CIND-221: Cadenas de Markov

Felipe Osorio

f.osoriosalgado@uandresbello.edu

Facultad de Ingeniería, UNAB

Ejemplo (Predicción del tiempo):

Suponga que la probabilidad de que llueva mañana depende del clima previo solo a través de si **hoy** está o no lloviendo, y **no** de las condiciones del tiempo en días anteriores.

Suponga que si hoy está lloviendo, entonces lloverá mañana con probabilidad a; y si no está lloviendo hoy, entonces mañana lloverá con probabilidad b.

De este modo, el proceso está en el estado 0 cuando llueve y 1 cuando no llueve, lo que lleva a una cadena de Markov con 2 estados $(S = \{0, 1\})$ con probabilidades de transición dadas por:

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

Ejemplo:

Sea $\{X_n : n \geq 0\}$ una cadena de Markov con estados $S = \{0, 1, 2\}$ y matriz de probabilidades de transición

$$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 & 0.00 \\ 0.25 & 0.50 & 0.25 \\ 0.00 & 0.75 & 0.25 \end{pmatrix},$$

con distribución inicial $\pi_i = \mathsf{P}(X_0 = i) = \frac{1}{3}$, para $i = 0, 1, 2$.

Entonces,

$$\begin{aligned} p_{21} &= \mathsf{P}(X_1 = 1 | X_0 = 2) = 0.75, \\ p_{12} &= \mathsf{P}(X_2 = 2 | X_1 = 1) = 0.25. \end{aligned}$$

Además,

$$\begin{aligned} \mathsf{P}(X_2 = 2, X_1 = 1 | X_0 = 2) &= \mathsf{P}(X_2 = 2 | X_1 = 1)\, \mathsf{P}(X_1 = 2 | X_0 = 1) \\ &= 0.25 \cdot 0.75 = \frac{3}{16}. \end{aligned}$$

Resultado 1 (Ecuaciones de Chapman-Kolmogorov):

Si la secuencia de variables aleatorias $\{X_n : n \geq 0\}$ es una cadena de Markov y si $k < m < n$, entonces tenemos que para todo $h, j \in S$,

$$\mathsf{P}(X_n = j | X_k = h) = \sum_{i \in S} \mathsf{P}(X_n = j | X_m = i)\, \mathsf{P}(X_m = i | X_k = k).$$

Definición 1 (Probabilidad de transición en m-pasos):

Sea $\{X_n : n \in \mathbb{N}\}$ una cadena de Markov, la probabilidad de transición en m-pasos desde el estado i al j, es dada por:

$$p_{ij}^{(m)} = \mathsf{P}(X_{n+m} = j | X_n = i).$$

Además,

$$p_{ij}^{(0)} = \mathsf{P}(X_n = j | X_n = i) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}.$$

Observación:

La probabilidad $p_{ij}^{(m)}$ es estacionaria si y solo si,

$$p_{ij}^{(m)} = \mathsf{P}(X_{n+m} = j | X_n = i) = \mathsf{P}(X_m = j | X_0 = i).$$

Observación:

Las probabilidades de transición de m-pasos, pueden ser escritas en la matriz de transición:[1]

$$\boldsymbol{P}^{(m)} = (p_{ij}^{(m)}).$$

[1] Evidentemente, $\boldsymbol{P}^{(0)} = \boldsymbol{I}$.

Definición 2 (Cadena de Markov homogénea):

Una cadena de Markov cuyas probabilidades de transición en m pasos son todas estacionarias es llamada cadena de Markov homogénea.

Proposición 1:

Las ecuaciones de Chapman-Kolmogorov proveen un procedimiento para calcular las probabilidades de transición de n-pasos. En efecto, sigue que:

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)},$$

para todo $n, m \geq 0$ y todo $i, j \in S$.

Observación:

En forma matricial la Proposición 1 puede ser escrita como:

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

donde $\cdot$ denota multiplicación matricial.

Note la condición inicial,

$$\boldsymbol{P}^{(2)} = \boldsymbol{P}^{(1+1)} = \boldsymbol{P} \cdot \boldsymbol{P} = \boldsymbol{P}^2,$$

continuando por inducción tenemos que la matriz de transición de n-pasos puede ser escrita como:

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

Corolario 1 (Matriz de transición de n-pasos):

Sea $X_0, X_1, \ldots$ una cadena de Markov con matriz de transición $\boldsymbol{P}$. La matriz de transición de n-pasos de la cadena es dada por:

$$\boldsymbol{P}^{(n)} = \boldsymbol{P}^n, \qquad n \geq 0,$$

con $p_{ij}^{(n)} = \mathsf{P}(X_n = j | X_0 = i)$, $\forall\, i, j$.

Ejemplo:

Considere el ejemplo de predicción del tiempo con $a = 0.7$ y $b = 0.4$. Se desea calcular la probabilidad de que llueva después de 4 días a partir de hoy, dado que hoy está lloviendo. Tenemos

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

(...continuación)

En efecto, $\boldsymbol{P}^{(2)} = \boldsymbol{P}^2$, con

$$\boldsymbol{P}^2 = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}.$$

De ahí que, $\boldsymbol{P}^{(4)} = (\boldsymbol{P}^2)^2$,

$$\boldsymbol{P}^4 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix} \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix} = \begin{pmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix}.$$

Es decir, la probabilidad deseada es $p_{00}^{(4)} = 0.5749$.

Ejercicio:

Considere la matriz de transición

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad q = 1 - p.$$

Obtenga $\boldsymbol{P}^3$ y verifique que es matriz de transición.

Ejercicio:

Considere una cadena de Markov con matriz de transición

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

donde $p, q \in (0, 1)$ y $p + q > 0$. Determine $\boldsymbol{P}^n$.

Ejercicio:

Considere

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

verifique (por inducción) que la n-ésima potencia de $\boldsymbol{P}$ es dada por:

$$\boldsymbol{P}^n = \begin{pmatrix} \frac{1}{2^n} & 0 & 1 - \frac{1}{2^n} \\ 0 & \frac{1}{2^n} & 1 - \frac{1}{2^n} \\ 0 & 0 & 1 \end{pmatrix}.$$

¿Qué sucede cuando $\lim_{n\to\infty} \boldsymbol{P}^n$?

Observación:

Existe diversos métodos prácticos[2] para obtener $\boldsymbol{P}^n$.

Sea $\boldsymbol{A} \in \mathbb{R}^{k\times k}$, un método para obtener $f(\boldsymbol{A})$ (útil para $\boldsymbol{A}$ matriz simétrica) está basado en la descomposición espectral:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1},$$

con $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1,\dots,\lambda_k)$. De este modo,

$$f(\boldsymbol{A}) = f(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}) = \boldsymbol{U}f(\boldsymbol{\Lambda})\boldsymbol{U}^{-1},$$

y debido a la forma de $\boldsymbol{\Lambda}$, permite un cálculo simple de $f(\boldsymbol{\Lambda})$.

Para calcular $\boldsymbol{A}^n$, podemos hacer:[3]

$$\boldsymbol{A}^n = \boldsymbol{U}\boldsymbol{\Lambda}^n\boldsymbol{U}^{-1},$$

donde $\boldsymbol{\Lambda}^n = \operatorname{diag}(\lambda_1^n,\dots,\lambda_k^n)$.

[2] Higham, N.J. (2008). Functions of Matrices: Theory and Computation. SIAM.

[3] Aunque este método no puede ser recomendado, pues puede ser muy impreciso.

Ejercicio:

Para la matriz

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Verifique que sus valores propios y vectores propios asociados son:

$$\lambda_1 = 1, \quad \boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \lambda_2 = 1-p-q, \quad \boldsymbol{u}_2 = \begin{pmatrix} -p \\ q \end{pmatrix}.$$

Así,

$$\boldsymbol{U} = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}, \qquad \boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}.$$

Obtenga $\boldsymbol{P}^n$.

Definición 3 (Distribución de X_n):

Sea $\{X_n : n \geq 0\}$ una cadena de Markov con matriz de transición $\boldsymbol{P}$ y distribución inicial $\boldsymbol{\pi}$. Para todo $n \geq 0$, la distribución de X_n es dada por $\boldsymbol{\pi}^\top \boldsymbol{P}^n$. Esto es:

$$\mathsf{P}(X_n = k) = \sum_j p_{jk}^{(n)} \pi_j, \qquad \forall\, k.$$

Demostración:

Note que

$$\begin{aligned}
\mathsf{P}(X_n = k) &= \sum_j \mathsf{P}(X_{n+1} = k, X_n = j) \\
&= \sum_j \mathsf{P}(X_{n+1} = k | X_n = j)\, \mathsf{P}(X_n = j) \\
&= \sum_j p_{jk}^{(n)}\, \mathsf{P}(X_n = j) \\
&= \sum_j p_{jk}^{(n)} \pi_j.
\end{aligned}$$

Ejemplo:

Para el ejemplo de predicción del tiempo, si $\pi_0 = 0.4$, $\pi_1 = 0.6$. Tenemos que la probabilidad de que llueva en el 4º día, despues de que comenzamos a llevar los registros es:

$$\begin{aligned}\mathsf{P}(X_4 = 0) &= 0.4\, p_{00}^{(4)} + 0.6\, p_{10}^{(4)} = 0.4 \cdot 0.5749 + 0.6 \cdot 0.5668 \\ &= 0.57004\end{aligned}$$

En efecto, $\boldsymbol{\pi}^\top \boldsymbol{P}^{(4)} = \boldsymbol{\pi}^\top \boldsymbol{P}^4$, lo que lleva a:

$$\boldsymbol{\pi}^\top \boldsymbol{P}^4 = (0.4, 0.6) \begin{pmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix} = (0.57004, 0.42996).$$

Es decir,

$$\mathsf{P}(X_4 = 0) = 0.57004, \qquad \mathsf{P}(X_4 = 1) = 0.42996.$$

Un aspecto clave en el estudio de cadenas de Markov, es el análisis de su comportamiento asintótico. Primeramente, debemos clasificar los estados de la cadena, a saber:

Definición 4 (Accesibilidad):

Un estado j se dice accesible desde el estado i si $p_{ij}^{(n)} > 0$ para algún $n \geq 0$ y escribimos $i \to j$.

Observación:

La relación "$\to$" es transitiva, es decir:

$$i \to j, \text{ y } j \to k \quad \Rightarrow \quad i \to k.$$

(esto se verifica usando las ecuaciones de Chapman-Kolmogorov).

Definición 5 (Comunicación de estados):

Los estados i y j se comunican si $i \to j$ y $j \to i$ lo que escribimos como $i \leftrightarrow j$.

Observación:

Note que cualquier estado se comunica consigo mismo, dado que por definición:

$$p_{ii}^{(0)} = \mathsf{P}(X_0 = i | X_0 = i) = 1.$$

Además, se puede verificar que "$i \leftrightarrow j$" es una relación de equivalencia sobre S, y de este modo las clases:[4]

$$C(i) := \{j \in S : i \leftrightarrow j\}, \qquad i \in S,$$

forman una partición de S.

[4] Dos estados que comunican se dice que están en la misma clase.

Ejemplo:

Considere una cadena de Markov con 3 estados $S = \{0, 1, 2\}$ y matriz de transición

$$\boldsymbol{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

Por ejemplo, es posible ir del estado 0 al 2. En efecto,

$$0 \to 1 \to 2$$

Es decir, podemos ir del estado 0 al 1, $0 \to 1$ con probabilidad $p_{01} = 1/2$ e ir del estado 1 al 2, $1 \to 2$ con probabilidad $p_{12} = 1/4$.

Definición 6 (Cadena irreducible):

Una cadena de Markov se dice irreducible si el espacio de estados consiste sólo de una clase, esto es, todos los estados se comunican entre si.

Definición 7 (Estado absorvente):

Un estado i se dice absorvente si $p_{ii} = 1$, o equivalentemente $p_{ij} = 0$ para todo $j \neq i$.

Ejemplo:

Considere una cadena de Markov con 4 estados $S = \{0, 1, 2, 3\}$ y matriz de transición

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

cuyas clases son $\{0, 1\}$ y $\{3\}$ (verificar). Note que el estado 3 es absorvente.

Definición 8 (Distribución límite):

Sea $\{X_n : n \geq 0\}$ una cadena de Markov con matriz de transición $\boldsymbol{P}$. Una distribución límite para la cadena de Markov es una distribución de probabilidades $\boldsymbol{\lambda}$ con la propiedad

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lambda_j, \qquad \forall\, i, j.$$

Observación:

La definición anterior es equivalente con las siguientes:

(i) Para cualquier distribución inicial y todo j,

$$\lim_{n\to\infty} \mathsf{P}(X_n = j) = \lambda_j.$$

(ii) Para cualquier distribución inicial,

$$\lim_{n\to\infty} \boldsymbol{\pi}^\top \boldsymbol{P}^n = \boldsymbol{\lambda}^\top.$$

(iii)

$$\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{\Lambda},$$

donde $\boldsymbol{\Lambda}$ es una matriz de probabilidades cuyas filas son todas igual a $\boldsymbol{\lambda}^\top$.

Ejemplo:

En un ejercicio anterior se debe obtener

$$\boldsymbol{P}^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}.$$

Como $|1-p-q| < 1$ $(p, q \in (0,1))$. Así,

$$\lim_{n\to\infty} \boldsymbol{P}^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} = \boldsymbol{\Lambda}.$$

De ahí que la distribución límite es dada por

$$\boldsymbol{\lambda}^\top = \Big(\frac{q}{p+q}, \frac{p}{p+q}\Big).$$

Definición 9 (Distribución estacionaria):

Sea $X_0, X_1, \ldots$ una cadena de Markov con matriz de transición $\boldsymbol{P}$. Una distribución estacionaria es una distribución de probabilidades $\boldsymbol{\pi}$ que satisface

$$\boldsymbol{\pi}^\top \boldsymbol{P} = \boldsymbol{\pi}^\top.$$

Esto es

$$\pi_j = \sum_i \pi_i p_{ij}, \qquad \forall\, j$$

Lema 1:

Suponga $\boldsymbol{\pi}$ la distribución límite de una cadena de Markov con matriz de transición $\boldsymbol{P}$, entonces $\boldsymbol{\pi}$ es una distribución estacionaria.

Observación:

Lo contrario del Lema anterior no es verdad, en general distribuciones estacionarias no necesariamente son distribuciones límite.

Definición 10 (Matriz de transición regular):

Una matriz de transición $\boldsymbol{P}$ se dice regular si para alguna potencia positiva de $\boldsymbol{P}$, digamos $\boldsymbol{P}^n$ tiene todas sus entradas positivas, $(n \geq 1)$.

Ejemplo:

Considere

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

es regular, pues

$$\boldsymbol{P}^4 = \begin{pmatrix} 9/16 & 5/16 & 2/16 \\ 2/8 & 3/8 & 3/8 \\ 8/16 & 5/16 & 3/16 \end{pmatrix},$$

tiene todas sus entradas positivas.

Proposición 2:

Una cadena de Markov cuya matriz de transición $\boldsymbol{P}$ es regular, tiene una distribución límite, la que es única, positiva, estacionaria. Esto es, existe un único vector de probabilidades $\boldsymbol{\pi} > 0$, tal que

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j, \qquad \forall\, i, j,$$

donde

$$\sum_i \pi_i p_{ij} = \pi_j.$$

Equivalentemente, existe una matriz estocástica $\boldsymbol{\Pi}$ con todos sus elementos positivos, tal que

$$\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{\Pi},$$

donde $\boldsymbol{\Pi}$ tiene filas iguales a $\boldsymbol{\pi}^\top$, y $\boldsymbol{\pi}$ es el único vector de probabilidades que satisface

$$\boldsymbol{\pi}^\top \boldsymbol{P} = \boldsymbol{\pi}^\top.$$

Método para hallar la distribución estacionaria:

Sea $\boldsymbol{A} \in \mathbb{R}^{k\times k}$, un vector propio (por la derecha), satisface:

$$\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u},$$

para algún escalar λ. Por otro lado, si tenemos que

$$\boldsymbol{v}^\top\boldsymbol{A} = \mu\boldsymbol{v}^\top,$$

$\boldsymbol{v}$ es llamado un vector propio por la izquierda.

Observación:

Además, un vector propio por la izquierda de $\boldsymbol{A}$ es un vector propio por la derecha de $\boldsymbol{A}^\top$.

Observación:

Sabemos que una matriz de transición debe satisfacer:

$$\boldsymbol{P}\mathbf{1} = \mathbf{1},$$

es decir, $\mathbf{1}$ es vector propio de $\boldsymbol{P}$ con valor propio asociado $\lambda = 1$.

Mientras que, si $\boldsymbol{\pi}$ es la distribución estacionaria de una cadena de Markov, satisface:

$$\boldsymbol{\pi}^\top \boldsymbol{P} = \boldsymbol{\pi}^\top,$$

es vector propio por la izquierda con $\lambda = 1$.