una cadena de Markov con espacio de estados $S = \{1, 2\}$ y matriz de transición:

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix},$$

donde $0 < \alpha, \beta < 1$. Esta es una cadena arbitraria simétrica o asimétrica, dependiendo de los valores de α y β .

Paso 1: Calcular los valores propios.

Para esto podemos encontrar las raíces del polinomio característico $\det(P - \lambda I) = 0$ o hacer caso a las observaciones anteriores. Al ser el determinante es $\det(P) = (1 - \alpha)(1 - \beta) - \alpha\beta = 1 - \alpha - \beta$ y 1 un valor propio, el otro tiene que necesariamente ser $1 - \alpha - \beta$.

Paso 2: Calcular los vectores propios.

Para $\lambda_1 = 1$:

$$\begin{pmatrix} 1 - \alpha - 1 & \alpha \\ \beta & 1 - \beta - 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

De la primera ecuación: $-\alpha v_1 + \alpha v_2 = 0 \implies v_1 = v_2$. Un vector propio es $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Para $\lambda_2 = 1 - \alpha - \beta$:

$$\begin{pmatrix} 1 - \alpha - (1 - \alpha - \beta) & \alpha \\ \beta & 1 - \beta - (1 - \alpha - \beta) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

De la primera ecuación: $\beta v_1 + \alpha v_2 = 0 \implies v_2 = -\frac{\beta}{\alpha} v_1$. Un vector propio es $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -\beta/\alpha \end{pmatrix}$ o, reescalando, $\mathbf{v}_2 = \begin{pmatrix} \alpha \\ -\beta \end{pmatrix}$.

Paso 3: Formar la matriz V y su inversa.

$$V = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}.$$

El determinante es $\det(V) = 1 \cdot (-\beta) - \alpha \cdot 1 = -\beta - \alpha = -(\alpha + \beta)$.

$$V^{-1} = \frac{1}{-(\alpha + \beta)} \begin{pmatrix} -\beta & -\alpha \\ -1 & 1 \end{pmatrix} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}.$$

Paso 4: Expresión para P^n .

$$P^n = V D^n V^{-1} = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1 - \alpha - \beta)^n \end{pmatrix} \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}.$$

Calculando:

$$V D^n = \begin{pmatrix} 1 & \alpha(1 - \alpha - \beta)^n \\ 1 & -\beta(1 - \alpha - \beta)^n \end{pmatrix}.$$

Multiplicando por V^{-1} :

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} 1 & \alpha(1 - \alpha - \beta)^n \\ 1 & -\beta(1 - \alpha - \beta)^n \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}.$$

El elemento $(1, 1)$ de P^n es:

$$(P^n)_{1,1} = \frac{1}{\alpha + \beta} [\beta + \alpha(1 - \alpha - \beta)^n].$$

De manera similar, podemos calcular los demás elementos:

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha(1 - \alpha - \beta)^n & \alpha - \alpha(1 - \alpha - \beta)^n \\ \beta - \beta(1 - \alpha - \beta)^n & \alpha + \beta(1 - \alpha - \beta)^n \end{pmatrix}.$$

Comportamiento asintótico cuando $n \rightarrow \infty$:

Dado que $0 < \alpha, \beta < 1$, tenemos $0 < \alpha + \beta < 2$, y por lo tanto $1 - \alpha - \beta \in (-1, 1)$. Esto implica que $(1 - \alpha - \beta)^n \rightarrow 0$ cuando $n \rightarrow \infty$.

Por consiguiente:

$$\lim_{n \rightarrow \infty} P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}.$$

Este es un fenómeno notablemente importante: independientemente del estado inicial, después de un número muy grande de pasos, la probabilidad de encontrarse en el estado j es esencialmente la misma. Esto significa que la “memoria” del estado inicial se disipa con el

tiempo, y la cadena alcanza un régimen estacionario donde la distribución de probabilidades se estabiliza.

Esto establece que la distribución límite satisface una ecuación de equilibrio: las probabilidades no cambian de un paso al siguiente. Este comportamiento es intuitivo desde una perspectiva física: después de suficiente tiempo, el sistema “olvida” de dónde comenzó y se comporta como si estuviera en un estado de equilibrio. En las secciones siguientes especificaremos las condiciones bajo las cuales una cadena de Markov garantiza la existencia y unicidad de una distribución estacionaria, así como condiciones que aseguran la convergencia.

Cálculo de la inversa de una matriz

Para completar el cálculo de P^n mediante diagonalización, frecuentemente necesitamos calcular la inversa de la matriz V de vectores propios. A continuación recordamos el método general usando cofactores y determinantes.

Definición 1.5. Sea A una matriz cuadrada $m \times m$. La **matriz de cofactores** C se define como:

$$C_{ij} = (-1)^{i+j} M_{ij},$$

donde M_{ij} es el determinante del menor obtenido al eliminar la fila i y la columna j de A . La **matriz adjunta** (o **matriz de adjuntos**) es la transpuesta de la matriz de cofactores:

$$\text{adj}(A) = C^T.$$

Teorema 1.3. Sea A una matriz cuadrada con $\det(A) \neq 0$. Entonces:

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A).$$

Observación 1.9. Este método es especialmente útil para matrices pequeñas (por ejemplo 3×3). Para matrices grandes, métodos computacionales como eliminación gaussiana o descomposición LU suelen ser más eficientes.

Caso particular: Matriz 3×3 .

Para una matriz $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, primero calculamos:

$$\det(A) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).$$

Luego, la matriz de cofactores es:

$$C = \begin{pmatrix} +(a_{22}a_{33} - a_{23}a_{32}) & -(a_{21}a_{33} - a_{23}a_{31}) & +(a_{21}a_{32} - a_{22}a_{31}) \\ -(a_{12}a_{33} - a_{13}a_{32}) & +(a_{11}a_{33} - a_{13}a_{31}) & -(a_{11}a_{32} - a_{12}a_{31}) \\ +(a_{12}a_{23} - a_{13}a_{22}) & -(a_{11}a_{23} - a_{13}a_{21}) & +(a_{11}a_{22} - a_{12}a_{21}) \end{pmatrix}.$$

La matriz adjunta es $\text{adj}(A) = C^T$, y finalmente:

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A).$$

Ejemplo 1.18 (18). *Cadena de Markov con tres estados.* Consideremos una cadena de Markov con espacio de estados $S = \{1, 2, 3\}$ y matriz de transición:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

1.7. Comunicación entre Estados

Una pregunta fundamental en la teoría de cadenas de Markov es: ¿puede la cadena alcanzar un estado partiendo desde otro? Esta pregunta nos conduce a las nociones de **accesibilidad** y **comunicación**, conceptos que clasifican los estados de acuerdo con su “conectividad” en el grafo de transiciones.

Definición 1.6. Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov homogénea con espacio de estados S . Decimos que un estado j es **accesible** desde un estado i (denotado $i \rightarrow j$) si existe un entero $n \geq 0$ tal que:

$$\mathbb{P}(X_n = j \mid X_0 = i) = p_{ij}^{(n)} > 0.$$

Equivalentemente, existe un camino (posiblemente de longitud cero) que conecta i con j con probabilidad positiva.

Definición 1.7. Dos estados i y j **se comunican** (denotado $i \leftrightarrow j$) si son mutuamente accesibles, es decir, si $i \rightarrow j$ y $j \rightarrow i$.

Observación 1.10. La relación “se comunica con” es una relación de equivalencia: es reflexiva ($i \leftrightarrow i$ para todo i), simétrica ($i \leftrightarrow j \implies j \leftrightarrow i$), y transitiva ($i \leftrightarrow j$ y $j \leftrightarrow k$ implica $i \leftrightarrow k$).

Por lo tanto, la relación de comunicación particiona el espacio de estados S en **clases de comunicación** disjuntas. Dos estados pertenecen a la misma clase si y sólo si se comunican entre sí.

Ejemplo 1.19 (19). *Movilidad social revisitada.* Reconsideremos la cadena de Markov del Ejemplo 4 que modela la movilidad socioeconómica con estados $S = \{B, M, A\}$ (baja, media, alta). Supongamos que la matriz de transición es:

$$P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0 & 0 & 1 \end{pmatrix},$$

donde las filas y columnas corresponden a los estados B , M y A , respectivamente.

Analicemos la accesibilidad:

- Desde B podemos alcanzar M con probabilidad $p_{BM} = 0.3 > 0$, de modo que $B \rightarrow M$.
- Desde M podemos alcanzar A con probabilidad $p_{MA} = 0.3 > 0$, de modo que $M \rightarrow A$.
- Desde A no podemos alcanzar B o M

En consecuencia, tenemos dos clases de comunicación:

1. Clase 1: $\{B, M\}$ donde $B \leftrightarrow M$ (ambos se comunican).
2. Clase 2: $\{A\}$ que es una clase unipuntual (estado absorbente).

Observe que una vez que la cadena entra la tercera clase $\{A\}$ (estado de clase alta), permanece ahí indefinidamente con probabilidad 1.

Definición 1.8. Una cadena de Markov se llama **irreducible** si todos sus estados pertenecen a una única clase de comunicación, es decir, si cualesquiera dos estados $i, j \in S$ se comunican mutuamente ($i \leftrightarrow j$). De forma equivalente, todo estado es accesible desde cualquier otro estado.

Ejemplo 1.20 (20). **Cadena con varias clases de comunicación.** Consideremos una cadena de Markov con espacio de estados $S = \{1, 2, 3, 4, 5, 6\}$ y matriz de transición:

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Analicemos la estructura de comunicación:

- **Clase 1:** Estados $\{1, 2\}$. Desde el estado 1 podemos alcanzar el estado 2 con $p_{1,2} = 0.5 > 0$, y desde el estado 2 podemos alcanzar el estado 1 con $p_{2,1} = 0.3 > 0$. Por lo tanto, $1 \leftrightarrow 2$. Observe que desde estos estados no podemos escapar: $p_{1,3} = p_{1,4} = p_{1,5} = p_{1,6} = 0$ y $p_{2,3} = p_{2,4} = \dots = p_{2,6} = 0$.
- **Clase 2:** Estado $\{3\}$. Este es un estado absorbente: $p_{3,3} = 1$ y $p_{3,j} = 0$ para $j \neq 3$. Una vez que la cadena entra en este estado, permanece ahí para siempre.
- **Clase 3:** Estado $\{4\}$. Desde el estado 4 podemos alcanzar el estado 5 con probabilidad $p_{4,5} = 0.6 > 0$, pero desde el estado 5 tenemos $p_{5,5} = 1$ y $p_{5,4} = 0$, lo que significa que una vez en el estado 5, la cadena permanece allí. Así pues, $4 \rightarrow 5$ pero NO $5 \rightarrow 4$, de modo que 4 y 5 NO se comunican. Como $\{4\}$ no se comunica con otros estados, forma entonces su propia clase de comunicación.
- **Clase 4:** Estado $\{5\}$. Este es un estado absorbente: $p_{5,5} = 1$ y $p_{5,j} = 0$ para $j \neq 5$.
- **Clase 5:** Estado $\{6\}$. Este es un estado absorbente: $p_{6,6} = 1$.

Esta cadena es **reducible** porque posee cinco clases de comunicación.

Ejemplo 1.21 (21). **Caminata aleatoria simétrica.** La caminata aleatoria simétrica del Ejemplo 2 con espacio de estados $S = \mathbb{Z}$ es un ejemplo de cadena irreducible. Desde cualquier entero i , podemos alcanzar cualquier otro entero j siguiendo una secuencia de pasos hacia

arriba (incrementos de $+1$) y hacia abajo (decrementos de -1). Por ejemplo, desde el estado 0 podemos alcanzar el estado 5 mediante la secuencia $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$, cada transición ocurriendo con probabilidad $\frac{1}{2}$. De hecho, existen múltiples caminos desde 0 hasta 5, por lo que $0 \rightarrow 5$ (y mutuamente $5 \rightarrow 0$).

Esta irreducibilidad es una propiedad estructural fundamental de la caminata aleatoria simétrica: la cadena “explora” potencialmente todo el espacio de estados.

Ejemplo 1.22 (22). *Cadena de las horas del reloj.* Recordemos el Ejemplo 14 de la cadena de las horas. Desde cualquier hora i , podemos alcanzar cualquier otra hora j avanzando $j - i$ (mód 12) horas. Por ejemplo, desde las 3 podemos alcanzar las 7 avanzando 4 horas ($3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$), o avanzando 16 horas ($3 \rightarrow 4 \rightarrow \dots \rightarrow 19 \equiv 7 \pmod{12}$). Además, podemos regresar desde las 7 a las 3 avanzando 8 horas. Por lo tanto, dos estados cualesquiera se comunican, y la cadena es irreducible.

1.8. Distribuciones Periódicas: Período de un Estado

Hemos visto que ciertos estados pueden ser accesibles desde otros, pero existe una propiedad adicional que afecta la estructura de la cadena: el **período** de un estado. El período mide la periodicidad con la cual la cadena puede regresar a un estado dado.

Definición 1.9. El periodo de un estado i es un número entero no negativo, denotado por d_i , y definido como

$$d_i = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\},$$

donde $\gcd$ denota el máximo común divisor. Si $p_{ii}^{(n)} = 0$ para toda $n \geq 1$, se define $d_i = 0$. En particular, se dice que un estado i es aperiódico si $d_i = 1$. Cuando $d_i = k \geq 2$, se dice que i es periódico de periodo k .

Ejemplo 1.23 (22). *Cadena de las horas cíclica.* Recordemos la cadena del Ejemplo 14, donde $X_{n+1} = (X_n + 1) \pmod{12}$. Debido a que cada estado se alcanza incrementando el reloj en 1 unidad, tenemos:

- La cadena retorna al estado 0 (las 12 en punto) después de exactamente 12 pasos: $p_{0,0}^{(12)} = 1$.
- La cadena no retorna al estado 0 en ningún otro número de pasos: $p_{0,0}^{(n)} = 0$ para $n \neq 12, 24, 36, \dots$
- Por lo tanto, el conjunto de retorno es $\{12, 24, 36, \dots\} = \{12k : k \geq 1\}$.
- El período es $d_0 = \gcd(12, 24, 36, \dots) = 12$.

De forma similar, para cualquier hora i , el período es $d_i = 12$. La cadena es **periódica con período 12**.

Ejemplo 1.24 (23). *Caminata aleatoria simétrica.* Para la caminata aleatoria simétrica del Ejemplo 2, considere un estado particular, digamos el 0. Tenemos:

- En 1 paso: $p_{0,0}^{(1)} = 0$ (no podemos regresar en 1 paso puesto que cada paso es ± 1).
- En 2 pasos: $p_{0,0}^{(2)} = \mathbb{P}(Y_1 = 1, Y_2 = -1) + \mathbb{P}(Y_1 = -1, Y_2 = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} > 0$.
- En 3 pasos: $p_{0,0}^{(3)} = 0$ (imposible regresar en un número impar de pasos).
- En 4 pasos: $p_{0,0}^{(4)} > 0$ (retornos como $+1, +1, -1, -1$ o $+1, -1, +1, -1$, etc.).

El conjunto de retorno es $\{2, 4, 6, 8, \dots\}$ (todos los números pares), y el período es:

$$d_0 = \gcd(2, 4, 6, 8, \dots) = 2.$$

Proposición 1.5. El período es una **propiedad de clase**: si dos estados i y j se comunican ($i \leftrightarrow j$), entonces tienen el mismo período: $d_i = d_j$.

Demostración. Supongamos que $i \leftrightarrow j$. Entonces existen $m, n \geq 1$ tales que

$$p_{ij}^{(m)} > 0, \quad p_{ji}^{(n)} > 0.$$

Definimos

$$A_i = \{r \geq 1 : p_{ii}^{(r)} > 0\}, \quad A_j = \{r \geq 1 : p_{jj}^{(r)} > 0\},$$

y $d_i = \gcd(A_i)$, $d_j = \gcd(A_j)$.

Paso 1: $d_j \mid d_i$.

Sea $r \in A_i$. Entonces $p_{ii}^{(r)} > 0$. Por Chapman–Kolmogorov,

$$p_{jj}^{(m+r+n)} \geq p_{ji}^{(n)} p_{ii}^{(r)} p_{ij}^{(m)} > 0,$$

luego $m + n + r \in A_j$.

Como también $p_{ii}^{(2r)} > 0$, se obtiene igualmente

$$m + n + 2r \in A_j.$$

Como d_j divide todos los elementos de A_j , divide $m + n + r$ y $m + n + 2r$, y por tanto su diferencia:

$$(m + n + 2r) - (m + n + r) = r.$$

Así, todo $r \in A_i$ es divisible por d_j , y por lo tanto

$$d_j \mid d_i.$$

Paso 2: $d_i \mid d_j$.

Intercambiando los papeles de i y j y repitiendo el argumento, se obtiene

$$d_i \mid d_j.$$

Concluimos que $d_i = d_j$. □

Proposición 1.6. Si un estado i en una cadena de Markov tiene período d_i , entonces existe un entero N (que depende de i) tal que para todo $n \geq N$:

$$p_{ii}^{(n \cdot d_i)} > 0.$$

Como corolario de la proposición anterior se tiene que si acaso es posible pasar de i a j en m pasos, entonces también es posible tal transición en $m + n \cdot d_j$ pasos con n suficientemente grande, suponiendo $d_j \geq 1$.

1.9. Primeras Visitas

Una pregunta fundamental en la teoría de cadenas de Markov es: ¿cuál es la probabilidad de que una cadena alcance un estado (o conjunto de estados) por primera vez exactamente en el paso n ? Esta pregunta lleva a la definición de **tiempo de primera visita** y a las correspondientes probabilidades, que son herramientas esenciales para estudiar la estructura y el comportamiento a largo plazo de las cadenas de Markov.

Definición 1.10. Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov con espacio de estados S , y sea $B \subseteq S$ un conjunto no vacío de estados. El **tiempo de primera visita** a B partiendo del estado i se define como:

$$\tau_B^{(i)} = \min\{n \geq 1 : X_n \in B \mid X_0 = i\},$$

donde convenimos que $\tau_B^{(i)} = \infty$ si la cadena nunca alcanza B partiendo de i .

Observación 1.11. Para el caso especial donde $B = \{j\}$ es un estado único, escribimos simplemente $\tau_j^{(i)}$ en lugar de $\tau_{\{j\}}^{(i)}$. El evento $\{\tau_j^{(i)} = n\}$ significa que la cadena alcanza el estado j por primera vez en exactamente n pasos, partiendo del estado i . Este evento puede escribirse como:

$$\{\tau_j^{(i)} = n\} = \{X_n = j, X_k \neq j \text{ para todo } k = 1, 2, \dots, n-1 \mid X_0 = i\}.$$

Definición 1.11. Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov con espacio de estados S , y sean $i, j \in S$. Definimos la **probabilidad de primera visita de i a j en n pasos** como:

$$f_{ij}^{(n)} = \mathbb{P}(\tau_j^{(i)} = n) = \mathbb{P}(X_n = j, X_k \neq j \text{ para } k = 1, \dots, n-1 \mid X_0 = i).$$

Alternativamente, es la probabilidad de que, comenzando en i , la cadena llegue a j por primera vez exactamente en n pasos.

Ejemplo 1.25 (25). *Caminata aleatoria sesgada en tres estados.* Considere una cadena de Markov con espacio de estados $S = \{1, 2, 3\}$ y matriz de transición:

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.2 & 0.4 & 0.4 \\ 0 & 0 & 1 \end{pmatrix}.$$

Aquí el estado 3 es absorbente. Queremos calcular $f_{1,3}^{(n)}$, la probabilidad de que, comenzando en el estado 1, la cadena alcance el estado 3 por primera vez en exactamente n pasos.

Para $n = 1$: $f_{1,3}^{(1)} = \mathbb{P}(X_1 = 3 \mid X_0 = 1) = p_{1,3} = 0$ (no hay transición directa desde 1 a 3).

Para $n = 2$: El único camino de 1 a 3 sin pasar por 3 en el paso 1 es $1 \rightarrow 2 \rightarrow 3$, con probabilidad $p_{1,2} \cdot p_{2,3} = 0.7 \times 0.4 = 0.28$. Por lo tanto, $f_{1,3}^{(2)} = 0.28$.

Para $n = 3$: Hay dos caminos:

- $1 \rightarrow 1 \rightarrow 2 \rightarrow 3$: probabilidad $p_{1,1} \cdot p_{1,2} \cdot p_{2,3} = 0.3 \times 0.7 \times 0.4 = 0.084$.

- $1 \rightarrow 2 \rightarrow 2 \rightarrow 3$: probabilidad $p_{1,2} \cdot p_{2,2} \cdot p_{2,3} = 0.7 \times 0.4 \times 0.4 = 0.112$.

Por lo tanto, $f_{1,3}^{(3)} = 0.084 + 0.112 = 0.196$.

Proposición 1.7. *La probabilidad de transición en n pasos desde i a j puede descomponerse en términos de la probabilidad de primera visita. Específicamente, si la cadena accede de i a j en exactamente n pasos, entonces debe haber alcanzado j por primera vez en algún paso $k \leq n$, y luego hacer la trayectoria completa de j a j en $n - k$ pasos. Formalmente:*

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} \cdot p_{jj}^{(n-k)},$$

donde convenimos que $p_{jj}^{(0)} = 1$ (sin dar pasos, permanecemos en j).

Demostración. Particionamos todos los caminos de i a j en n pasos de acuerdo al tiempo en el que se realiza la primera visita a j :

$$\{X_n = j \mid X_0 = i\} = \bigcup_{k=1}^n \{X_n = j, \tau_j^{(i)} = k \mid X_0 = i\}.$$

Estos eventos son disjuntos para diferentes valores de k . Por lo tanto:

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i) = \sum_{k=1}^n \mathbb{P}(X_n = j, \tau_j^{(i)} = k \mid X_0 = i).$$

Ahora, si $\tau_j^{(i)} = k$, entonces $X_k = j$ y la cadena no ha visitado j en los pasos $1, \dots, k-1$. De forma condicional a $X_k = j$, el evento $\{X_n = j\}$ depende únicamente de la evolución de la cadena desde el paso k en adelante, que es independiente del pasado (propiedad de Markov). Así:

$$\mathbb{P}(X_n = j, \tau_j^{(i)} = k \mid X_0 = i) = \mathbb{P}(\tau_j^{(i)} = k \mid X_0 = i) \cdot \mathbb{P}(X_n = j \mid X_k = j) = f_{ij}^{(k)} \cdot p_{jj}^{(n-k)}.$$

Sumando sobre k , obtenemos la fórmula deseada. □

Definición 1.12. *Definimos la **probabilidad de eventual alcance** de j partiendo de i como:*

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = \mathbb{P}(\tau_j^{(i)} < \infty \mid X_0 = i).$$

Esta es la probabilidad total de que la cadena alcance el estado j en algún momento del futuro, partiendo del estado i .

Observación 1.12. *La cantidad f_{ij} representa la probabilidad de que la cadena alguna vez visite el estado j comenzando desde i . Tiene interpretaciones muy útiles:*

- Si $f_{ij} = 1$, decimos que j es **accesible con certeza** desde i : con probabilidad total, la cadena eventualmente alcanzará j .

- Si $f_{ij} < 1$, existe una probabilidad positiva $1 - f_{ij}$ de que la cadena nunca alcance j partiendo de i .
- En particular, si $i = j$, entonces f_{ii} es la probabilidad de que la cadena regrese al estado i en algún momento futuro. Esta cantidad es crucial para definir nociones de **recurrencia** y **transitoriedad** de los estados.

Ejemplo 1.26 (26). Caminata aleatoria simétrica: Probabilidad de retorno. Para la caminata aleatoria simétrica en $\mathbb{Z}$ del Ejemplo 2, calculemos $f_{0,0}$, la probabilidad de que la cadena regrese alguna vez al estado 0 partiendo de 0.

Recordemos que el período es $d_0 = 2$, así que la cadena solo puede retornar al 0 en tiempos pares. Para $n = 2k$ (número par), la probabilidad de retorno en exactamente 2 pasos es:

$$f_{0,0}^{(2)} = \mathbb{P}(X_2 = 0, X_1 \neq 0 \mid X_0 = 0) = \mathbb{P}(Y_1 = 1, Y_2 = -1) + \mathbb{P}(Y_1 = -1, Y_2 = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Para el retorno en 4 pasos, debemos evitar retornar en los primeros 2 pasos. Los caminos de retorno en 4 pasos sin tocar 0 intermedariamente incluyen secuencias como $(+1, +1, -1, -1)$, $(+1, -1, +1, -1)$, etc. El cálculo directo es más complejo, pero mediante técnicas analíticas (funciones generadoras), puede demostrarse que:

$$f_{0,0} = \sum_{n=1}^{\infty} f_{0,0}^{(2n)} = 1.$$

Este es un resultado célebre: la caminata aleatoria simétrica en $\mathbb{Z}$ es **recurrente**, es decir, con probabilidad 1, la cadena regresa al estado inicial (y de hecho, visita cualquier estado infinitamente a menudo).

Ejemplo 1.27 (27). Caminata aleatoria sesgada: Transitoriedad. Consideremos una caminata aleatoria en $\mathbb{Z}$ con $\mathbb{P}(Y_i = 1) = p$ y $\mathbb{P}(Y_i = -1) = q = 1 - p$, donde $p \neq \frac{1}{2}$.

Supongamos $p = 0.6$ (sesgada hacia la derecha). Intuitivamente, si comenzamos en el estado 0, con probabilidad 1 la cadena “fluye” hacia $+\infty$, por lo que la probabilidad de retornar al 0 debe ser menor que 1. De hecho, mediante análisis de funciones generadoras, puede demostrarse que:

$$f_{0,0} = \left(\frac{q}{p}\right) = \frac{0.4}{0.6} = \frac{2}{3} < 1.$$

Por lo tanto, existe una probabilidad positiva $1 - f_{0,0} = \frac{1}{3}$ de que la cadena nunca retorne al estado 0. En este caso, decimos que la cadena es **transitoria**: comenzando en cualquier estado, existe una probabilidad positiva de nunca visitarlo nuevamente.

Ejemplo 1.28 (28). Cadena con estado absorbente. Reconsideremos el Ejemplo 25 con matriz de transición:

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.2 & 0.4 & 0.4 \\ 0 & 0 & 1 \end{pmatrix}.$$

Ya calculamos que:

- $f_{1,3}^{(1)} = 0$ (no hay transición directa desde 1 a 3).
- $f_{1,3}^{(2)} = 0.28$ (única ruta $1 \rightarrow 2 \rightarrow 3$).
- $f_{1,3}^{(3)} = 0.196$ (dos rutas).

La probabilidad de eventual alcance es $f_{1,3} = \sum_{n=1}^{\infty} f_{1,3}^{(n)}$. Dado que el estado 3 es absorbente y la cadena no puede “escapar” a infinito, intuitivamente esperamos que $f_{1,3} = 1$: comenzando en cualquier estado 1 o 2, eventualmente la cadena será absorbida por el estado 3 con certeza.

De hecho, de la matriz podemos ver que desde el estado 1 o 2, la única forma de no alcanzar el estado 3 es permanecer indefinidamente en los estados 1 y 2. Pero estos estados se comunican (viajan entre sí) con probabilidad positiva, y eventualmente la cadena saldrá de este “grupo” y será absorbida por 3. Así, $f_{1,3} = 1$.

— Termina material del primer parcial —

1.10. Estados Recurrentes y Transitorios

Veremos a continuación que los estados de una cadena de Markov pueden ser clasificados, en una primera instancia, en dos tipos, dependiendo de si la cadena es capaz de regresar con certeza al estado de partida. Esta dicotomía es esencial para entender el comportamiento a largo plazo de la cadena: en algunos modelos, la cadena vuelve inevitablemente a ciertos estados; en otros, existe una probabilidad positiva de abandonar un estado para no regresar nunca.

En esta sección utilizaremos la noción de **primeras visitas** introducida previamente. En particular, para un estado $i \in S$ recordemos que

$$f_{ii} = \mathbb{P}(\tau_i^{(i)} < \infty \mid X_0 = i)$$

es la probabilidad de volver a visitar el estado i en algún tiempo futuro (en un tiempo finito), partiendo de i .

Definición 1.13 (Recurrencia y transitoriedad (I)). *Se dice que un estado i es **recurrente** si la probabilidad de eventualmente regresar a i , partiendo de i , es uno; es decir, si*

$$\mathbb{P}(X_n = i \text{ para alguna } n \geq 1 \mid X_0 = i) = 1.$$

Un estado que no es recurrente se llama **transitorio**; en tal caso, la probabilidad anterior es estrictamente menor que uno.

De manera intuitiva, un estado es recurrente si con probabilidad uno la cadena es capaz de regresar eventualmente a ese estado. Más aún, cuando la cadena regresa en algún tiempo finito, por la propiedad de Markov el proceso se *reinicia* desde i y, en ese sentido, puede regresar a i una y otra vez con probabilidad uno. En cambio, el estado se llama transitorio si existe una probabilidad positiva de que la cadena, iniciando en él, ya no regrese nunca a ese estado.

La definición anterior puede enunciarse de manera equivalente en términos de las probabilidades de primera visita.

Definición 1.14 (Recurrencia y transitoriedad (II)). *Un estado i es **recurrente** si $f_{ii} = 1$, es decir, si la probabilidad de regresar a i en un tiempo finito es uno. Análogamente, un estado i es **transitorio** si $f_{ii} < 1$.*

Observación 1.13. *La cantidad f_{ii} se interpreta como una probabilidad de retorno eventual. En contraste, las cantidades $p_{ii}^{(n)}$ miden la probabilidad de estar en i exactamente después de n pasos. El criterio siguiente conecta ambas familias de cantidades.*

Además de la definición, tenemos el siguiente criterio útil para determinar si un estado es recurrente o transitorio.

Proposición 1.8 (Criterio de recurrencia/transitoriedad). *Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov homogénea con espacio de estados S y sea $i \in S$. Entonces:*

(a) *El estado i es **recurrente** si, y sólo si,*

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty.$$

(b) *El estado i es **transitorio** si, y sólo si,*

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty.$$

Demostración. Fijemos un estado $i \in S$ y supongamos que la cadena inicia en $X_0 = i$. Definamos la variable aleatoria N_i como el número total de visitas al estado i después del instante inicial:

$$N_i = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=i\}}.$$

Por definición, la probabilidad de regresar al menos una vez al estado i en un tiempo finito es f_{ii} . Si el proceso regresa, por la propiedad fuerte de Markov evaluada en ese primer instante de retorno, la cadena se reinicia. Por ende, la probabilidad de realizar al menos k retornos, dado que ya presenciamos $k - 1$ retornos, sigue siendo f_{ii} . Utilizando este argumento de manera inductiva, la suma de los primeros retornos nos da que la probabilidad de que haya al menos k visitas es:

$$\mathbb{P}(N_i \geq k \mid X_0 = i) = f_{ii}^k, \quad k \geq 1.$$

Si reescribimos esto en su forma puntual, $\mathbb{P}(N_i = k \mid X_0 = i) = \mathbb{P}(N_i \geq k) - \mathbb{P}(N_i \geq k+1) = f_{ii}^k(1 - f_{ii})$. Es decir, vemos fácilmente que **la variable aleatoria N_i tiene una distribución geométrica** sobre el conjunto $\{0, 1, 2, \dots\}$ con parámetro de "éxito" $1 - f_{ii}$.

Procedamos a calcular $\mathbb{E}[N_i \mid X_0 = i]$ de dos maneras distintas. Usando la definición por variables indicadoras, la linealidad de la esperanza y el teorema de convergencia monótona nos otorgan:

$$\mathbb{E}[N_i \mid X_0 = i] = \mathbb{E} \left[\sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=i\}} \mid X_0 = i \right] = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i \mid X_0 = i) = \sum_{n=1}^{\infty} p_{ii}^{(n)}.$$

Alternativamente, como N_i es una variable aleatoria no negativa, su esperanza puede obtenerse sumando las probabilidades de sus colas:

$$\mathbb{E}[N_i \mid X_0 = i] = \sum_{k=1}^{\infty} \mathbb{P}(N_i \geq k \mid X_0 = i) = \sum_{k=1}^{\infty} f_{ii}^k.$$

Para concluir, veamos los dos casos posibles:

- Si $f_{ii} < 1$ (es decir, i es transitorio), la suma de las colas es una serie geométrica convergente con valor $\frac{f_{ii}}{1-f_{ii}} < \infty$. Igualando ambos cálculos, encontramos que $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.
- Si $f_{ii} = 1$ (es decir, i es recurrente), los términos en la serie no decaen, por lo que la suma de las colas diverge ($\mathbb{E}[N_i \mid X_0 = i] = \infty$). Por equivalencia, esto implica directamente que $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

Esto demuestra los incisos (a) y (b) estableciendo el criterio de sumabilidad. $\square$

Observación 1.14. Como se evidenció en la prueba, $1 + \mathbb{E}[N_i \mid X_0 = i] = \sum_{n=0}^{\infty} p_{ii}^{(n)}$ representa el número total esperado de visitas al estado i , partiendo de i e incluyendo el tiempo $t = 0$. El criterio establece, de manera natural e intuitiva, que un estado es transitorio si su número total esperado de visitas es finito, lo cual concuerda con una probabilidad de retorno estrictamente menor a uno; por el contrario, un estado es recurrente si esperamos visitarlo un número infinito de veces.

Ejemplo 1.29. Un estado absorbente es recurrente. Si i es absorbente, entonces $p_{ii} = 1$ y por lo tanto $p_{ii}^{(n)} = 1$ para todo $n \geq 1$. En consecuencia,

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty,$$

de modo que i es recurrente.

Ejemplo 1.30. Un ejemplo transitorio en dos estados. Considere la cadena con $S = \{0, 1\}$ y

$$P = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}, \quad 0 < p < 1.$$

Partiendo de 0, se tiene $p_{00}^{(n)} = (1 - p)^n$, y por lo tanto

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} (1 - p)^n = \frac{1 - p}{p} < \infty.$$

Así, el estado 0 es transitorio. (En cambio, el estado 1 es absorbente y por ende recurrente.)

El criterio anterior nos permite establecer una tercera caracterización, quizás la más intuitiva, basada en el número promedio de retornos.

Definición 1.15 (Recurrencia y transitoriedad (III)). Sea N_i el número total de visitas al estado i después del tiempo 0. Decimos que el estado i es **recurrente** si, y sólo si, el número promedio de retornos a él es infinito:

$$\mathbb{E}[N_i \mid X_0 = i] = \infty.$$

Equivalentemente, i es **transitorio** si, y sólo si, $\mathbb{E}[N_i \mid X_0 = i] < \infty$.

Ejemplo 1.31 (Caminata aleatoria simple en $\mathbb{Z}$). Consideremos la caminata aleatoria donde $p_{i,i+1} = p$ y $p_{i,i-1} = 1 - p$. Del análisis combinatorio (o utilizando la fórmula de Stirling), conocemos el comportamiento asintótico de $p_{00}^{(n)}$.

Caso simétrico ($p = 1/2$): Sabemos que $p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}$. Al sumar estos términos, obtenemos una serie armónica generalizada (tipo $1/n^{1/2}$) que diverge:

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty.$$

Por lo tanto, la caminata aleatoria simétrica es **recurrente**.

Caso asimétrico ($p \neq 1/2$): En este caso, la probabilidad de retorno decae, grosso modo, exponencialmente: $p_{00}^{(2n)} \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$. Como $p \neq 1/2$, se tiene $4p(1-p) < 1$, por lo que la serie converge:

$$\sum_{n=1}^{\infty} p_{00}^{(n)} < \infty.$$

Por lo tanto, la caminata asimétrica es **transitoria**.

A continuación demostramos que la recurrencia es una propiedad compartida por todos los estados de una misma clase de comunicación.

Proposición 1.9 (La recurrencia es propiedad de clase). Sean $i, j \in S$ dos estados que se comunican ($i \leftrightarrow j$).

- (a) Si i es recurrente, entonces j es recurrente.
- (b) Si i es transitorio, entonces j es transitorio.

Demostración. Demostraremos (a). Supongamos que i es recurrente. Como $i \leftrightarrow j$, existen enteros $n, m \geq 0$ tales que $p_{ij}^{(n)} > 0$ y $p_{ji}^{(m)} > 0$. Para cualquier entero $k \geq 0$, consideremos la probabilidad de ir de j a j en $m + k + n$ pasos, forzando la trayectoria a pasar por i . Por la desigualdad de Chapman-Kolmogorov:

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

Sumando sobre todos los $k \geq 0$:

$$\sum_{k=0}^{\infty} p_{jj}^{(m+k+n)} \geq p_{ji}^{(m)} p_{ij}^{(n)} \sum_{k=0}^{\infty} p_{ii}^{(k)}.$$

Como i es recurrente, $\sum p_{ii}^{(k)} = \infty$. Dado que $p_{ji}^{(m)}$ y $p_{ij}^{(n)}$ son positivos, el lado derecho es infinito (producto de constantes positivas por infinito). Esto implica que $\sum_t p_{jj}^{(t)} = \infty$, por lo que j es recurrente. La parte (b) se sigue por el contrarrecíproco de (a). $\square$

Esto significa que, al descomponer la cadena en clases de comunicación, podemos clasificar cada clase entera como recurrente o transitoria. Si la cadena es irreducible, entonces **toda la cadena** es recurrente o **toda la cadena** es transitoria.

Ejemplo 1.32. En la cadena de dos estados si $a, b \in (0, 1)$, entonces la cadena es irreducible y recurrente. Calculemos explícitamente f_{00} o recordemos $p_{00}^{(n)}$.

Ejemplo 1.33 (Cadena de rachas). Consideremos la cadena de rachas de éxitos (obtener sol) en lanzamientos de una moneda (con prob. de éxito p). Esta cadena es irreducible en $S = \{0, 1, 2, \dots\}$. No es difícil demostrar, por ejemplo calculando $f_{0,0}$, que el estado 0 es recurrente. Como es irreducible, **todos** los estados son recurrentes. Comentario: Incluso un estado “difícil” de alcanzar como 9999 (una racha de 9999 éxitos seguidos) es recurrente y será visitado en promedio infinitas veces con probabilidad 1. Cabe preguntarse ¿Qué tan frecuentes serán esas visitas?.

Proposición 1.10 (Existencia de recurrentes en espacios finitos). Toda cadena de Markov con espacio de estados finito tiene por lo menos un estado recurrente.

Demostración. Demostración por contradicción. Supongamos que todos los estados son transitorios. Sea S finito. Si todos son transitorios, entonces para todo i, j , $\sum_{n=0}^{\infty} p_{ij}^{(n)} < \infty$. Consideremos la suma sobre j : para cada n , $\sum_{j \in S} p_{ij}^{(n)} = 1$. Sumando sobre n :

$$\sum_{n=0}^{\infty} \sum_{j \in S} p_{ij}^{(n)} = \sum_{n=0}^{\infty} 1 = \infty.$$

Intercambiando la suma (válido por ser S finito):

$$\sum_{j \in S} \left(\sum_{n=0}^{\infty} p_{ij}^{(n)} \right) = \infty.$$

Esto contradice el hecho de que cada suma interna $\sum_n p_{ij}^{(n)}$ es finita (pues todo j es transitorio) y la suma de un número finito de términos finitos debe ser finita. Por lo tanto, debe existir al menos un estado recurrente. $\square$

La siguiente proposición nos da información sobre el comportamiento límite de las probabilidades de transición hacia estados transitorios.

Proposición 1.11. *Si j es un estado transitorio, entonces para todo i :*

$$\sum_{n=1}^{\infty} p_{i,j}^{(n)} < \infty \quad \Rightarrow \quad \lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0.$$

Demostración. Sea $f_{ij}^{(n)} = \mathbb{P}_i(T_j = n)$ la probabilidad de que la cadena visite j por primera vez en el tiempo n . Para $n \geq 1$,

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}.$$

Sumando en n desde 1 hasta ∞ obtenemos

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}.$$

Como todos los términos son no negativos, podemos cambiar el orden de suma:

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \sum_{m=0}^{\infty} p_{jj}^{(m)}.$$

Observamos que

$$\sum_{k=1}^{\infty} f_{ij}^{(k)} = f_{ij} \leq 1.$$

Además, como j es transitorio,

$$\sum_{m=0}^{\infty} p_{jj}^{(m)} < \infty.$$

Por lo tanto,

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = f_{ij} \sum_{m=0}^{\infty} p_{jj}^{(m)} < \infty.$$

Concluimos que la serie $\sum_{n=1}^{\infty} p_{ij}^{(n)}$ converge. □

1.10.1. Visitas a un estado: el caso general N_{ij}

Generalicemos la noción de visitas. Sea N_{ij} el número total de visitas al estado j hasta el tiempo n partiendo desde i :

$$N_{ij}(n) = \sum_{k=1}^n \mathbf{1}_{\{X_k=j\}} \quad \text{dado que } X_0 = i.$$

y definamos $N_{ij} = \lim_{n \rightarrow \infty} N_{ij}(n)$ como el número total de visitas a j a lo largo de toda la evolución de la cadena, partiendo de i .

Proposición 1.12. *Para cualesquiera estados $i, j \in S$. La variable N_{ij} con soporte en $\{0, 1, 2, \dots\} \cup \{\infty\}$ satisface las siguientes propiedades:*

1. **La cola de la distribución de N_{ij} :**

$$\mathbb{P}(N_{ij} \geq k) := \mathbb{P}(N_i \geq k \mid X_0 = i) = f_{ij} f_{jj}^{k-1}, \quad \text{para } k \geq 1.$$

2. **La función de masa de N_{ij} :**

$$\mathbb{P}(N_{ij} = k) := \mathbb{P}(N_j = k \mid X_0 = i) = \begin{cases} 1 - f_{ij} & \text{si } k = 0 \\ f_{ij}(1 - f_{jj})f_{jj}^{k-1} & \text{si } k \geq 1 \end{cases}$$

3. **La esperanza:**

$$\mathbb{E}[N_{ij}] := \mathbb{E}[N_j \mid X_0 = i] = \frac{f_{ij}}{1 - f_{jj}}.$$

(Interpretamos $1/0$ como ∞ si $f_{jj} = 1$).

4. **Probabilidad de infinitas visitas:**

- Si j es transitorio ($f_{jj} < 1$): $\mathbb{P}(N_{ij} = \infty) = 0$.
- Si j es recurrente ($f_{jj} = 1$): $\mathbb{P}(N_{ij} = \infty) = f_{ij}$.

5. **Probabilidad de finitas visitas:**

- Si j es transitorio: $\mathbb{P}(N_{ij} < \infty) = 1$.
- Si j es recurrente: $\mathbb{P}(N_{ij} < \infty) = 1 - f_{ij}$ (es la probabilidad de nunca visitarlo).

Demostración. Demostraremos únicamente el primer inciso. Procedemos condicionando sobre el tiempo de la primera visita al estado j , denotado por τ_j .

$$\mathbb{P}(N_j \geq k \mid X_0 = i) = \sum_{n=1}^{\infty} \mathbb{P}(N_j \geq k \mid \tau_j = n, X_0 = i) \mathbb{P}(\tau_j = n \mid X_0 = i). \quad (1.1)$$

Si la primera visita ocurre en el tiempo n , por la propiedad fuerte de Markov el proceso se reinicia en j . Para acumular al menos k visitas en total, se requieren al menos $k - 1$ visitas adicionales (retornos) partiendo ahora desde j . Por la demostración del Criterio de Recurrencia (donde vimos que N_{jj} es geométrico), sabemos que $\mathbb{P}(N_{jj} \geq k - 1 \mid X_0 = j) = f_{jj}^{k-1}$. Sustituyendo esto en la suma:

$$\mathbb{P}(N_{ij} \geq k \mid X_0 = i) = \sum_{n=1}^{\infty} f_{jj}^{k-1} f_{ij}^{(n)} = f_{jj}^{k-1} \sum_{n=1}^{\infty} f_{ij}^{(n)} = f_{jj}^{k-1} f_{ij}.$$

Los demás incisos se siguen directamente de este resultado. □

Observación 1.15 (Propiedad fuerte de Markov). *En la última demostración hemos utilizado la **propiedad fuerte de Markov**. Enunciémosla brevemente.*

Sea τ un tiempo de paro respecto a la filtración natural de la cadena, es decir, una variable aleatoria con valores en $\mathbb{N} \cup \{\infty\}$ tal que el evento $\{\tau = n\}$ depende únicamente de $X_0, X_1, \dots, X_n$. La propiedad fuerte de Markov establece que, condicionado en que $\tau < \infty$ y en el valor $X_\tau = i$, el proceso posterior $\{X_{\tau+n}\}_{n \geq 0}$ tiene la misma distribución que la cadena iniciando en i :

$$\mathbb{P}(X_{\tau+n} = j \mid X_\tau = i, \tau < \infty) = p_{ij}^{(n)}.$$

¿En qué se diferencia de la propiedad de Markov ordinaria? La propiedad de Markov aplica en tiempos deterministas n : el futuro depende del presente pero no del pasado. La propiedad fuerte extiende esto a tiempos de paro aleatorios, como el tiempo de primera visita τ_j . Esto no es evidente: el tiempo τ_j puede ser arbitrariamente grande y la cadena vista antes de τ_j puede revelar información complicada sobre el pasado.

Antes de enunciar el teorema ergódico, resumamos las implicaciones de la clasificación en recurrente/transitorio para el número de visitas a un estado j partiendo de un estado i cualquiera.

- **Transitoriedad y número de visitas:** Si j es transitorio, entonces sin importar el estado inicial i , con probabilidad uno la cadena realiza sólo un número finito de visitas al estado j . El número esperado de visitas es siempre finito:

$$\mathbb{E}[N_{ij}] = \sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty \implies \lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0.$$

- **Recurrencia y número de visitas:** Si j es recurrente y se inicia en j , entonces con probabilidad uno la cadena regresa a j una infinidad de veces, y el número esperado de visitas es infinito. Si la cadena inicia en cualquier otro estado i , existe la posibilidad de que nunca visite j ($f_{ij} < 1$). Pero si la cadena visita j alguna vez (con prob. f_{ij}), entonces regresará a j una infinidad de veces.

1.11. Tiempo medio de retorno

Hemos visto que si una cadena de Markov inicia en un estado recurrente, entonces regresa a él una infinidad de veces con probabilidad uno. Ya definimos τ_j como el tiempo de primera visita. Ahora definiremos el tiempo medio de retorno.

Definición 1.16 (Tiempo medio de retorno). *Para un estado recurrente j , definimos el tiempo medio de retorno μ_{jj} como la esperanza de τ_j partiendo de j :*

$$\mu_{jj} = \mathbb{E}[\tau_j \mid X_0 = j] = \sum_{n=1}^{\infty} n f_{jj}^{(n)}.$$

Ejemplo 1.34 (Cadena de dos estados). Consideremos nuevamente la cadena de dos estados $(0, 1)$ con matriz de transición

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Queremos calcular el tiempo medio de retorno al estado 0 usando la definición $\mu_{00} = \sum_{n=1}^{\infty} n f_{00}^{(n)}$. Primero calculemos las probabilidades de primer retorno $f_{00}^{(n)}$:

- Para $n = 1$: El único camino es $0 \rightarrow 0$. Así, $f_{00}^{(1)} = p_{00} = 1 - a$.
- Para $n > 1$: El camino debe ser de la forma $0 \rightarrow 1 \rightarrow 1 \cdots \rightarrow 1 \rightarrow 0$. Es decir, salimos de 0 a 1 (probabilidad a), permanecemos en 1 durante $n-2$ pasos (probabilidad $(1-b)^{n-2}$), y finalmente regresamos a 0 (probabilidad b).

$$f_{00}^{(n)} = a(1-b)^{n-2}b, \quad n \geq 2.$$

Ahora calculamos la esperanza:

$$\begin{aligned} \mu_{00} &= 1 \cdot f_{00}^{(1)} + \sum_{n=2}^{\infty} n f_{00}^{(n)} \\ &= (1-a) + \sum_{n=2}^{\infty} n \cdot ab(1-b)^{n-2}. \end{aligned}$$

Haciendo el cambio de variable $k = n - 2$ (de modo que $n = k + 2$), la suma se convierte en:

$$ab \sum_{k=0}^{\infty} (k+2)(1-b)^k = ab \left[\sum_{k=0}^{\infty} k(1-b)^k + 2 \sum_{k=0}^{\infty} (1-b)^k \right].$$

Recordando las series geométricas para $|1-b| < 1$ (pues $b \in (0, 1)$): $\sum_{k=0}^{\infty} (1-b)^k = \frac{1}{1-(1-b)} = \frac{1}{b}$ y $\sum_{k=0}^{\infty} k(1-b)^k = \frac{1-b}{b^2}$. Sustituyendo estos valores:

$$\mu_{00} = (1-a) + ab \left[\frac{1-b}{b^2} + \frac{2}{b} \right] = (1-a) + a \frac{1-b}{b} + 2a = 1 + a + \frac{a}{b} - a = 1 + \frac{a}{b}.$$

Concluimos que el tiempo medio de retorno es finito. Por simetría, $\mu_{11} = 1 + \frac{b}{a}$.

Ejemplo 1.35 (Recurrencia nula en la caminata aleatoria simétrica). aleatoria simétrica simple en $\mathbb{Z}$ ($p = 1/2$). Ya sabemos que este estado es recurrente ($\sum p_{00}^{(n)} = \infty$). Veremos ahora que su tiempo medio de retorno es infinito, lo que se conoce como **recurrencia nula**.

Una forma intuitiva de verlo es mediante el siguiente argumento de esperanza condicional. Sea $\mu = \mathbb{E}[\tau_0 \mid X_0 = 0]$ el tiempo medio de retorno. Condicionando en el primer paso:

$$\mu = 1 + \frac{1}{2} \mathbb{E}[\text{tiempo para ir de } 1 \rightarrow 0] + \frac{1}{2} \mathbb{E}[\text{tiempo para ir de } -1 \rightarrow 0].$$

Por simetría, el tiempo esperado para ir de $1 \rightarrow 0$ es igual al de $-1 \rightarrow 0$. Llamemos a este valor x . Entonces $\mu = 1 + x$. Ahora bien, para ir de 1 a 0, damos un paso.

- Con probabilidad $1/2$ vamos a 0 (tardamos 1 paso).
- Con probabilidad $1/2$ vamos a 2. Estando en 2, para llegar a 0 primero debemos llegar a 1 (tiempo esperado x) y luego de 1 a 0 (tiempo esperado x).

Esto nos da la ecuación para x :

$$x = 1 + \frac{1}{2}(0) + \frac{1}{2}(2x) = 1 + x.$$

Obtenemos la ecuación $x = 1 + x$, la cual no tiene solución finita. Esto sugiere que $x = \infty$, y por tanto el tiempo medio de retorno $\mu = 1 + x$ también es infinito.

1.11.1. Análisis del primer paso

El **análisis del primer paso** es una técnica general para calcular tiempos esperados (de visita, de retorno, de absorción, etc.) en cadenas de Markov. La idea es condicionar sobre el valor de X_1 —el primer paso— y usar la propiedad de Markov para traducir el problema en un sistema de ecuaciones lineales.

Método general. Sea $h_i = \mathbb{E}[\tau_j \mid X_0 = i]$ el tiempo esperado para alcanzar por primera vez el estado objetivo j , partiendo de i . Condicionando sobre X_1 :

$$h_i = \mathbb{E}[\tau_j \mid X_0 = i] = \sum_{k \in S} p_{ik} \mathbb{E}[\tau_j \mid X_0 = i, X_1 = k].$$

Por la propiedad de Markov, dado $X_1 = k$, estamos en el estado k y el tiempo restante para llegar a j tiene la misma distribución que τ_j partiendo de k . Como ya gastamos 1 paso, obtenemos:

$$h_i = 1 + \sum_{k \neq j} p_{ik} h_k, \quad i \neq j,$$

con la condición de frontera $h_j = 0$.

Este sistema lineal, uno por cada estado $i \neq j$, determina los tiempos esperados. En particular, para calcular el tiempo medio de retorno ($i = j$), basta notar que desde j damos un paso obligatorio a algún vecino k , de modo que:

$$\mu_{jj} = 1 + \sum_{k \neq j} p_{jk} h_k^{(j)},$$

donde $h_k^{(j)}$ es la solución del sistema anterior con objetivo j .

Ejemplo 1.36 (Cadena de tres estados). Consideremos la cadena con $S = \{0, 1, 2\}$ y matriz de transición

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

Queremos calcular el tiempo medio de retorno al estado 0, es decir, μ_{00} .

Definimos $h_i = \mathbb{E}[\tau_0 \mid X_0 = i]$ con $h_0 = 0$. Las ecuaciones de primer paso para $i = 1$ e $i = 2$ son:

$$\begin{aligned} h_1 &= 1 + p_{10}h_0 + p_{11}h_1 + p_{12}h_2 = 1 + \frac{1}{4}(0) + \frac{1}{2}h_1 + \frac{1}{4}h_2, \\ h_2 &= 1 + p_{21}h_1 + p_{22}h_2 = 1 + \frac{1}{2}h_1 + \frac{1}{2}h_2. \end{aligned}$$

De la segunda ecuación, despejando h_2 :

$$h_2 - \frac{1}{2}h_2 = 1 + \frac{1}{2}h_1 \implies h_2 = 2 + h_1.$$

Sustituyendo en la primera:

$$h_1 = 1 + \frac{1}{2}h_1 + \frac{1}{4}(2 + h_1) = 1 + \frac{1}{2}h_1 + \frac{1}{2} + \frac{1}{4}h_1 = \frac{3}{2} + \frac{3}{4}h_1.$$

Despejando: $\frac{1}{4}h_1 = \frac{3}{2}$, por lo que $h_1 = 6$ y $h_2 = 8$.

Por último, el tiempo medio de retorno desde 0 es:

$$\mu_{00} = 1 + p_{01}h_1 + p_{02}h_2 = 1 + \frac{1}{2}(6) + 0 = 4.$$

La cadena tarda en promedio 4 pasos en regresar al estado 0.

Ejemplo 1.37. Considere una sucesión de lanzamientos de una moneda justa. Estamos interesados en la cantidad de lanzamientos que se necesitan para obtener la secuencia “Sol, Sol, Sol” por primera vez. Modelamos esto como una cadena de Markov con estados que representan el progreso hacia la secuencia objetivo. Cada estado de la cadena medirá el número de éxitos consecutivos obtenidos hasta el momento. El estado 0 representa no haber obtenido ningún éxito, el estado 1 representa haber obtenido un éxito, el estado 2 representa haber obtenido dos éxitos consecutivos, y el estado 3 representa haber obtenido la secuencia completa (tres éxitos consecutivos). El estado 3 es absorbente. Dando lugar a una matriz de transición:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1 \end{pmatrix}.$$

Para calcular el tiempo esperado para alcanzar el estado 3 partiendo de 0, definimos $h_i = \mathbb{E}[\tau_3 \mid X_0 = i]$ con $h_3 = 0$. Las ecuaciones de primer paso son: Resolviendo el sistema no es difícil encontrar que $h_0 = 14$.

1.12. Teorema ergódico

El Teorema Ergódico es el resultado central que describe el comportamiento asintótico de una cadena de Markov irreducible. Conecta el tiempo medio de retorno μ_{jj} con la fracción de tiempo que la cadena pasa en el estado j a largo plazo.

Antes de enunciarlo, recordemos que para una cadena irreducible, ya establecimos que todos los estados son del mismo tipo (recurrentes o transitorios), y que si son recurrentes, la cadena visita cada estado una infinidad de veces. La pregunta natural es: ¿con qué frecuencia visita cada estado?

Teorema 1.4 (Teorema ergódico para cadenas de Markov). *Sea $\{X_n\}$ una cadena de Markov irreducible. Para cualesquiera estados $i, j \in S$ se cumple*

$$\lim_{n \rightarrow \infty} \frac{N_{ij}(n)}{n} = \frac{1}{\mu_{jj}} \quad \text{c.s.,}$$

donde $N_{ij}(n) = \sum_{k=1}^n \mathbf{1}_{\{X_k=j\}}$ es el número de visitas a j hasta el tiempo n , y $\mu_{jj} = \mathbb{E}[\tau_j \mid X_0 = j]$ es el tiempo medio de retorno. Si $\mu_{jj} = \infty$, el límite es 0.

Demostración. Caso transitorio. Si la cadena es transitoria, entonces $N_{ij}(n) < \infty$ casi seguramente para todo i, j , por lo que $N_{ij}(n)/n \rightarrow 0$ casi seguramente. Además, $\mu_{jj} = \infty$ en este caso, de modo que ambos lados son iguales a cero.

Caso recurrente. Supongamos que la cadena es recurrente. Sea $\tau_{ij} = \min\{n \geq 1 : X_n = j \mid X_0 = i\}$ el tiempo de primera visita de i a j . Por irreducibilidad y recurrencia, $\mathbb{P}(\tau_{ij} < \infty) = 1$.

Notemos que para $n \geq 1$ se cumple la identidad

$$N_{ij}(n) = 1 + N_{jj}(n - \tau_{ij}),$$

es decir, el número de visitas que hace a j partiendo desde i hasta el tiempo n es igual al número de retornos que hace a j partiendo desde j hasta el tiempo $n - \tau_{ij}$ más la primera visita, pues la primera visita tarda exactamente τ_{ij} pasos. Por lo tanto, es suficiente demostrar la convergencia para $N_{jj}(n)/n$, ya que

$$\frac{N_{ij}(n)}{n} = \frac{N_{jj}(n - \tau_{ij})}{n - \tau_{ij}} \cdot \frac{n - \tau_{ij}}{n} \xrightarrow{n \rightarrow \infty} \frac{1}{\mu_{jj}} \cdot 1,$$

donde en el último paso usamos que τ_{ij} es finito c.s. y la convergencia de $N_{jj}(n)/n$ que demostraremos a continuación.

Convergencia de $N_{jj}(n)/n$. Sea $Y^{(k)}$ la longitud del k -ésimo intervalo entre visitas consecutivas al estado j , es decir, el tiempo transcurrido entre la visita $(k-1)$ -ésima y la k -ésima. Por la propiedad fuerte de Markov (aplicada en cada tiempo de visita), las variables $Y^{(1)}, Y^{(2)}, \dots$ son independientes e idénticamente distribuidas con media μ_{jj} .

Sea $N = N_{jj}(n)$ el número de visitas hasta el tiempo n . Entonces la suma de los primeros N intervalos no excede n , mientras que la de los primeros $N+1$ sí lo hace:

$$Y^{(1)} + \dots + Y^{(N)} \leq n < Y^{(1)} + \dots + Y^{(N+1)}.$$

Dividiendo entre N :

$$\frac{Y^{(1)} + \dots + Y^{(N)}}{N} \leq \frac{n}{N} < \frac{Y^{(1)} + \dots + Y^{(N+1)}}{N}.$$

Como la cadena es recurrente, $N = N_{jj}(n) \rightarrow \infty$ cuando $n \rightarrow \infty$. Por la **Ley de los Grandes Números**, ambos extremos de la desigualdad convergen a μ_{jj} casi seguramente. Por el teorema del sandwich, concluimos que $n/N \rightarrow \mu_{jj}$ c.s., es decir,

$$\lim_{n \rightarrow \infty} \frac{N_{jj}(n)}{n} = \frac{1}{\mu_{jj}} \quad \text{c.s.}$$

□

Observación 1.16 (Interpretación). El número $\pi_j = 1/\mu_{jj}$ representa la **fracción de tiempo** que la cadena pasa en el estado j a largo plazo. Si $\mu_{jj} < \infty$, $\pi_j > 0$ y el estado es visitado con frecuencia positiva. Si $\mu_{jj} = \infty$, $\pi_j = 0$. Note que $\mu_{jj} = \infty$ no implica que el estado j sea transitorio. Más aún, no significa que no sea visitado, sino que es visitado con frecuencia cada vez menor a medida que pasa el tiempo. Este fenómeno se conoce como **recurrencia nula**.

Tomando esperanza en el límite y usando el Teorema de Convergencia Dominada, para una cadena irreducible se cumple que

$$\frac{1}{\mu_{jj}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)}.$$

Este es el resultado que relaciona el tiempo medio de retorno con el promedio de las probabilidades de transición, y es la base de la teoría de distribuciones estacionarias que estudiaremos en secciones siguientes.

Ejemplo 1.38. Preferencias musicales a largo plazo. Una persona alterna entre tres géneros musicales: Pop (P), Rock (R) y Jazz (J), con preferencias modeladas por la cadena de Markov:

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0.5 & 0 & 0.5 \end{pmatrix}.$$

Estamos interesados en calcular qué proporción de tiempo, a largo plazo, la persona escuchará Pop.

Sea μ_{PP} el tiempo medio de retorno a Pop. Definimos:

- h_R : tiempo esperado para llegar a P empezando en R
- h_J : tiempo esperado para llegar a P empezando en J

Por análisis de primer paso:

$$h_R = 1 + 0.5 h_R + 0.25 h_J$$

$$h_J = 1 + 0.5 h_J$$

De la segunda ecuación:

$$h_J = 2$$

Sustituyendo en la primera:

$$h_R = 1 + 0.5h_R + 0.25(2) = 1 + 0.5h_R + 0.5$$

$$0.5h_R = 1.5 \Rightarrow h_R = 3$$

Ahora, desde P:

$$\mu_{PP} = 1 + 0.5h_R + 0 \cdot h_J = 1 + 0.5(3) = 2.5 = \frac{5}{2}$$

Aplicando el Teorema Ergódico:

$$\pi_P = \frac{1}{\mu_{PP}} = \frac{2}{5} = 0.4$$

Por lo tanto, a largo plazo la persona escuchará Pop aproximadamente el 40 % del tiempo.

Ejemplo 1.39. Tiempo medio de retorno en la cadena de rachas. Consideremos la cadena de rachas con espacio de estados $S = \{0, 1, 2, \dots\}$, donde j representa una racha actual de j éxitos consecutivos, con probabilidades de transición $p_{j,j+1} = p$ y $p_{j,0} = q := 1 - p$. Queremos calcular $\mu_{jj} = 1/\pi_j$ vía el Teorema Ergódico; para ello calculamos $\lim_{n \rightarrow \infty} p_{ij}^{(n)}$.

Fijado el estado objetivo j , aplicamos convenientemente Chapman-Kolmogorov separando toda trayectoria de longitud m en un primer tramo de $m - j$ pasos y un segundo tramo de j pasos:

$$p_{ij}^{(m)} = \sum_{k \in S} p_{ik}^{(m-j)} p_{kj}^{(j)}.$$

Observemos que $p_{kj}^{(j)} > 0$ únicamente cuando $k = 0$: la única manera de alcanzar j en exactamente j pasos es partir de 0 y acumular j éxitos consecutivos, lo cual ocurre con probabilidad p^j . Para cualquier otro k , o bien la trayectoria es geoméricamente imposible en j pasos, o bien requiere pasar por j antes de completarlos. Por tanto, en el límite $m \rightarrow \infty$ solo sobrevive ese término:

$$\lim_{m \rightarrow \infty} p_{ij}^{(m)} = \lim_{m \rightarrow \infty} p_{i,0}^{(m-j)} \cdot p^j.$$

Ahora bien, hemos demostrado previamente que $p_{i,0}^{(r)} = q$ para todo $r \geq 1$ y todo estado inicial i , por lo que $\lim_{m \rightarrow \infty} p_{i,0}^{(m-j)} = q$. De modo que

$$\lim_{m \rightarrow \infty} p_{ij}^{(m)} = qp^j.$$

Dado que esta sucesión converge, converge también en el sentido de Cesàro (pues la convergencia ordinaria implica la convergencia de las medias aritméticas), y por el Teorema Ergódico:

$$\frac{1}{\mu_{jj}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n p_{ij}^{(m)} = \frac{1}{qp^j}.$$

Para ilustrar la magnitud de este resultado, tomemos $p = 1/2$ y $j = 5000$:

$$\mu_{5000,5000} = \frac{1}{q p^{5000}} = \frac{1}{\frac{1}{2} \cdot \left(\frac{1}{2}\right)^{5000}} = 2^{5001}.$$

Este número posee más de 1500 dígitos decimales y supera con creces la edad del universo medida en nanosegundos ($\approx 4 \times 10^{26}$). Sin embargo, el Teorema Ergódico garantiza que el estado 5000 es visitado infinitas veces con probabilidad 1; simplemente, la espera entre visitas consecutivas es astronómica.

Ejemplo 1.40. El mono y Shakespeare. Un mono teclea caracteres de manera independiente y uniforme sobre un teclado de m teclas. Deseamos estimar el tiempo esperado para que produzca la obra completa de Shakespeare, de longitud N caracteres.

El proceso puede modelarse con la cadena de rachas: el estado $j \in \{0, 1, \dots, N\}$ representa haber tecleado correctamente los últimos j caracteres consecutivos de la obra. Las probabilidades de transición son $p_{j,j+1} = p := 1/m$ (acierta el siguiente carácter) y $p_{j,0} = q := 1 - 1/m$ (falla), con estado absorbente N una vez completada la obra. Para el estado $j = N - 1$, la transición exitosa lleva a N también con probabilidad p .

Aproximamos $h_0 := \mathbb{E}_0[\tau_N]$ usando el análisis del primer paso junto con $\mu_{N,N}$. Por el resultado del ejemplo anterior, el tiempo medio de retorno al estado N es $\mu_{N,N} = 1/(q p^N)$. Ahora bien, por análisis del primer paso desde el estado $N - 1$:

$$h_{N-1} = 1 + p \cdot 0 + q h_0 \implies h_{N-1} = 1 + q h_0.$$

Dado que $\mu_{N,N} = h_{N-1} + (\text{tiempo de } N \text{ a } N - 1) \approx h_0$ cuando N es grande (pues el tiempo para ir de N a $N - 1$ es despreciable frente al tiempo de reconstruir la racha), concluimos:

$$\mathbb{E}_0[\tau_N] \approx \mu_{N,N} = \frac{1}{q p^N} = \frac{m^N}{1 - 1/m}.$$

Para $m = 50$ teclas y $N = 5 \times 10^6$ caracteres:

$$\mathbb{E}_0[\tau_N] \approx \frac{50^{5 \times 10^6}}{1 - 1/50} = \frac{50}{49} \cdot 50^{5 \times 10^6},$$

un número con más de 8.5 millones de dígitos decimales. A pesar de esto, el Teorema Ergódico garantiza que el mono producirá la obra con probabilidad 1 si teclea indefinidamente.

1.13. Recurrencia positiva y nula

Hemos visto que si una cadena de Markov inicia en un estado recurrente, entonces regresa a él una infinidad de veces con probabilidad uno. Sin embargo, esta recurrencia puede presentarse de dos formas: cuando el tiempo promedio de retorno es finito, o cuando es infinito. Esto lleva a la noción de **recurrencia positiva** y **recurrencia nula**, respectivamente.

Consideremos un estado recurrente j . Recordemos que el tiempo de primera visita a j a partir de cualquier estado i es la variable aleatoria

$$\tau_{ij} := \min\{n \geq 1 : X_n = j \mid X_0 = i\}.$$

La esperanza de esta variable aleatoria es el **tiempo medio de recurrencia**.

Definición 1.17. El tiempo medio de recurrencia del estado recurrente j a partir del estado i se define como la esperanza de τ_{ij} , y se denota por μ_{ij} :

$$\mu_{ij} := \mathbb{E}[\tau_{ij}] = \sum_{n=1}^{\infty} n f_{ij}^{(n)},$$

donde $f_{ij}^{(n)} = P(X_n = j, X_k \neq j, 1 \leq k < n \mid X_0 = i)$ es la probabilidad de visitar j por primera vez en el instante n . Decimos que j es **recurrente positivo** si $\mu_{jj} < \infty$, y **recurrente nulo** si $\mu_{jj} = \infty$.

La intuición es la siguiente: en ambos casos la cadena regresa a j infinitas veces con probabilidad 1, pero el ritmo al que lo hace difiere cualitativamente. Si j es recurrente positivo, las visitas sucesivas están separadas por tiempos con media finita y la fracción de tiempo que la cadena pasa en j es positiva ($\pi_j = 1/\mu_{jj} > 0$). Si j es recurrente nulo, los retornos se vuelven cada vez más espaciados y esa fracción se diluye hasta cero ($\pi_j = 0$), aunque la cadena siga regresando siempre.

Proposición 1.13. La recurrencia positiva y la recurrencia nula son propiedades de clase: si $i \leftrightarrow j$ y i es recurrente positivo (respectivamente recurrente nulo), entonces j también lo es.

Demostración. Supongamos que $i \leftrightarrow j$. Entonces existen enteros $r, s \geq 1$ tales que $p_{ij}^{(r)} > 0$ y $p_{ji}^{(s)} > 0$. Por Chapman-Kolmogorov:

$$p_{jj}^{(r+n+s)} \geq p_{ji}^{(s)} p_{ii}^{(n)} p_{ij}^{(r)}.$$

Sumando sobre n e invocando el Teorema Ergódico, si $\pi_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ii}^{(k)} > 0$, entonces

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{jj}^{(k)} \geq p_{ji}^{(s)} \pi_i p_{ij}^{(r)} > 0,$$

de modo que $\mu_{jj} = 1/\pi_j < \infty$ y j es recurrente positivo. El argumento es simétrico intercambiando i y j , de modo que i es recurrente positivo si y solo si j lo es. El caso nulo se obtiene por contraposición. $\square$

Ejemplo 1.41. Caminata aleatoria simple simétrica: recurrencia nula. En el ejemplo de recurrencia nula de la caminata aleatoria simétrica (véase el Ejemplo 1.35), demostramos que el tiempo medio de retorno al origen satisface la ecuación $\mu_{00} = 1 + x$ con $x = 1 + x$, que no tiene solución finita, luego $\mu_{00} = \infty$. Como la cadena es irreducible (todos los estados se comunican), la recurrencia positiva y la nula son propiedades de clase, y basta con este cálculo en un único estado para concluir que todos los estados de la caminata aleatoria simétrica son recurrentes nulos. Intuitivamente, la cadena regresa siempre al origen, pero lo hace cada vez más tarde: cuanto más tiempo ha transcurrido, más lejos se ha alejado en promedio y mayor es el tiempo que necesitará para regresar, de modo que el promedio de los tiempos de retorno diverge.

Ejemplo 1.42. Cadena de rachas: recurrencia positiva. En el Ejemplo de tiempo medio de retorno en la cadena de rachas demostramos que $\mu_{jj} = 1/(qp^j) < \infty$ para todo estado $j \geq 0$. En particular $\mu_{00} = 1/q < \infty$, de modo que el estado 0 es recurrente positivo. Al ser la cadena irreducible, todos sus estados son recurrentes positivos.

Observación 1.17. Todo estado de una cadena de Markov pertenece exactamente a una de las siguientes tres categorías: **transitorio**, **recurrente positivo** o **recurrente nulo**. Esta tricotomía es exhaustiva y las categorías son mutuamente excluyentes. Dado que la recurrencia, la recurrencia positiva y la recurrencia nula son todas propiedades de clase, la clasificación es constante dentro de cada clase de comunicación.

Recordemos que en una cadena de Markov con espacio de estados finito hemos demostrado que debe existir al menos un estado recurrente. Este hecho, combinado con la finitud del espacio de estados, impone una restricción adicional sobre el tipo de recurrencia posible.

Proposición 1.14. En una cadena de Markov con espacio de estados finito, no existen estados recurrentes nulos.

Demostración. Sea j un estado recurrente y consideremos el conjunto C de estados que se comunican con j . Intuitivamente, una vez que la cadena entra en C , no puede escapar de él: en efecto, si desde algún estado de C se pudiera alcanzar un estado fuera de C , entonces ese estado también tendría que ser recurrente, lo cual contradice la definición de C .

En términos de probabilidades de transición, esto implica que para todo $i \in C$ y todo $n \geq 1$,

$$\sum_{k \in C} p_{ik}^{(n)} = 1.$$

Promediando,

$$\frac{1}{n} \sum_{m=1}^n \sum_{k \in C} p_{ik}^{(m)} = 1,$$

y como C es finito, podemos intercambiar suma y promedio:

$$\sum_{k \in C} \left(\frac{1}{n} \sum_{m=1}^n p_{ik}^{(m)} \right) = 1.$$

Haciendo $n \rightarrow \infty$ y usando el Teorema Ergódico,

$$\sum_{k \in C} \frac{1}{\mu_k} = 1.$$

En particular, esta suma no puede ser nula, por lo que debe existir al menos un estado $k \in C$ tal que $\mu_k < \infty$, es decir, k es recurrente positivo. Como la recurrencia positiva es una propiedad compartida entre estados que se comunican, se concluye que todo estado en C , y en particular j , es recurrente positivo. $\square$

Ejemplo 1.43. Cadena de 5 estados con dos clases de comunicación. Consideremos la cadena con $S = \{1, 2, 3, 4, 5\}$ y matriz de transición

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Las clases de comunicación son:

- $C_1 = \{1, 2\}$: los estados 1 y 2 se comunican entre sí, pero desde el estado 3 se puede llegar a C_1 sin que C_1 pueda llegar a 3; en realidad desde 1 o 2 no se llega a 3. Inspeccionando: $p_{13} = 1/2 > 0$ pero $p_{31} = 1/4 > 0$ y $p_{12} = 1/2 > 0$, $p_{21} = 1/2 > 0$, $p_{23} = 1/2 > 0$. Entonces $1 \rightarrow 3 \rightarrow 4$ pero $4 \not\rightarrow 1$. Así $C_1 = \{1, 2\}$ es una clase transitoria.
- $C_2 = \{4, 5\}$: $p_{45} = 1/2$, $p_{54} = 1/2$, $p_{44} = 1/2$, $p_{55} = 1/2$; no hay salida de $\{4, 5\}$ pues $p_{4k} = 0$ para $k \notin \{4, 5\}$. Es una clase recurrente (positiva, por ser finita).
- El estado 3 es transitorio: desde 3 se puede ir a $\{1, 2\}$ o a $\{4, 5\}$ con probabilidad positiva, pero no hay retorno desde ninguno de esos conjuntos a 3.

Tiempos de retorno en C_2 . Desde el estado 4 la cadena va a 4 o a 5 con igual probabilidad $1/2$, y desde 5 ocurre exactamente lo mismo. Por simetría total, a largo plazo la cadena no tiene motivo alguno para preferir un estado sobre el otro, de modo que pasa la misma fracción de tiempo en 4 que en 5. Como ambas fracciones deben sumar 1, cada una vale $\pi_4 = \pi_5 = 1/2$. Por el Teorema Ergódico:

$$\mu_{44} = \frac{1}{\pi_4} = 2, \quad \mu_{55} = \frac{1}{\pi_5} = 2.$$

En promedio, partiendo del estado 4 (o del 5), la cadena regresa a ese mismo estado tras 2 pasos.

1.14. Evolución de la distribución de X_n

Hasta ahora hemos estudiado el comportamiento de la cadena desde un estado inicial fijo: la probabilidad de estar en j en el paso n partiendo de i es $p_{ij}^{(n)}$. Sin embargo, en muchas situaciones el estado inicial no se conoce con certeza sino que está determinado por una distribución de probabilidad. Nos interesa entonces describir cómo evoluciona esa distribución con el tiempo y; en particular, si es que, a largo plazo, converge y hacia qué límite.

Denotamos por $\pi^{(n)}$ el vector fila de distribución de X_n , es decir,

$$\pi_j^{(n)} := \mathbb{P}(X_n = j), \quad j \in S.$$

En particular, $\pi^{(0)}$ es la **distribución inicial** de la cadena. Aprovechamos aquí para enmendar un abuso de notación que hemos cometido en secciones anteriores: definimos $\pi_j := 1/\mu_{jj}$ como la proporción de tiempo asintótica que la cadena pasa en j . Para evitar confusión con la distribución de X_n , reservamos de ahora en adelante la notación $\pi^{(0)}$ para la distribución inicial y escribimos π (sin supraíndice) para el vector de proporciones asintóticas $(\pi_j)_{j \in S}$.

Establecer una recurrencia para la evolución de la distribución es simple: por la ley de la probabilidad total,

$$\pi_j^{(n)} = \mathbb{P}(X_n = j) = \sum_{i \in S} \mathbb{P}(X_{n-1} = i) p_{ij} = \sum_{i \in S} \pi_i^{(n-1)} p_{ij},$$

que en notación matricial se escribe

$$\pi^{(n)} = \pi^{(n-1)} P.$$

Iterando, $\pi^{(n)} = \pi^{(0)} P^n$.

Ejemplo 1.44 (Convergencia de la distribución). *Consideremos la cadena de dos estados $S = \{0, 1\}$ con matriz*

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

y distribución inicial $\pi^{(0)} = (1, 0)$ (la cadena comienza en 0 con certeza). Las primeras iteraciones dan:

$$\begin{aligned} \pi^{(1)} &= (0.7, 0.3), \\ \pi^{(2)} &= (0.7 \cdot 0.7 + 0.3 \cdot 0.4, 0.7 \cdot 0.3 + 0.3 \cdot 0.6) = (0.61, 0.39), \\ \pi^{(3)} &\approx (0.574, 0.426), \\ \pi^{(10)} &\approx (0.572, 0.428). \end{aligned}$$

La sucesión $(\pi^{(n)})$ parece estabilizarse alrededor de $(4/7, 3/7) \approx (0.571, 0.429)$, independientemente del punto de partida.

Ejemplo 1.45 (Distribución que oscila). *Consideremos la cadena en $S = \{0, 1\}$ con matriz*

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

y distribución inicial $\pi^{(0)} = (1, 0)$. Entonces:

$$\pi^{(1)} = (0, 1), \quad \pi^{(2)} = (1, 0), \quad \pi^{(3)} = (0, 1), \quad \dots$$

La distribución oscila indefinidamente entre $(1, 0)$ y $(0, 1)$: la cadena es periódica de período 2 y la sucesión $(\pi^{(n)})$ no converge.

Estos ejemplos muestran que la convergencia de $\pi^{(n)}$ no está garantizada en general. Nuestro objetivo en lo que sigue es entender cuándo y hacia qué límite converge esta distribución. El punto de partida natural es identificar qué distribuciones son *invariantes* bajo la acción de P , es decir, distribuciones que, una vez alcanzadas, ya no cambian.

Definición 1.18. Un vector de probabilidad $\pi = (\pi_j)_{j \in S}$ (con $\pi_j \geq 0$ y $\sum_j \pi_j = 1$) se llama **distribución invariante** o **distribución estacionaria** si

$$\pi = \pi P,$$

es decir, $\pi_j = \sum_{i \in S} \pi_i p_{ij}$ para todo $j \in S$. Si la cadena se inicia con $\pi^{(0)} = \pi$, entonces $\pi^{(n)} = \pi$ para todo $n \geq 0$.