eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología

Matrices no negativas y cadenas de Markov

Trabajo Fin de Grado
Grado en Matemáticas

Amaia Lobato Izagirre

Trabajo dirigido por
Silvia Marcaida Bengoechea

Leioa, 27 de junio de 2017

Índice general

Introducción

La teoría de los procesos estocásticos se creó principalmente para satisfacer las necesidades de los físicos. Comenzó con el estudio de los fenómenos físicos como fenómenos aleatorios que cambiaban con el tiempo. Hoy en día la teoría de los procesos estocásticos se utiliza en un amplio abanico de ámbitos como la biología y la física (evolución de enfermedades, modelos climatológicos, comportamiento de partículas...), la economía (inversión en Bolsa, evaluación de riesgos...), etc. En este trabajo, hablaremos sobre un proceso estocástico en particular, llamado cadena de Markov. Las cadenas de Markov permiten predecir el comportamiento a corto y a largo plazo de sistemas que pueden cambiar de estado en cada instante de tiempo y que cumplen la propiedad de Markov, es decir, que el futuro del sistema, a partir de un presente conocido, es independiente del pasado. Las cadenas de Markov deben su nombre al matemático Andréi Márkov.

Andréi Márkov (1856-1922) fue un matemático ruso conocido por sus trabajos en la teoría de números y la teoría de la probabilidad. Cursó sus estudios universitarios de matemáticas entre 1874 y 1878, siendo premiado con una medalla de oro al finalizarlos. Realizó su carrera académica en la Universidad de San Petesburgo. Aunque su tesis doctoral trataba sobre la teoría de números, con la retirada de Chebyshev en 1883, Márkov pasó a encargarse del curso de teoría de la probabilidad que había impartido hasta entonces Chebyshev. Aunque fueron muchas las aportaciones de Márkov en las matemáticas, quizás la más conocida sea su trabajo en el campo de los procesos estocásticos. En concreto, las cadenas de Markov.

La teoría de las matrices no negativas también tiene muchas aplicaciones en ámbitos variados como la economía, la física, la probabilidad, etc. En particular es de gran utilidad en la investigación de las cadenas de Markov. Muchos de los resultados sobre cadenas de Markov se pueden obtener a través de propiedades de las matrices no negativas. En 1907 Perron descubrió algunas propiedades significativas de las matrices cuadradas positivas. Más adelante Frobenius amplió los resultados de Perron a las matrices no negativas. Desde entonces la teoría de las matrices no negativas ha sido una de las áreas más activas del álgebra lineal.

En este trabajo se tratan estos dos temas: las matrices no negativas (en particular irreducibles y estocásticas) y las cadenas de Markov. En primer lugar se habla sobre las matrices, ya que muchos de los resultados que se han conseguido en esta parte han servido después para desarrollar la teoría de las cadenas de Markov. Después de finalizar tanto con la teoría sobre matrices no negativas como con la teoría de las cadenas de Markov, se expone una aplicación de las cadenas de Markov: PageRank, que es la herramienta que utiliza el buscador Google para clasificar las páginas web.

Uno de los objetivos de este trabajo es ampliar los conocimientos que se han adquirido en varias asignaturas del Grado en Matemáticas como probabilidad y procesos estocásticos y ampliación de métodos numéricos.

En el Capítulo 1 se define el concepto de matriz irreducible y a tráves de una serie de resultados que se demuestran a lo largo del capítulo, se consigue una relación entre el grafo dirigido asociado a una matriz y la irreducibilidad de dicha matriz. Además, se definen los conceptos vector y valor propio de una matriz y se demuestra el teorema de Perron-Frobenius para matrices irreducibles; resultado que será útil en el capítulo que trata sobre las cadenas de Markov. Para concluir este capítulo, se define el concepto de matriz estocástica y se demuestra que el radio espectral de cualquier matriz estocástica es uno. Para desarrollar este capítulo se han utilizado [7], [8], [6] y [2].

En el Capítulo 2 se da la definición de una cadena de Markov y de algunos conceptos básicos como el espacio de estados y la matriz de transición de una cadena de Markov. El objetivo principal de este capítulo es estudiar el comportamiento de una cadena a largo plazo. Para escribir el capítulo se han utilizado [1], [11], [4] y [5].

Finalmente, en el Capítulo 3, se proporciona una aplicación de las cadenas de Markov: PageRank. La información de este capítulo se ha obtenido de [10], [9] y [3].

Capítulo 1

Matrices no negativas

Hoy en día, la teoría de las matrices no negativas es una de las áreas más activas del álgebra lineal. Tiene muchas aplicaciones en diversos campos como economía (modelos de Leontief), probabilidad (cadenas de Markov)...

Muchos de los resultados que demostremos en este capítulo se utilizarán posteriormente en el Capítulo 2 para desarrollar la teoría de las cadenas de Markov. Para elaborar este capítulo se ha utilizado como referencia principalmente [7].

1.1. Matrices Irreducibles

En esta sección se introduce el concepto de matriz irreducible y utilizando una serie de resultados, se proporciona una relación entre la irreducibilidad de una matriz y el grafo dirigido asociado a dicha matriz. Finalmente se definen los conceptos de valor propio y vector propio y se habla sobre los vectores y valores propios de una matriz irreducible y algunas de sus características.

1.1.1. Introducción

Definición 1.1.1. Dada una matriz $P \in \mathbb{R}^{n\times n}$, se dice que P es una *matriz de permutación* si en cada fila y columna de P hay un elemento igual a 1 y el resto de los elementos son cero.

Es muy fácil probar el siguiente resultado.

Proposición 1.1.1. *Sea $P \in \mathbb{R}^{n\times n}$ una matriz de permutación. Entonces se cumplen las siguientes propiedades:*

(i) $P^{-1} = P^T$.

(ii) P^T *es una matriz de permutación.*

Definición 1.1.2. Sea $n \geq 2$, entonces dada una matriz $A \in \mathbb{R}^{n\times n}$, se dice que A es una matriz *reducible* si existe una matriz de permutación $P \in \mathbb{R}^{n\times n}$ tal que

$$P^T AP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

donde A_{11} y A_{22} son dos matrices cuadradas de orden menor que n. En caso de que no exista dicha matriz P, se dice que A es una matriz *irreducible*.

Ahora demostraremos algunas propiedades de las matrices reducibles.

Proposición 1.1.2. *Sea $A \in \mathbb{R}^{n\times n}$ una matriz reducible. Entonces para cualquier $p \in \mathbb{N}$, A^p es una matriz reducible.*

Demostración. Sea $B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix} \in \mathbb{R}^{n\times n}$, siendo $B_{11} \in \mathbb{R}^{r\times r}$, $B_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$ y $1 \leq r < n$. Demostremos por inducción que

$$B^n = \begin{bmatrix} \widetilde{B_{11}} & \widetilde{B_{12}} \\ 0 & \widetilde{B_{22}} \end{bmatrix}, \quad \forall n \in \mathbb{N}, \tag{1.1}$$

donde $\widetilde{B_{11}} \in \mathbb{R}^{r\times r}$ y $\widetilde{B_{22}} \in \mathbb{R}^{(n-r)\times(n-r)}$.

- $n = 1$, es trivial.
- Hipótesis de inducción: Supongamos que $B^{n-1} = \begin{bmatrix} \widetilde{\widetilde{\widetilde{B_{11}}}} & \widetilde{\widetilde{\widetilde{B_{12}}}} \\ 0 & \widetilde{\widetilde{\widetilde{B_{22}}}} \end{bmatrix}$, con $\widetilde{\widetilde{\widetilde{B_{11}}}} \in \mathbb{R}^{r\times r}$ y $\widetilde{\widetilde{\widetilde{B_{22}}}} \in \mathbb{R}^{(n-r)\times(n-r)}$.
- Aplicando la hipótesis de inducción,

$$B^n = B^{n-1}B = \begin{bmatrix} \widetilde{\widetilde{\widetilde{B_{11}}}} & \widetilde{\widetilde{\widetilde{B_{12}}}} \\ 0 & \widetilde{\widetilde{\widetilde{B_{22}}}} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \widetilde{\widetilde{\widetilde{B_{11}}}}B_{11} & \widetilde{\widetilde{\widetilde{B_{11}}}}B_{12} + \widetilde{\widetilde{\widetilde{B_{12}}}}B_{22} \\ 0 & \widetilde{\widetilde{\widetilde{B_{22}}}}B_{22} \end{bmatrix}.$$

Como $\widetilde{\widetilde{\widetilde{B_{11}}}}, B_{11} \in \mathbb{R}^{r\times r}$ y $\widetilde{\widetilde{\widetilde{B_{22}}}}, B_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$, se tiene que $\widetilde{\widetilde{\widetilde{B_{11}}}}B_{11} \in \mathbb{R}^{r\times r}$ y $\widetilde{\widetilde{\widetilde{B_{22}}}}B_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$.

A continuación, demostraremos que para cualquier $p \in \mathbb{N}$, A^p es una matriz reducible. Por hipótesis A es reducible, por tanto, $\exists P \in \mathbb{R}^{n\times n}$ una matriz de permutación tal que $P^T AP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ siendo $A_{11} \in \mathbb{R}^{r\times r}$

y $A_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$ con $1 \leq r < n$. Por tanto, $(P^TAP)^p = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}^p$.
Como P es una matriz de permutación, $P^T = P^{-1}$, luego

$$(P^TAP)^p = P^TAPP^TAP\dots P^TAP = P^TAI_nAI_n\dots I_nAP = P^TA^pP.$$

Por otro lado, aplicando el resultado (1.1), $\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}^p = \begin{bmatrix} \widetilde{A_{11}} & \widetilde{A_{12}} \\ 0 & \widetilde{A_{22}} \end{bmatrix}$, con $\widetilde{A_{11}} \in \mathbb{R}^{r\times r}$ y $\widetilde{A_{22}} \in \mathbb{R}^{(n-r)\times(n-r)}$.

Así, $P^TA^pP = \begin{bmatrix} \widetilde{A_{11}} & \widetilde{A_{12}} \\ 0 & \widetilde{A_{22}} \end{bmatrix}$, siendo P una matriz de permutación y $\widetilde{A_{11}}$ y $\widetilde{A_{22}}$ dos matrices cuadradas de orden r y $n-r$ respectivamente. Luego A^p es una matriz reducible. □

Proposición 1.1.3. *Sea $A \in \mathbb{R}^{n\times n}$. Entonces A es reducible si y solo si A^T es reducible.*

Demostración. Supongamos que A es reducible. Entonces $\exists P \in \mathbb{R}^{n\times n}$ una matriz de permutación tal que $P^TAP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ siendo A_{11} y A_{22} dos matrices cuadradas de orden r y $n-r$ respectivamente, con $1 \leq r < n$. Por tanto,

$$(P^TAP)^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}^T \Leftrightarrow P^TA^TP = \begin{bmatrix} A_{11}^T & 0 \\ A_{12}^T & A_{22}^T \end{bmatrix}. \tag{1.2}$$

Sea $P_0 = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \iddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ 0 & \cdots & 0 & I_{n-2r} & 0 & \cdots & 0 \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \iddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$. Si a la matriz A se le multiplica la matriz P por la izquierda, se intercambian la primera y la n-ésima fila de A, la segunda y la $(n-1)$-ésima fila de A y así sucesivamente hasta intercambiar la r-ésima y la $(n-r+1)$-ésima fila de A. En cambio, si a la matriz A se le multiplica la matriz P por la derecha, se intercambian la primera y la n-ésima columna de A, la segunda y la $(n-1)$-ésima columna de A y así sucesivamente hasta intercambiar la r-ésima y la $(n-r+1)$-ésima columna de A. Por tanto, multiplicando la matriz P_0 por la derecha y por la izquierda en la igualdad (1.2), se tiene que

$$P_0P^TA^TPP_0 = P_0\begin{bmatrix} A_{11}^T & 0 \\ A_{12}^T & A_{22}^T \end{bmatrix}P_0 = \begin{bmatrix} \widetilde{A_{12}^T} & \widetilde{A_{22}^T} \\ \widetilde{A_{11}^T} & 0 \end{bmatrix}P_0 = \begin{bmatrix} \widetilde{\widetilde{A_{22}^T}} & \widetilde{\widetilde{A_{12}^T}} \\ 0 & \widetilde{\widetilde{A_{11}^T}} \end{bmatrix},$$

con $\widetilde{\widetilde{\widetilde{A_{22}^T}}} \in \mathbb{R}^{(n-r)\times(n-r)}$ y $\widetilde{\widetilde{\widetilde{A_{11}^T}}} \in \mathbb{R}^{r\times r}$.

Nótese que P_0 es una matriz simétrica y por tanto, $P_0 = P_0^T$, luego $(PP_0)^T = (PP_0^T)^T = P_0P^T$. Y como el producto entre dos matrices de permutación es una matriz de permutación, se tiene que A^T es una matriz reducible.

Supongamos que A^T es una matriz reducible. Entonces se tiene que $(A^T)^T = A$ es reducible.

$\square$

Corolario 1.1.4. *Sea $A \in \mathbb{R}^{n\times n}$. Entonces A es irreducible si y solo si A^T es irreducible.*

Demostración. Se deduce de la Proposición 1.1.3. $\square$

1.1.2. El grafo dirigido asociado a una matriz irreducible

En primer lugar, definiremos varios conceptos que vamos a utilizar a lo largo de esta sección.

Definición 1.1.3. Un *grafo dirigido G* está definido por dos conjuntos finitos llamados V y A, $G = (V, A)$. V es un conjunto cuyos elementos se llaman *vértices* o *nodos* y A es un conjunto de pares ordenados de vértices que se llaman *arcos*.

Observación 1.1.1. Llamaremos *líneas dirigidas* a los arcos de un grafo.

Definición 1.1.4. Dado $G = (V, A)$ un grafo dirigido, un *camino dirigido* es una secuencia de una o más líneas dirigidas tal que existe un vértice entre cada línea dirigida y la siguiente.

Definición 1.1.5. Dado $G = (V, A)$ un grafo dirigido, se dice que el grafo es *fuertemente conexo* si para cada par de nodos $P_i, P_j \in V$ con $i \neq j$, hay un camino dirigido que conecta P_i con P_j. Dicho camino puede ser directo de P_i a P_j o puede comenzar en P_i y pasar por otros nodos del grafo antes de llegar a P_j.

Observación 1.1.2. Los nodos P_i y P_j pueden estar conectados por un camino dirigido mientras que P_j y P_i no lo están.

Ejemplo 1.1.1. Como se puede observar en la Figura 1.1, existe un camino dirigido que conecta los nodos P_1 y P_3, pero en cambio no existe un camino dirigido que conecta los nodos P_3 y P_1.

Definición 1.1.6. Sean $P_1, P_2, ..., P_n$ distintos puntos de la recta real y sea $A \in \mathbb{R}^{n\times n}$. Por cada elemento distinto de cero a_{ij} de A, se conectan los puntos P_i y P_j con una línea dirigida. La figura resultante en la recta real es un *grafo dirigido asociado a A*.

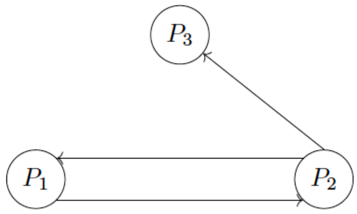

Figura 1.1: Grafo dirigido

Proposición 1.1.5. *Sea $A \in \mathbb{R}^{n\times n}$. Entonces si todos los elementos de una fila (o columna) de A que están fuera de la diagonal principal son cero, el grafo dirigido asociado a A no es fuertemente conexo.*

Demostración. Sea $A \in \mathbb{R}^{n\times n}$. Supongamos que los elementos de la r-ésima fila de A que no están en la diagonal principal son cero. Es decir, $a_{rj} = 0$ $\forall j \neq r$. Por tanto, no existe ni una línea dirigida que salga desde el nodo P_r y se dirija a un nodo P_j tal que $j \neq r$. Por tanto, no existe un camino dirigido que una el nodo P_r con el nodo P_j $\forall j \neq r$. Entonces el grafo dirigido asociado a A no es fuertemente conexo. Se puede hacer el mismo razonamiento por columnas. □

Proposición 1.1.6. *Sea $A \in \mathbb{R}^{n\times n}$ tal que $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, $A_{11} \in \mathbb{R}^{k\times k}$, $A_{12} \in \mathbb{R}^{k\times(n-k)}$, $A_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$ y $1 \le k \le n-1$. Entonces el grafo dirigido asociado a A no es fuertemente conexo.*

Demostración. Consideremos un camino dirigido que comience desde el nodo P_i tal que $i > k$. Como $i > k$, en la i-ésima fila de A los primeros k elementos son 0. Por tanto, el nodo P_i solamente se puede conectar con un nodo P_j mediante una línea dirigida si $j > k$. Del mismo modo, como $j > k$, los primeros k elementos de la j-ésima fila de A son 0. Luego el nodo P_j solamente se puede conectar con el nodo P_z mediante una línea dirigida si $z > k$. Siguiendo este procedimiento, llegamos a la conclusión de que si $i > k$ y $j \le k$, no existe un camino dirigido de P_i a P_j. Por tanto, el grafo dirigido asociado a A no es fuertemente conexo. □

Proposición 1.1.7. *Sea $A \in \mathbb{R}^{n\times n}$ y sea $P \in \mathbb{R}^{n\times n}$ una matriz de permutación. Entonces el grafo dirigido asociado a A es fuertemente conexo si y solo si el grafo dirigido asociado a PAP^T es fuertemente conexo.*

Demostración. Supongamos que el grafo dirigido asociado a A es fuertemente conexo. Observamos que el grafo dirigido de PAP^T se obtiene renumerando los nodos del grafo de A. Es decir, si la matriz de permutación P pone la i-ésima fila de A en la j-ésima, entonces el nodo P_i del grafo de A, será el nodo P_j del grafo de PAP^T. Por tanto, como el grafo dirigido asociado a A es fuertemente conexo, el grafo dirigido asociado a PAP^T también lo será.

Supongamos que el grafo dirigido asociado a PAP^T es fuertemente conexo. Como P es una matriz de permutación, P^T también lo es, y por tanto, el grafo dirigido asociado a $P^T PAP^T (P^T)^T = A$ es fuertemente conexo. □

Definición 1.1.7. Dada una matriz $A \in \mathbb{R}^{m\times n}$, se dice que A es *no negativa* si no hay ni un elemento de A que sea negativo.

Definición 1.1.8. Dada una matriz $A \in \mathbb{R}^{m\times n}$, se dice que A es *positiva* si todos los elementos de A son positivos.

Observación 1.1.3. Sea $A \in \mathbb{R}^{m\times n}$. Para expresar que A es una matriz no negativa, escribiremos $A \geq 0$. Si A es una matriz positiva, escribiremos $A > 0$. Y finalmente para expresar que A no es la matriz nula escribiremos $A \neq 0$.

Ahora, demostraremos algunas propiedades de las matrices no negativas e irreducibles.

Proposición 1.1.8. *Sea $A \in \mathbb{R}^{n\times n}$ una matriz no negativa e irreducible. Entonces $(I_n + A)^{n-1} > 0$.*

Demostración. Sea $y \in \mathbb{R}^n$ un vector tal que $y \geq 0$ e $y \neq 0$. Se define

$$z = (I_n + A)y = y + Ay. \tag{1.3}$$

Como $y \geq 0$ y por hipótesis $A \geq 0$, se tiene que $Ay \geq 0$. Por tanto, z tiene al menos tantos elementos no nulos como y. Veamos que si algún elemento de y es cero, entonces z tiene al menos un elemento no nulo más que y. Sea P una matriz de permutación tal que $Py = \begin{bmatrix} u \\ 0 \end{bmatrix}$ y $u > 0$. Utilizando la igualdad (1.3), se tiene que

$$Pz = P(y + Ay) = Py + PAy = Py + PAP^T Py = \begin{bmatrix} u \\ 0 \end{bmatrix} + PAP^T \begin{bmatrix} u \\ 0 \end{bmatrix}. \tag{1.4}$$

Se dividen Pz y PAP^T de acuerdo con $\begin{bmatrix} u \\ 0 \end{bmatrix}$. Es decir, $Pz = \begin{bmatrix} v \\ w \end{bmatrix}$ siendo u y v dos vectores del mismo tamaño y $PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ siendo A_{11} una matriz cuadrada del mismo tamaño que u. Debido a la ecuación (1.4),

$$v = u + A_{11}u \quad \text{y} \quad w = A_{21}u. \tag{1.5}$$

Como $A \geq 0$, $P \geq 0$ y $P^T \geq 0$, se tiene que $PAP^T \geq 0$. En particular $A_{11} \geq 0$ y $A_{21} \geq 0$. Como A es irreducible, se tiene que $A_{21} \neq 0$. Por tanto, debido a las ecuaciones de (1.5) y que $u > 0$, se tiene que $v > 0$ y $w \geq 0$ pero $w \neq 0$. Luego $Pz = \begin{bmatrix} v \\ w \end{bmatrix}$ tendrá al menos un elemento no nulo más que y. Como $z = P^T \begin{bmatrix} v \\ w \end{bmatrix}$, z tiene los mismos elementos que Pz pero en distinto orden, y por tanto, z tiene al menos un elemento no nulo más que y.

Si $(I_n + A)y = z$ todavía tiene algún elemento nulo, se considera el vector $(I_n + A)y \neq 0$ y se repite el mismo proceso con $z = (I_n + A)[(I_n + A)y] = (I_n + A)^2 y$ y se consigue que $(I_n + A)^2 y$ tiene al menos un elemento no nulo más que $(I_n + A)y$, es decir, que $(I_n + A)^2 y$ tiene al menos dos elementos no nulos más que y. Se repite el proceso hasta conseguir $k_0 \in \mathbb{N}$ tal que $(I_n + A)^{k_0} y > 0$. Como $y \in \mathbb{R}^n$, el proceso se repetirá como mucho $n - 1$ veces. Vamos a diferenciar dos casos: $k_0 = n - 1$ y $k_0 < n - 1$. Si $k_0 = n - 1$, se tiene que $(I_n + A)^{n-1} y > 0$. Si $k_0 < n - 1$, se tiene que $(I_n + A)^{n-1} y = (I_n + A)^{n-1-k_0}(I_n + A)^{k_0} y > 0$, ya que los elementos de la diagonal principal de $(I_n + A)^{n-1-k_0}$ son positivos por ser $A \geq 0$ y $(I_n + A)^{k_0} y > 0$. Por tanto,

$$(I_n + A)^{n-1} y > 0 \quad \forall y \in \mathbb{R}^n \quad \text{tal que} \quad y \geq 0, y \neq 0.$$

Por tanto, $(I_n + A)^{n-1} e_j > 0$ $\forall j \in \{1, \ldots, n\}$. Es decir, los elementos de la j-ésima columna de $(I_n + A)^{n-1}$ son positivos $\forall j \in \{1, \ldots, n\}$. Por tanto, $(I_n + A)^{n-1} > 0$. $\square$

Proposición 1.1.9. *Sea $A \in \mathbb{R}^{n \times n}$ una matriz no negativa e irreducible donde $a_{ij}^{(q)}$ es el (i, j)-ésimo elemento de A^q. Entonces para cada $i, j \in \{1, \ldots, n\}$ existe $q \in \mathbb{N}$ tal que $a_{ij}^{(q)} > 0$.*

Demostración. Como $A \in \mathbb{R}^{n \times n}$ es una matriz no negativa e irreducible, por la Proposición 1.1.8, se tiene que $(I_n + A)^{n-1} > 0$. Veamos que $A(I_n + A)^{n-1} > 0$. Sea $A(I_n + A)^{n-1} = [c_{ij}]_{1 \leq i,j \leq n}$. Supongamos por reducción al absurdo que $\exists i_0, j_0 \in \{1, \ldots, n\}$ tales que $c_{i_0 j_0} = 0$. Sean $A = [a_{ij}]_{1 \leq i,j \leq n}$ y $(I_n + A)^{n-1} = [b_{ij}]_{1 \leq i,j \leq n}$. Como $A \geq 0$, se tiene que $a_{ij} \geq 0$ $\forall i, j \in \{1, \ldots, n\}$ y como $(I_n + A)^{n-1} > 0$, se tiene que $b_{ij} > 0$ $\forall i, j \in \{1, \ldots, n\}$. Como $c_{i_0 j_0} = 0$ es el (i_0, j_0)-ésimo elemento de $A(I_n + A)^{n-1}$, se tiene que

$$c_{i_0 j_0} = \sum_{k=1}^{n} a_{i_0 k} b_{k j_0} = 0. \tag{1.6}$$

Como $a_{i_0 k} \geq 0$ y $b_{k j_0} > 0$ $\forall k \in \{1, \ldots, n\}$, para que se cumpla la igualdad (1.6), tiene que cumplirse que $a_{i_0 k} = 0$, $\forall k \in \{1, \ldots, n\}$. Es decir, la i_0-ésima fila de A está compuesta por ceros. Sea $P \in \mathbb{R}^{n \times n}$ la matriz de permutación que se obtiene al intercambiar la i_0-ésima y la n-ésima filas de la matriz I_n.

Por tanto,

$$PAP^T = \begin{bmatrix} a_{11} & \cdots & a_{1n} & \cdots & a_{1i_0} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ a_{n1} & \cdots & a_{nn} & \cdots & a_{ni_0} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ a_{i_01} & \cdots & a_{i_0n} & \cdots & a_{i_0i_0} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & \cdots & a_{1i_0} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ a_{n1} & \cdots & a_{nn} & \cdots & a_{ni_0} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$
$$= \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \text{con } A_{11} \in \mathbb{R}^{(n-1)\times(n-1)}, A_{12} \in \mathbb{R}^{(n-1)\times 1} \text{ y } A_{22} \in \mathbb{R}^{1\times 1}.$$

Entonces A es una matriz reducible y se obtiene una contradicción ya que A es una matriz irreducible por hipótesis. Por lo tanto, $c_{ij} > 0$ $\forall i, j \in \{1, \dots, n\}$, es decir, $A(I_n + A) > 0$. Como $AI_n = I_nA$, utilizaremos la fórmula del binomio para calcular $(I_n + A)^{n-1}$.

$$\begin{aligned} A(I_n + A)^{n-1} &= A\left[\binom{n-1}{0}A^{n-1} + \binom{n-1}{1}I_nA^{n-2} + \ldots + \binom{n-1}{n-1}I_n^{n-1}\right] \\ &= A^n + (n-1)A^{n-1} + \ldots + (n-1)A^2 + A > 0. \end{aligned} \tag{1.7}$$

Por la igualdad (1.7), para cada $i, j \in \{1, \dots, n\}$ $\exists q \in \{1, \dots, n\}$ tal que $a_{ij}^{(q)} > 0$. □

Proposición 1.1.10. *Sea $A \in \mathbb{R}^{n\times n}$ una matriz no negativa e irreducible y sea $a_{ij}^{(q)}$ el (i, j)-ésimo elemento de A^q. Entonces $a_{ij}^{(q)} > 0$ si y solo si existe una secuencia de q líneas dirigidas en el grafo dirigido asociado a A que conectan los nodos P_i y P_j.*

Demostración. Demostremos por inducción que

$$a_{ij}^{(q)} = \sum_{k_{q-1}=1}^{n} \sum_{k_{q-2}=1}^{n} \cdots \sum_{k_2=1}^{n} \sum_{k_1=1}^{n} a_{ik_1} a_{k_1k_2} \ldots a_{k_{q-2}k_{q-1}} a_{k_{q-1}j}\,, \quad \forall q \geq 2.$$

- $q = 2$, $a_{ij}^{(2)} = \sum_{k_1=1}^{n} a_{ik_1} a_{k_1j}$.
- $q = 3$,

$$a_{ij}^{(3)} = \sum_{k_2=1}^{n} a_{ik_2}^{(2)} a_{k_2j} = \sum_{k_2=1}^{n} \left(\sum_{k_1=1}^{n} a_{ik_1} a_{k_1k_2} \right) a_{k_2j} = \sum_{k_2=1}^{n} \sum_{k_1=1}^{n} a_{ik_1} a_{k_1k_2} a_{k_2j}.$$

- Hipótesis de inducción: Supongamos que

$$a_{ij}^{(q-1)} = \sum_{k_{q-2}=1}^{n} \sum_{k_{q-3}=1}^{n} \cdots \sum_{k_2=1}^{n} \sum_{k_1=1}^{n} a_{ik_1} a_{k_1k_2} \ldots a_{k_{q-3}k_{q-2}} a_{k_{q-2}j}.$$

- Aplicando la hipótesis de inducción,

$$\begin{aligned}
a_{ij}^{(q)} &= \sum_{k_{q-1}=1}^{n} a_{ik_{q-1}}^{(q-1)} a_{k_{q-1}j} \\
&= \sum_{k_{q-1}=1}^{n} \left(\sum_{k_{q-2}=1}^{n} \sum_{k_{q-3}=1}^{n} \cdots \sum_{k_2=1}^{n} \sum_{k_1=1}^{n} a_{ik_1} a_{k_1k_2} \ldots a_{k_{q-3}k_{q-2}} a_{k_{q-2}k_{q-1}} \right) a_{k_{q-1}j} \\
&= \sum_{k_{q-1}=1}^{n} \sum_{k_{q-2}=1}^{n} \cdots \sum_{k_2=1}^{n} \sum_{k_1=1}^{n} a_{ik_1} a_{k_1k_2} \ldots a_{k_{q-2}k_{q-1}} a_{k_{q-1}j}.
\end{aligned} \tag{1.8}$$

Por hipótesis A es una matriz no negativa, por tanto, considerando la igualdad (1.8), se tiene que $a_{ij}^{(q)} > 0$ si y solo si $\exists k_{1_0}, \ldots, k_{q-1_0} \in \{1, \ldots, n\}$ tales que $a_{ik_{1_0}} a_{k_{1_0}k_{2_0}} \ldots a_{k_{q-1_0}j} > 0$. Como A es una matriz no negativa, $a_{ik_{1_0}} a_{k_{1_0}k_{2_0}} \ldots a_{k_{q-1_0}j} > 0$ si y solo si $a_{ik_{1_0}}, a_{k_{1_0}k_{2_0}}, \ldots, a_{k_{q-1_0}j} > 0$. Es decir, que hay una secuencia de q líneas dirigidas en el grafo dirigido de A que conectan los nodos P_i y P_j. □

Ahora, estamos listos para demostrar el siguiente teorema.

Teorema 1.1.11. *Sea $A \in \mathbb{R}^{n\times n}$ una matriz no negativa. Entonces A es irreducible si y solo si el grafo dirigido asociado a A es fuertemente conexo.*

Demostración. Supongamos que $A \in \mathbb{R}^{n\times n}$ es una matriz irreducible. Como A también es una matriz no negativa, por la Proposición 1.1.9, para cada $i, j \in \{1, \ldots, n\}$ $\exists q \in \mathbb{N}$ tal que $a_{ij}^{(q)} > 0$. Por la Proposición 1.1.10 $\forall i, j \in \{1, \ldots, n\}$ los nodos P_i y P_j están conectados. Por tanto, el grafo dirigido asociado a A es fuertemente conexo.

Supongamos que $A \in \mathbb{R}^{n\times n}$ es una matriz reducible. Entonces, $\exists P \in \mathbb{R}^{n\times n}$ una matriz de permutación tal que $P^T AP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, siendo A_{11} y A_{22} dos matrices cuadradas de orden menor que n. Por la Proposición 1.1.6, se tiene que el grafo dirigido de $P^T AP$ no es fuertemente conexo. Al ser $P^T \in \mathbb{R}^{n\times n}$ una matriz de permutación, debido a la Proposición 1.1.7, se tiene que el grafo dirigido asociado a A no es fuertemente conexo. □

1.1.3. Teorema de Perron-Frobenius

En primer lugar, vamos a definir algunos conceptos básicos.

Definición 1.1.9. Sea $A \in \mathbb{R}^{n\times n}$, se dice que $\lambda \in \mathbb{C}$ es un *valor propio* de A si existe un vector no nulo $x \in \mathbb{C}^n$ tal que $Ax = \lambda x$.

Definición 1.1.10. Dado λ un valor propio de $A \in \mathbb{R}^{n\times n}$, se dice que $x \in \mathbb{C}^n$ es un *vector propio* de A asociado al valor propio λ si $x \neq 0$ y $Ax = \lambda x$.

Definición 1.1.11. Dados $A \in \mathbb{R}^{n\times n}$ y λ un valor propio de A, el *subespacio propio* de A asociado a λ es $Ker(\lambda I_n - A) = \{x \in \mathbb{C}^n \mid Ax = \lambda x\}$.

Definición 1.1.12. Dados $A \in \mathbb{R}^{n\times n}$ y λ un valor propio de A, la *multiplicidad geométrica* de λ es la dimensión del subespacio propio de A asociado a λ.

Definición 1.1.13. Dada $A \in \mathbb{R}^{n\times n}$, se dice que $\det(\lambda I_n - A) = (\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\dots(\lambda-\lambda_s)^{m_s}$ es el *polinomio característico* de A, siendo $\lambda_1,\dots,\lambda_s$ los valores propios de A y $\lambda_i \neq \lambda_j$. Además, m_i es la *multiplicidad algebraica* de λ_i para cada $i = 1,\dots,s$.

Definición 1.1.14. Sea $\mathcal{S}_r = Ker(\lambda I_n - A)^r$. Es fácil demostrar que $\mathcal{S}_r \subset \mathcal{S}_{r+1}$ para cada $r \in \mathbb{N} \cup \{0\}$. A $\mathcal{S}_r$ se le llama *subespacio propio generalizado* de A *de orden* r asociado a λ. Además, cualquier vector no nulo $x \in \mathbb{C}$ tal que $x \in \mathcal{S}_r$ pero $x \notin \mathcal{S}_r$ es un *vector propio generalizado* de A *de orden* r asociado a λ.

Observaciones. (i) Dados $x, y \in \mathbb{R}^n$, para expresar que $x_i \leq y_i$ para cada $i \in \{1,\dots,n\}$, escribiremos $x \leq y$, siendo x_i e y_i las i-ésimas componentes de los vectores x e y respectivamente. Si para cada $i \in \{1,\dots,n\}$ $x_i < y_i$, escribiremos $x < y$.

(ii) Dada $A = [a_{ij}] \in \mathbb{R}^{n\times n}$, definimos $|A| = [|a_{ij}|] \in \mathbb{R}^{n\times n}$.

Sea $A \in \mathbb{R}^{n\times n}$ una matriz no negativa e irreducible y sea $x \in \mathbb{R}^n$ un vector no negativo y no nulo. Consideramos la siguiente función:

$$r(x) = \min_{1\leq i\leq n,\, x_i\neq 0} \frac{(Ax)_i}{x_i},$$

siendo $(Ax)_i$ el i-ésimo elemento del vector Ax. Como $A \geq 0$ y $x \geq 0$, se tiene que $r(x) \geq 0$. Además, por la definición de $r(x)$, se cumple que $r(x)x_i \leq (Ax)_i$ para cualquier $i = 1,\dots,n$. Por tanto, $r(x)x \leq Ax$. Veamos que $r(x)$ es el mayor entre todos los números ρ que cumplen la desigualdad $\rho x \leq Ax$. Supongamos por reducción al absurdo que existe $\rho > r(x)$ tal que $\rho x \leq Ax$. Entonces:

$$\rho x \leq Ax \Leftrightarrow \rho x_i \leq (Ax)_i \ \forall i \in \{1,\dots,n\} \Rightarrow \rho \leq \frac{(Ax)_i}{x_i}, \forall x_i \neq 0 \text{ tal que } i \in \{1,\dots,n\}$$
$$\Rightarrow \rho \leq \min_{1\leq i\leq n,\, x_i\neq 0} \frac{(Ax)_i}{x_i} = r(x) \Rightarrow \rho \leq r(x).$$

Y esto no es posible ya que hemos supuesto que $\rho > r(x)$. Por tanto, queda demostrado que $r(x)$ es el mayor entre todos los números ρ que cumplen la desigualdad $\rho x \leq Ax$.

Sea $\mathcal{L} = \{x \in \mathbb{R}^n : x \geq 0, x \neq 0\}$ el dominio de la función $r(x)$ y definimos r como:

$$r = \sup_{x \in \mathcal{L}} r(x).$$

Por la definición de $r(x)$, r no varía si x se sustituye por αx para cualquier $\alpha > 0$ ya que $r(\alpha x) = \text{mín}_{1 \leq i \leq n,\, \alpha x_i \neq 0} \frac{(A\alpha x)_i}{\alpha x_i} = \text{mín}_{1 \leq i \leq n,\, x_i \neq 0} \frac{\alpha (Ax)_i}{\alpha x_i} = \text{mín}_{1 \leq i \leq n,\, x_i \neq 0} \frac{(Ax)_i}{x_i} = r(x)$. Por lo tanto, para obtener el supremo, solamente consideraremos los vectores del conjunto cerrado $\mathcal{M} = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x \geq 0, x \neq 0\} \subset \mathcal{L}$. Luego $r = \sup_{x \in \mathcal{M}} r(x)$. Como $\mathcal{M}$ es un conjunto cerrado y acotado, si la función $r(x)$ fuese continua en $\mathcal{M}$, se podría sustituir el supremo por el máximo. Pero no podemos garantizar que la función $r(x)$ sea continua en $\mathcal{M}$ ya que podría no ser continua en los puntos en los que algún elemento de x se hace cero. Por tanto, consideraremos el conjunto de vectores $\mathcal{N} = \{y = (I_n + A)^{n-1} x : x \in \mathcal{M}\}$. Como A es una matriz no negativa e irreducible, por la Proposición 1.1.8, tenemos que $(I_n + A)^{n-1} > 0$. Además, $x \geq 0$ y $x \neq 0$. Entonces $y = (I_n + A)^{n-1} x > 0$. Por lo tanto, $\mathcal{N} \subset \mathcal{L}$.

$\mathcal{N}$ es la imagen del conjunto acotado y cerrado $\mathcal{M}$ mediante una función continua.

$$\begin{array}{rccl} f\colon & \mathcal{M} & \longrightarrow & \mathcal{N} \\ & x & \longmapsto & y = (I_n + A)^{n-1} x. \end{array}$$

Entonces $\mathcal{N}$ es un conjunto cerrado y acotado. Además $r(y)$ es una función continua en $\mathcal{N}$. Por lo tanto, $r(y)$ alcanzará el máximo para algún $y \in \mathcal{N}$. Para cada $x \in \mathcal{M}$ e $y = (I_n + A)^{n-1} x \in \mathcal{N}$, tenemos que,

$$\begin{aligned} r(x) y = r(x)(I_n + A)^{n-1} x &= (I_n + A)^{n-1} r(x) x \\ &\leq (I_n + A)^{n-1} A x = A (I_n + A)^{n-1} x = Ay. \end{aligned}$$

Como $r(y)$ es el mayor número ρ tal que $\rho y \leq Ay$, tenemos que $r(x) \leq r(y)$. Por lo tanto, $r = \sup_{x \in \mathcal{M}} r(x) \leq \text{máx}_{y \in \mathcal{N}} r(y)$. Pero como $\mathcal{N} \subset \mathcal{L}$, tenemos que $\text{máx}_{y \in \mathcal{N}} r(y) \leq \sup_{x \in \mathcal{L}} r(x) = \sup_{x \in \mathcal{M}} r(x) = r$. Entonces,

$$r = \underset{y \in \mathcal{N}}{\text{máx}}\, r(y).$$

Podría pasar que hubiera otros vectores en $\mathcal{L}$ para los cuales $r(x)$ alcanzase el valor r. Aquellos vectores que cumplen esta condición son los *vectores extremales* de A. Por tanto, un vector no nulo $z \geq 0$ es un vector extremal de A si $r(z) = r$ o lo que es equivalente, si $rz \leq Az$.

Proposición 1.1.12. *Sea* $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ *una matriz no negativa e irreducible y* $x = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$. *Entonces* $r(x) = \text{mín}_{1 \leq i \leq n} \sum_{k=1}^n a_{ik}$.

Demostración. Como $x_i = 1$ para cualquier $i \in \{1, \dots, n\}$,

$$r(x) = \min_{1 \le i \le n,\, x_i \neq 0} \frac{(Ax)_i}{x_i} = \min_{1 \le i \le n} \frac{\sum_{k=1}^{n} a_{ik}}{1} = \min_{1 \le i \le n} \sum_{k=1}^{n} a_{ik}.$$

□

Proposición 1.1.13. *Sea $A \in \mathbb{R}^{n \times n}$ una matriz no negativa e irreducible. Entonces $r = \sup_{x \in \mathcal{L}} r(x)$ es positivo y es un valor propio de la matriz A. Además, cada vector extremal de A es positivo y es un vector propio por la derecha de A asociado al valor propio r.*

Demostración. Sea $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ y $x_0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$. Por la Proposición 1.1.12 sabemos que $r(x_0) = \min_{1 \le i \le n} \sum_{k=1}^{n} a_{ik}$. En la demostración de la Proposición 1.1.9 hemos visto que si una matriz contiene una fila formada por ceros, dicha matriz es reducible. Como A es irreducible, en cada fila de A debe haber al menos un elemento no nulo. En consecuencia, $\sum_{k=1}^{n} a_{ik} > 0$ para cualquier $i \in \{1, \dots, n\}$. Luego $r(x_0) = \min_{1 \le i \le n} \sum_{k=1}^{n} a_{ik} > 0$. Por lo tanto, $r = \sup_{x \in \mathcal{L}} r(x) \ge r(x_0) > 0$.

Sea z un vector extremal y sea $w = (I_n + A)^{n-1} z$. Sin pérdida de generalidad, podemos suponer que $z \in \mathcal{M}$ ya que $r(z) = r(\alpha z)$ para cualquier $\alpha > 0$. Como A es una matriz no negativa e irreducible, por la Proposición 1.1.8 $(I_n + A)^{n-1} > 0$. Además $z \ge 0$ y $z \neq 0$. Por lo tanto, $w = (I_n + A)^{n-1} z > 0$ y $w \in \mathcal{N}$. Por ser z un vector extremal, $rz \le Az$, es decir, $Az - rz \ge 0$. Veamos que $Az - rz = 0$. Supongamos por reducción al absurdo que $Az - rz \neq 0$. Entonces como $(I_n + A)^{n-1} > 0$, tenemos que $(I_n + A)^{n-1}(Az - rz) > 0$. Pero $(I_n + A)^{n-1}(Az - rz) = (I_n + A)^{n-1} Az - (I_n + A)^{n-1} rz = A(I_n + A)^{n-1} z - r(I_n + A)^{n-1} z = Aw - rw$. Luego $Aw - rw > 0$, es decir, $rw < Aw$ lo que implica que $r < r(w)$ y esto no tiene sentido. Por lo tanto, $Az - rz = 0$. Entonces r es un valor propio de A y cualquier vector extremal es un vector propio por la derecha de A asociado al valor propio r.

Finalmente, para demostrar que cada vector extremal, z, de A es positivo, vamos a ver que $(I_n + A)^k z = (1 + r)^k z$ para cualquier $k \in \mathbb{N}$. Como z es un vector extremal de A, sabemos que $Az = rz$. Sumando z, tenemos que $(I_n + A)z = (1 + r)z$. Para $k = 2$, $(I_n + A)^2 z = (I_n + A)(I_n + A)z = (I_n + A)(1 + r)z = (1 + r)(I_n + A)z = (1 + r)^2 z$. Supongamos que $(I_n + A)^{k-1} z = (1 + r)^{k-1} z$. Entonces, $(I_n + A)^k = (I_n + A)(I_n + A)^{k-1} z = (I_n + A)(1 + r)^{k-1} z = (1 + r)^{k-1}(I_n + A)z = (1 + r)^k z$. Por lo tanto, $(1 + r)^{n-1} z = (I_n + A)^{n-1} z = w > 0$. Y por ser $r > 0$, tenemos que $z > 0$. □

Definición 1.1.15. Dada $A \in \mathbb{R}^{n \times n}$ se define el *radio espectral* de A de la siguiente manera:

$$\text{máx}\{|\lambda_i| : \lambda_i \text{ es un valor propio de } A\}.$$

Teorema 1.1.14 (Teorema de Perron-Frobenius para matrices irreducibles). *Sea $A \in \mathbb{R}^{n\times n}$ una matriz no negativa e irreducible. Entonces:*

(i) *La matriz A tiene un valor propio positivo, r, igual al radio espectral de A.*

(ii) *Hay un vector positivo por la derecha asociado al valor propio r.*

(iii) *El valor propio r tiene multiplicidad geométrica 1.*

(iv) *El valor propio r tiene multiplicidad algebraica 1.*

Demostración. (i) Por la Proposición 1.1.13, sabemos que $r = \sup_{x\in\mathcal{L}} r(x)$ es un valor propio positivo de A. Veamos que el valor absoluto de cualquier otro valor propio de A es menor o igual que r. Sea α un valor propio de A. Por tanto, existe un vector no nulo, $y \in \mathbb{R}^n$ tal que $Ay = \alpha y$. Como $A \geq 0$, $|\alpha||y| = |\alpha y| = |Ay| \leq |A||y| = A|y|$. Entonces, $|\alpha||y| \leq A|y|$. Por lo tanto, $|\alpha| \leq r(|y|) \leq r$. Y queda demostrado que r es el radio espectral de A.

(ii) Por la Proposición 1.1.13, sabemos que existe un vector propio por la derecha positivo asociado a r (los vectores extremales).

(iii) Supongamos que z es cualquier vector propio por la derecha de A asociado a r. Entonces, $Az = rz$ y como en el apartado (i), se puede demostrar que $r|z| \leq A|z|$. En consecuencia, $|z|$ es un vector extremal de A y por la Proposición 1.1.13, $|z| > 0$. Por lo tanto, $z_i \neq 0$, $i = 1, \ldots, n$, siendo z_i la i-ésima componente del vector z. Veamos que la dimensión del subespacio propio de A asociado a r es 1. Supongamos por reducción al absurdo que la dimensión es mayor que 1. Entonces, existen dos vectores linealmente independientes, v_1 y v_2, y existen $\alpha, \beta \in \mathbb{R}$ tales que $\alpha v_1 + \beta v_2$ tiene al menos un elemento nulo. Pero como $\alpha v_1 + \beta v_2$ es una combinación lineal de v_1 y v_2, es un vector propio por la derecha de A asociado a r. Y que tenga al menos un elemento nulo no es posible. Por lo tanto, la dimensión del subespacio propio de r es 1, es decir, que la multiplicidad geométrica de r es 1.

(iv) Como la multiplicidad geométrica de r es 1, para demostrar que la multiplicidad algebraica de r es 1 vamos a ver que no existe ningún vector propio generalizado de r de orden 2. Sean $x_1 > 0$ e $y > 0$ vectores propios asociados al valor propio r de A y A^T respectivamente. Entonces, $Ax_1 = rx_1$ y $A^Ty = ry$. Por lo que $(rI_n - A)x_1 = 0$ y $(rI_n - A^T)y = 0$. Supongamos por reducción al absurdo que existe un vector propio generalizado de orden 2, $x_2 \neq 0$ asociado al valor propio r tal que $(rI_n - A)x_2 = x_1$. Como $(rI_n - A^T)y = 0$, se tiene que $y^T(rI_n - A) = 0^T$ y al multiplicar por x_2, se consigue $y^T(rI_n - A)x_2 = 0$.

Es decir, $y^T x_1 = 0$, pero no puede ser ya que $x_1 > 0$ e $y > 0$. Por lo tanto, no existe un vector propio generalizado de orden 2 asociado al valor propio r. Entonces, la multiplicidad algebraica de r es 1. $\square$

Teorema 1.1.15. *Sea $A \in \mathbb{R}^{n\times n}$ una matriz no negativa e irreducible con radio espectral 1. Entonces las siguientes afirmaciones son equivalentes:*

(i) $A^k \xrightarrow{k\to\infty} P$ *para alguna matriz* P.

(ii) *Para algún entero positivo* p, $A^p > 0$.

La demóstración se encuentra en Bapat, Raghavan [1].

1.2. Matrices estocásticas

Una matriz estocástica es una matriz cuadrada no negativa donde la suma de los elementos de cada fila es uno.

Definición 1.2.1. Dada $P = [p_{ij}]_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$, se dice que P es una matriz *estocástica* si $P \ge 0$ y $\sum_{j=1}^n p_{ij} = 1$, para todo $i = 1,\dots,n$.

Teorema 1.2.1. *Sea $P = [p_{ij}]_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$ una matriz estocástica. Entonces $\lambda = 1$ es un valor propio de P y además el radio espectral de P es 1.*

Demostración. Sea $v = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$. Veamos que $Pv = v$:

$$Pv = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n p_{1j} \\ \vdots \\ \sum_{j=1}^n p_{nj} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = v.$$

Entonces $\lambda_0 = 1$ es un valor propio de P. Sea λ un valor propio cualquiera de P y sea $v \ne 0$ un vector propio asociado a λ. Definimos $|v_i| = \max_{1\le k\le n}\{|v_k|\}$. Como $Pv = \lambda v$, $\sum_{j=1}^n p_{kj}v_j = \lambda v_k$, $k = 1,\dots,n$. En particular, $\sum_{j=1}^n p_{ij}v_j = \lambda v_i$. Entonces $|\lambda||v_i| = |\sum_{j=1}^n p_{ij}v_j| \le \sum_{j=1}^n |p_{ij}v_j| = \sum_{j=1}^n p_{ij}|v_j|$. Por lo tanto,

$$|\lambda| \le \sum_{j=1}^n p_{ij}\frac{|v_j|}{|v_i|} \le \sum_{j=1}^n p_{ij} = 1,$$

ya que $|v_i| = \max_{1\le k\le n}\{|v_k|\}$ y P es una matriz estocástica. Como $|\lambda| \le 1$, y $\lambda_0 = 1$ es un valor propio de P, el radio espectral de P es 1. $\square$

Capítulo 2

Cadenas de Markov

Las cadenas de Markov se utilizan para predecir la evolución y el comportamiento de ciertos sistemas a corto y a largo plazo. Deben su nombre al matemático ruso Andréi Márkov (1856-1922). Las cadenas de Markov tienen un amplio abanico de aplicacciones en varias áreas de la ciencia y la tecnología (por ejemplo física, teoría de colas, teoría de programación dinámica...).

Para desarrollar la teoría de este capítulo utilizaremos varios resultados que han sido demostrados previamente en el Capítulo 1. En la elaboración del Capítulo 2 se han utilizado [1] y [11] entre otros.

2.1. Introducción

Vamos a definir algunos conceptos que utilizaremos a lo largo del capítulo.

Definición 2.1.1. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $X : \Omega \to \mathbb{R}$ una aplicación. Entonces X es una *variable aleatoria* sobre el espacio medible $(\Omega, \mathcal{F})$ si $\forall B \in \beta, X^{-1}(B) \in \mathcal{F}$.

Definición 2.1.2. Un *proceso estocástico* es una colección de variables aleatorias $\{X_t : t \in T\}$ ordenadas según el subíndice t que en general se suele identificar con el tiempo.

Si el conjunto T es continuo, se dice que el proceso estocástico es de *tiempo continuo*. En cambio si T es discreto (finito o infinito contable), se dice que el proceso estocástico es de *tiempo discreto*.

Los procesos estocásticos también se pueden clasificar en función del espacio de estados en el que toman valores las variables aleatorias. Si el espacio de estados es continuo, se dice que el proceso estocástico es de *estado continuo*. En cambio si el espacio de estados es discreto, se dice que el proceso estocástico es de *estado discreto*.

Definición 2.1.3. Sea $S = \{s_0, s_1, s_2, \ldots\}$ un conjunto contable llamado espacio de estados y sea $\{X_n\}_{n=0,1,\ldots}$ una secuencia de variables aleatorias que toman valores en S y que cumplen la siguiente propiedad de probabilidad:

$$P(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = i_{n+1} \mid X_n = i_n),$$
$$\forall n \geq 0, \forall i_0, \ldots, i_n, i_{n+1} \in S \text{ si } P(X_n = i_n, \ldots, X_0 = i_0) > 0. \quad (2.1)$$

La propiedad (2.1) se conoce como la *propiedad de Markov* y un proceso estocástico $\{X_n\}_{n=0,1,\ldots}$ que cumple esta propiedad es una *cadena de Markov de tiempo discreto*.

Tal y como dice [11], la propiedad de Markov se interpreta como "la futura evolución probabilística del proceso está determinada una vez que se conoce el *pasado inmediato*".

Observaciones. (i) Trabajaremos con cadenas de Markov de tiempo discreto.

(ii) El conjunto S es el *espacio de estados* y puede ser finito o infinito contable, pero a partir de este momento consideraremos que S es finito. Entonces diremos que la cadena de Markov es *finita*.

(iii) Para facilitar la notación, vamos a suponer que el conjunto de estados está formado por números enteros no negativos, es decir, $S = \{0, 1, 2, \ldots, r\}$.

Definición 2.1.4. Dada una cadena de Markov $\{X_n\}_{n=0,1,\ldots}$ que toma valores en el espacio de estados S y dados $i, j \in S$, la *probabilidad de transición* del estado i al estado j *en el instante* n es $P(X_n = j \mid X_{n-1} = i)$, y se denota mediante $p_{ij}(n)$. $P_n = [p_{ij}(n)]_{i,j \in S}$ es la *matriz de transición* de la cadena de Markov *en el instante* n.

Definición 2.1.5. Sea P_n la matriz de transición de una cadena de Markov en el instante n. Si $P_1 = P_2 = ... = P_n = \ldots$, se dice que la cadena de Markov es *homogénea*. De lo contrario, la cadena de Markov es *no homogénea*.

Observaciones. (i) Trabajaremos con cadenas de Markov homogéneas, por tanto, denotaremos la probabilidad de transición del estado i al estado j mediante p_{ij} y $P = [p_{ij}]_{i,j \in S}$ será la matriz de transición.

(ii) Nótese que la matriz de transición P es una matriz estocástica. Como los elementos de P son probabilidades, son no negativos y además, la suma de los elementos de cada fila es uno ya que

$$\begin{aligned}\sum_{j \in S} p_{ij} &= \sum_{j \in S} P(X_n = j \mid X_{n-1} = i) = P(\uplus_{j \in S} X_n = j \mid X_{n-1} = i) \\ &= P(X_n \in S \mid X_{n-1} = i) = 1, \quad \forall i \in S.\end{aligned}$$

Definición 2.1.6. Dada una cadena de Markov $\{X_n\}_{n=0,1,\ldots}$ que toma valores en el espacio de estados S y dados $i,j \in S$ y $m \geq 0$, la *probabilidad de transición en m pasos* del estado i al estado j es $P(X_{n+m} = j \mid X_n = i)$ y se denota mediante $p_{ij}^{(m)}$.

Observación 2.1.1. Nótese que $p_{ii}^{(0)} = P(X_n = i \mid X_n = i) = 1$ para todo $i \in S$ y $p_{ij}^{(0)} = P(X_n = j \mid X_n = i) = 0$ para todo $i,j \in S$ tal que $i \neq j$.

Lema 2.1.1. *Dada una cadena de Markov $\{X_n\}_{n=0,1,\ldots}$ que toma valores en el espacio de estados S, para todo $i,j,k \in S$ y $n,m \geq 1$ se tiene que $P(X_{n+m} = j \mid X_n = k, X_0 = i) = P(X_{n+m} = j \mid X_n = k)$.*

Demostración. Vamos a demostrarlo para $m = 1$.

$$P(X_{n+1} = i_{n+1} \mid X_n = i_n, X_0 = i_0) = \frac{P(X_{n+1} = i_{n+1}, X_n = i_n, X_0 = i_0)}{P(X_n = i_n, X_0 = i_0)}. \quad (2.2)$$

Introducimos el suceso seguro $\Omega = (\cup_{i_{n-1}\in S} X_{n-1} = i_{n-1}, \ldots, \cup_{i_1 \in S} X_1 = i_1)$. Luego

$$\begin{aligned}(2.2) &= \frac{P(X_{n+1} = i_{n+1}, X_n = i_n, \cup_{i_{n-1}\in S} X_{n-1} = i_{n-1}, \ldots, \cup_{i_1\in S} X_1 = i_1, X_0 = i_0)}{P(X_n = i_n, \cup_{i_{n-1}\in S} X_{n-1} = i_{n-1}, \ldots, \cup_{i_1\in S} X_1 = i_1, X_0 = i_0)} \\ &= \frac{\sum_{i_1\in S}\sum_{i_2\in S}\cdots\sum_{i_{n-1}\in S} P(X_{n+1} = i_{n+1}, X_n = i_n, \ldots, X_0 = i_0)}{\sum_{i_1\in S}\sum_{i_2\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)}.\end{aligned}$$

Usamos el teorema de la probabilidad compuesta:

$$(2.2) = \frac{\sum_{i_1\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, \ldots, X_0 = i_0)\cdot P(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0)}{\sum_{i_1\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, \ldots, X_0 = i_0)}.$$

Usamos la propiedad de Markov:

$$(2.2) = \frac{\sum_{i_1\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, \ldots, X_0 = i_0)\cdot P(X_{n+1} = i_{n+1} \mid X_n = i_n)}{\sum_{i_1\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, \ldots, X_0 = i_0)}.$$

Sacamos factor común,

$$\begin{aligned}(2.2) &= \frac{\sum_{i_1\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, \ldots, X_0 = i_0)}{\sum_{i_1\in S}\cdots\sum_{i_{n-1}\in S} P(X_n = i_n, \ldots, X_0 = i_0)} P(X_{n+1} = i_{n+1} \mid X_n = i_n) \\ &= P(X_{n+1} = i_{n+1} \mid X_n = i_n).\end{aligned}$$

□

Proposición 2.1.2. (*Ecuación de Chapman-Kolmogorov). Dada una cadena de Markov $\{X_n\}_{n=0,1,\ldots}$ que toma valores en el espacio de estados S, para todo $i,j \in S$ se tiene que*

$$p_{ij}^{(m+n)} = \sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(n)}, \ \forall n, m \geq 1.$$

Demostración. Para demostrar la ecuación de Chapman-Kolmogorov, utilizaremos la propiedad de Markov y el Lema 2.1.1.

$$\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j \mid X_0 = i) = P((X_{m+n} = j) \cap (\cup_{k\in S} X_m = k) \mid X_0 = i) \\
&= P(\cup_{k\in S}(X_{m+n} = j, X_m = k) \mid X_0 = i) \\
&= \sum_{k\in S} P(X_{m+n} = j, X_m = k \mid X_0 = i) \\
&= \sum_{k\in S} P(X_m = k \mid X_0 = i) \cdot P(X_{m+n} = j \mid X_m = k, X_0 = i) \\
&= \sum_{k\in S} P(X_m = k \mid X_0 = i) \cdot P(X_{m+n} = j \mid X_m = k) = \sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(n)}.
\end{aligned}$$

□

A lo largo del capítulo vamos a ir completando un ejemplo que se propone en [5].

Ejemplo 2.1.1. Sea $\{X_n\}_{n=0,1,\ldots}$ una cadena de Markov que toma valores en el espacio de estados $S = \{1,2,3\}$. Para cada $n \in \{0,1,\ldots\}$ X_n es una variable aleatoria que representa la clase social a la que pertenece una familia en la n-ésima generación. Si $X_n = 1$ la n-ésima generación es de clase baja, si $X_n = 2$ de clase media y por último si $X_n = 3$ de clase alta. Sea

$$P = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{bmatrix}$$

la matriz de transición. Si en la actualidad una familia es de clase media, ¿cuál es la probabilidad de que la segunda generación sea de clase alta?

Queremos saber cuál es la probabilidad de pasar del estado 2 (clase media) al estado 3 (clase alta) en dos pasos (2 generaciones). Por lo tanto aplicamos la ecuación de Chapman-Kolmogorov:

$$p_{23}^{(2)} = \sum_{k\in S} p_{2j}p_{j3} = p_{21}p_{13} + p_{22}p_{23} + p_{23}p_{33} = 0{,}3 \cdot 0{,}1 + 0{,}5 \cdot 0{,}2 + 0{,}2 \cdot 0{,}4 = 0{,}21.$$

Teorema 2.1.3. *Sea $\{X_n\}_{n=0,1,\ldots}$ una cadena de Markov que toma valores en el espacio de estados S y sea P su matriz de transición. Dados $i,j \in S$ la probabilidad de transición en m pasos del estado i al estado j, $P(X_{n+m} = j \mid X_n = i) = p_{ij}^{(m)}$ es el (i,j)-ésimo elemento de P^m.*

Demostración. Se define $A_m = [p_{ij}^{(m)}]_{i,j\in S}$. Denotaremos mediante $(B)_{i,j}$ el (i,j)-ésimo elemento de cualquier matriz B. Queremos demostrar que $A_m = P^m$, $\forall m \in \mathbb{N}$. Lo demostramos por inducción sobre m:

- $m = 1$, $(A_1)_{i,j} = p_{ij}$, luego $A_1 = P = P^1$.
- Hipótesis de inducción: Supongamos que $A_{m-1} = P^{m-1}$.
- Por la ecuación de Chapman-Kolmogorov, se tiene que

$$(A_m)_{i,j} = p_{ij}^{(m)} = \sum_{k\in S} p_{ik}^{(m-1)} p_{kj} = \sum_{k\in S} (A_{m-1})_{i,k} \cdot p_{kj} = (A_{m-1}P)_{i,j}. \quad (2.3)$$

  Aplicando la hipótesis de inducción en la igualdad (2.3), se tiene que

$$(A_m)_{i,j} = (A_{m-1}P)_{i,j} = (P^{m-1}P)_{i,j} = (P^m)_{i,j}$$

  Por tanto, $A_m = P^m$.

□

Observación 2.1.2. Para cualquier $m \in \mathbb{N}$, P^m es una matriz estocástica. Como los elementos de P^m son probabilidades, son no negativos y además, la suma de los elementos de cada fila es uno ya que

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

Ejemplo 2.1.2. Veamos con la matriz de transición P del Ejemplo 2.1.1 que P^2 es una matriz estocástica.

$$P^2 = PP = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{bmatrix} \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{bmatrix} = \begin{bmatrix} 0{,}57 & 0{,}28 & 0{,}15 \\ 0{,}4 & 0{,}39 & 0{,}21 \\ 0{,}34 & 0{,}4 & 0{,}26 \end{bmatrix}.$$

Todos los elementos de P^2 son no negativos y es fácil ver que la suma de todos los elementos de cada fila es uno. Por lo tanto, P^2 es una matriz estocástica. Además, tal y como se ha demostrado en el Teorema 2.1.3 podemos observar que el (2, 3)-ésimo elemento de P^2 coincide con $p_{23}^{(2)}$ que hemos calculado en el Ejercicio 2.1.1 previamente.

2.2. Clasificación de estados

Vamos a ver lo que significa que un estado *conduzca* a otro estado y que un estado esté *comunicado* con otro estado. Gracias a estas definiciones distinguiremos dos tipos de estados: *transitorios* y *recurrentes*. Además demostraremos que la relación *comunicarse* es una relación de equivalencia.

Definición 2.2.1. Dados $i, j \in S$ se dice que el estado i *conduce* al estado j o que j es *accesible* desde el estado i, si existe $n \geq 0$ tal que $p_{ij}^{(n)} > 0$. En tal caso, escribiremos $i \to j$.

Definición 2.2.2. Dados $i, j \in S$ se dice que el estado i se *comunica* con el estado j $(i \leftrightarrow j)$ si i conduce a j y j conduce a i. Es decir, si existen $n, m \geq 0$ tales que $p_{ij}^{(n)} > 0$ y $p_{ji}^{(m)} > 0$.

Observación 2.2.1. En las dos definiciones anteriores podríamos haber sustituido la condición $n \geq 0$ por $n \geq 1$ cuando $i \neq j$ pues $p_{ij}^{(0)} = 0$.

Proposición 2.2.1. *La relación $\leftrightarrow$ es una relación de equivalencia.*

Demostración. Veamos que la relación $\leftrightarrow$ es reflexiva, simétrica y transitiva:

(i) $p_{ii}^{(0)} = P(X_n = i \mid X_n = i) = \frac{P(X_n=i, X_n=i)}{P(X_n=i)} = \frac{P(X_n=i)}{P(X_n=i)} = 1 > 0 \Rightarrow i \leftrightarrow i$. Por tanto, $\leftrightarrow$ es reflexiva.

(ii) Si $i \leftrightarrow j \Rightarrow i \to j \wedge j \to i \Rightarrow j \leftrightarrow i$. Por tanto, $\leftrightarrow$ es simétrica.

(iii) Supongamos que $i \leftrightarrow j$ y $j \leftrightarrow k$. Entonces, existen $n, m, r, s \geq 1$ tales que $p_{ij}^{(n)} > 0$, $p_{ji}^{(m)} > 0$, $p_{jk}^{(r)} > 0$ y $p_{kj}^{(s)} > 0$. Por la ecuación de Chapman-Kolmogorov tenemos que:

$$p_{ik}^{(n+r)} = \sum_{t \in S} p_{it}^{(n)} p_{tk}^{(r)} \geq p_{ij}^{(n)} p_{jk}^{(r)} > 0 \Rightarrow i \to k,$$

$$p_{ki}^{(s+m)} = \sum_{t \in S} p_{kt}^{(s)} p_{ti}^{(r)} \geq p_{kj}^{(s)} p_{ji}^{(m)} > 0 \Rightarrow k \to i.$$

Por tanto, $i \leftrightarrow k$ y $\leftrightarrow$ es transitiva.

□

Ejemplo 2.2.1. Retomando el Ejemplo 2.1.1, como la matriz de transición es

$$P = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{bmatrix},$$

todos los estados se comunican entre sí ya que $p_{ij} > 0$ para cualquier $i, j \in \{1, 2, 3\}$.

Definición 2.2.3. Se dice que una cadena de Markov es *irreducible* si su matriz de transición es irreducible.

Ejemplo 2.2.2. La matriz de transición del Ejemplo 2.1.1 es

$$P = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{bmatrix}.$$

Como se puede apreciar en la Figura 2.1 el grafo dirigido asociado a P es fuertemente conexo. Por tanto, por el Teorema 1.1.11 P es una matriz irreducible. Entonces la cadena de Markov es irreducible.

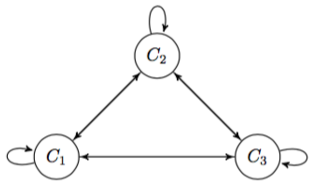

Figura 2.1: Grafo dirigido asociado a P

Proposición 2.2.2. *Una cadena de Markov es irreducible si y solo si todos sus estados están comunicados entre sí.*

Demostración. Supongamos que la cadena es irreducible. Sea $P = [p_{ij}]_{i,j \in S}$ la matriz de transición. Como la cadena de Markov es irreducible, P es una matriz irreducible. Por la Proposición 1.1.9, para cada $i, j \in S$, existe $q \in \mathbb{N}$ tal que $(P^q)_{i,j} > 0$ y por el Teorema 2.1.3 $p_{ij}^{(q)} = (P^q)_{i,j} > 0$. Por tanto, existen $q_1, q_2 \in \mathbb{N}$ tales que $(P^{q_1})_{i,j} = p_{ij}^{(q_1)} > 0$ y $(P^{q_2})_{j,i} = p_{ji}^{(q_2)} > 0$. Es decir, $i \leftrightarrow j$.

Supongamos que todos los estados están comunicados entre sí. Entonces para todo $i, j \in S$ existe $m \geq 1$ tal que $p_{ij}^{(m)} > 0$. Por la Proposición 1.1.10 existe una secuencia de m líneas dirigidas en el grafo dirigido asociado a P que conectan los nodos P_i y P_j. Por tanto, el grafo dirigido asociado a P es fuertemente conexo y por el Teorema 1.1.11, P es una matriz irreducible. Entonces la cadena de Markov es irreducible. □

Corolario 2.2.3. $\{X_n\}_{n=0,1,\ldots}$ *es una cadena de Markov irreducible si y solo si todos sus estados están en una única clase de equivalencia.*

Demostración. Se deduce de la Proposición 2.2.2. □

Definición 2.2.4. Dado $i \in S$ se dice que el estado i es *transitorio* si existe $j \in S$ tal que $i \to j$ pero $j \nrightarrow i$.

Proposición 2.2.4. *Dado $i \in S$, las siguientes afirmaciones son equivalentes:*

(i) *i es transitorio.*

(ii) $f_i = P(\textit{Volver a tomar el valor } i \mid X_0 = i) < 1$.

Demostración. Demostraremos solo la implicación que utilizaremos.
$(i) \Rightarrow (ii)$: Como $i \in S$ es transitorio, existe $j \in S$ tal que $i \to j$ pero $j \nrightarrow i$. Es decir, existe $m \in \mathbb{N}$ tal que $p_{ij}^{(m)} > 0$ y para cada $q \in \mathbb{N}$, $p_{ji}^{(q)} = 0$. Vamos a definir dos sucesos: $A \equiv$ " entrar en j " y $B \equiv$ " no volver a i ".

Como $j \nrightarrow i$, tenemos que $A \subset B$. Luego $P(A \mid X_0 = i) \leq P(B \mid X_0 = i)$. Queremos ver que $P(B \mid X_0 = i) > 0$. Para ello veamos que $P(A \mid X_0 = i) > 0$. $P(A \mid X_0 = i) = P(\cup_{k=1}^{\infty} X_k = j \mid X_0 = i) \geq P(X_m = j \mid X_0 = i) = p_{ij}^{(m)} > 0$. Por tanto,

$$\begin{aligned} f_i &= P(Volver\ a\ tomar\ el\ valor\ i \mid X_0 = i) = 1 - P(No\ volver\ a\ i \mid X_0 = i) \\ &= 1 - P(B \mid X_0 = i) < 1. \end{aligned}$$

□

Observación 2.2.2. Si $i \in S$ es un estado transitorio, se tiene que $f_i < 1$. Por la propiedad de Markov, $P(Volver\ a\ tomar\ el\ valor\ i\ k\ veces \mid X_0 = i) = f_i^k$. Como $f_i < 1$, $f_i^k \xrightarrow{k \to \infty} 0$. Por tanto, la cadena tomará el valor i un número finito de veces.

Definición 2.2.5. Dado $i \in S$ se dice que el estado i es *recurrente* si no es transitorio, es decir, si para todo $j \in S$ tal que $i \to j$ se tiene que $j \to i$.

Ejemplo 2.2.3. En el Ejemplo 2.1.1 todos los estados son recurrentes ya que para todo $i, j \in S$, $p_{ij} > 0$.

Definición 2.2.6. Dado $i \in S$ se dice que i es un estado *absorbente* si $p_{ii} = 1$.

Proposición 2.2.5. *Sea $\{X_n\}_{n=0,1,\ldots}$ una cadena de Markov. Entonces existe al menos un estado recurrente.*

Demostración. Sea $S = \{1, \ldots, r\}$ el espacio de estados. Supongamos por reducción al absurdo que todos los estados son transitorios. Por la Observación 2.2.2 para cada $i \in S$, la cadena tomará el valor i un número finito de veces. Sea N_i el número de pasos desde el principio hasta que se toma el valor i por última vez. Sea $N = \max_{1 \leq i \leq r} N_i$. Entonces, después del paso N, la cadena no tomará ningún valor de S, pero esto es imposible. Por tanto, al menos uno de los estados debe ser recurrente. □

Proposición 2.2.6. *Dados $i, j \in S$ si i es recurrente e $i \leftrightarrow j$ entonces j también es recurrente.*

Demostración. Sea $k \in S$ tal que $j \to k$. Queremos ver que $k \to j$, es decir, que existe $n \in \mathbb{N}$ tal que $p_{kj}^{(n)} > 0$. Como $j \to k$, existe $q \in \mathbb{N}$ tal que $p_{jk}^{(q)} > 0$ y como $i \leftrightarrow j$, existe $m \in \mathbb{N}$ tal que $p_{ij}^{(m)} > 0$. Por la ecuación de Chapman-Kolmogorov tenemos que:

$$p_{ik}^{(m+q)} = \sum_{t \in S} p_{it}^{(m)} p_{tk}^{(q)} \geq p_{ij}^{(m)} p_{jk}^{(q)} > 0.$$

Por tanto, $i \to k$, y como i es recurrente, $k \to i$. Por tanto, tenemos que $k \leftrightarrow i$ e $i \leftrightarrow j$. Por la propiedad transitiva de la relación $\leftrightarrow$, tenemos que $k \leftrightarrow j$. Por tanto, $k \to j$. □

Corolario 2.2.7. *Todos los estados de una cadena de Markov irreducible son recurrentes.*

Demostración. Por la Proposición 2.2.5 sabemos que al menos un estado es recurrente ($i \in S$). Por la Proposición 2.2.2 sabemos que todos los estados están comunicados entre sí. En concreto, para todo $j \in S$ tenemos que $i \leftrightarrow j$. Finalmente, por la Proposición 2.2.6 j también es un estado recurrente. □

Definición 2.2.7. Dado $i \in S$ se considera el conjunto $\{n \geq 1 : p_{ii}^{(n)} > 0\}$. Si este conjunto es vacío, se dice que i tiene *periodo* $d = d(i) = 0$. Si el conjunto es distinto del vacío, se define el *periodo* de i como $d = d(i) = \text{m.c.d.}\{n \geq 1 : p_{ii}^{(n)} > 0\}$.

Definición 2.2.8. Dado $i \in S$ se dice que el estado i es *aperiódico* si $d = d(i) = 1$. Además, se dice que una cadena de Markov es *aperiódica* si todos sus estados son aperiódicos.

Ejemplo 2.2.4. Vamos a calcular el periodo del estado 1 del Ejemplo 2.1.1. Se cumple que $d(1) = \text{m.c.d.}\{n \geq 1 : p_{11}^{(n)} > 0\}$. Al ser $p_{11} = p_{11}^{(1)} > 0$, tenemos que $d(1) \mid 1$ y como $d(1) \geq 1$ se tiene que $d(1) = 1$. Como $p_{22} > 0$ y $p_{33} > 0$ se demuestra de la misma manera que $d(2) = 1$ y $d(3) = 1$. Por tanto, todos los estados de S son aperiódicos y la cadena de Markov es aperiódica.

Proposición 2.2.8. *Dados $i, j \in S$, si $i \leftrightarrow j$, entonces, $d(i) = d(j)$.*

Demostración. Si $i = j$ el resultado es trivial. Entonces, supongamos que $i \neq j$. Como $i \leftrightarrow j$, existen $m, n \in \mathbb{N}$ tales que $p_{ij}^{(m)} > 0$ y $p_{ji}^{(n)} > 0$. Por la ecuación de Chapman-Kolmogorov tenemos que $p_{ii}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{ki}^{(n)} \geq p_{ij}^{(m)} p_{ji}^{(n)} > 0$. Sea $s \geq 1$ cualquiera tal que $p_{ii}^{(s)} > 0$. Por la ecuación de Chapman-Kolmogorov:

$$p_{jj}^{(m+n+s)} = \sum_{k \in S} p_{jk}^{(n)} p_{kj}^{(m+s)} \geq p_{ji}^{(n)} p_{ij}^{(m+s)} = p_{ji}^{(n)} \sum_{k \in S} p_{ik}^{(s)} p_{kj}^{(m)} \geq p_{ji}^{(n)} p_{ii}^{(s)} p_{ij}^{(m)} > 0.$$

De la misma manera se puede demostrar que $p_{jj}^{(m+n+2s)} > 0$. Como $d(j) = \text{m.c.d.}\{n \geq 1 : p_{jj}^{(n)} > 0\}$, tenemos que $d(j) \mid m + n + s$ y $d(j) \mid m + n + 2s$. Por lo tanto, existen $b > a > 0$ tales que $d(j)b = m + n + 2s$ y $d(j)a = m + n + s$. Restando las dos expresiones, tenemos que $d(j)(b - a) = s$. Como $b > a$, $d(j) \mid s$ para cualquier $s \in \mathbb{N}$ tal que $p_{ii}^{(s)} > 0$. Y como $d(i) = \text{m.c.d.}\{n \geq 1 : p_{ii}^{(n)} > 0\}$, $d(j) \mid d(i)$. Por simetría, $d(i) \mid d(j)$, luego $d(i) = d(j)$. □

Observación 2.2.3. Todos los estados de una misma clase de equivalencia tienen el mismo periodo.

2.3. Comportamiento de una cadena a largo plazo

Queremos ver qué sucede con $p_{ij}^{(m)}$ cuando m tiende a infinito. Como en el Teorema 2.1.3 hemos visto que para cada $m \in \mathbb{N}$ $p_{ij}^{(m)}$ y el (i,j)-ésimo elemento de la matriz P^m son iguales, vamos a ver si existe el límite de P^m cuando m tiende a infinito.

Lema 2.3.1. *Sea $d = m.c.d.\{n_1, ..., n_k\}$. Entonces, existe un entero positivo, h, tal que para cualquier número entero $m \geq h$, $md = \sum_{j=1}^{k} c_j n_j$ para algunos enteros no negativos $c_1, ..., c_k$.*

Demostración. Por la identidad de Bézout, existen enteros $\pi_1, ..., \pi_k$ tales que $d = \sum_{j=1}^{k} \pi_j n_j$. Dividiendo entre d, $1 = \sum_{j=1}^{k} \pi_j (\frac{n_j}{d})$. Sea $d^* = \sum_{j=1}^{k} |\pi_j| n_j$ y se define $h = (d^*)^2$. Por la manera en la que se ha definido h, para cada número entero $m \geq h$, se tiene que $m = qd^* + r$ siendo $q \geq d^*$ y $d^* > r \geq 0$. Entonces, como $d^* = \sum_{j=1}^{k} |\pi_j| n_j$ y $1 = \sum_{j=1}^{k} \pi_j (\frac{n_j}{d})$:

$$md = (qd^* + r)d = qd \sum_{j=1}^{k} |\pi_j| n_j + rd \sum_{j=1}^{k} \pi_j (\frac{n_j}{d}) = \sum_{j=1}^{k} (qd|\pi_j| + r\pi_j) n_j.$$

Como $q > r$ y $d \geq 1$, se cumple que $qd|\pi_j| \geq q|\pi_j| > r|\pi_j| \geq -r\pi_j$. Por lo tanto, $c_j = (qd|\pi_j| + r\pi_j) > 0$. □

Lema 2.3.2. *Sea $\{X_n\}_{n=0,1,...}$ una cadena de Markov con matriz de transición P y sean q y m enteros no negativos cualesquiera. Entonces $p_{jj}^{(qm)} \geq (p_{jj}^{(m)})^q$ para todo $j \in S$.*

Demostración. Sean q y m enteros no negativos cualesquiera. Por el Lema 2.1.1:

$$\begin{aligned}
p_{jj}^{(qm)} &= P(X_{qm} = j \mid X_0 = j) \\
&\geq P(X_{qm} = j, X_{(q-1)m} = j, \ldots, X_{2m} = j, X_m = j \mid X_0 = j) \\
&= P(X_m = j \mid X_0 = j) \cdot P(X_{2m} = j \mid X_0 = j, X_m = j) \ldots P(X_{qm} = j \mid X_0 = j, \\
&\quad X_m = j, \ldots, X_{(q-1)m} = j) \\
&= P(X_m = j \mid X_0 = j) \cdot P(X_{2m} = j \mid X_m = j) \ldots P(X_{qm} = j \mid X_{(q-1)m} = j) \\
&= (p_{jj}^{(m)})^q.
\end{aligned}$$

□

Teorema 2.3.3. *Sea $\{X_n\}_{n=0,1,...}$ una cadena de Markov con matriz de transición P. Entonces la cadena es irreducible y aperiódica si y solo si $P^N > 0$ para algún $N \in \mathbb{N}$.*

Demostración. Supongamos que la cadena es irreducible y aperiódica. Como P es irreducible, por la Proposición 2.2.2, todos los estados de la cadena están comunicados entre sí. Es decir, para cada $i, j \in S$ existe $m = m(i, j)$ tal que $p_{ij}^{(m)} > 0$. Como la cadena es aperiódica, $d = d(i) = \text{m.c.d.}\{n \geq 1 : p_{ii}^{(n)} > 0\} = 1$, $\forall i \in S$. Sea $p_{ii}^{(n_s)} > 0$, para $s = 1, 2, ..., t$. Sin pérdida de generalidad, supongamos que $d(i) = \text{m.c.d.}\{n_s : s = 1, ..., t\}$. Por el Lema 2.3.1, existe un entero positivo $h = h(i)$ tal que para cualquier $n \geq h(i)$, $n = \sum_{s=1}^{t} \pi_s n_s$, siendo π_s un entero no negativo. Entonces por la ecuación de Chapman-Kolmogorov y el Lema 2.3.2:

$$\begin{aligned} p_{ij}^{(m+n)} &= \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)} \geq p_{ii}^{(n)} p_{ij}^{(m)} = p_{ii}^{(\sum_{s=1}^{t} \pi_s n_s)} p_{ij}^{(m)} \\ &= p_{ij}^{(m)} \prod_{s=1}^{t} p_{ii}^{(\pi_s n_s)} \geq p_{ij}^{(m)} \prod_{s=1}^{t} (p_{ii}^{(n_s)})^{\pi_s} > 0. \end{aligned}$$

Entonces, para todo $n \geq h(i)$, se tiene que $p_{ij}^{(m+n)} > 0$. Sea $N = \text{máx}_{k,l \in S}\{m(k,l) + h(k)\}$. Entonces, $p_{ij}^{(N)} > 0$ para todo $i, j \in S$.

Recíprocamente supongamos que $P^N > 0$ para algún $N \in \mathbb{N}$. Por lo tanto, para todo $i, j \in S$, $p_{ij}^{(N)} > 0$. Entonces, todos los estados están comunicados entre sí y por la Proposición 2.2.2, la cadena de Markov es irreducible y por tanto, la matriz de transición P es irreducible. Veamos por inducción que $P^{N+m} > 0$ para cualquier $m \geq 1$:

- $m = 1$: $P^{N+1} = P^N P$. Supongamos por recucción al absurdo que P^{N+1} no es positiva. Como $P^N > 0$, la única opción para que P^{N+1} no sea positiva es que una columna de P esté formada por ceros. Entonces, por la Proposición 1.1.5 el grafo dirigido asociado a la matriz P no es fuertemente conexo y por el Teorema 1.1.11 , P no es una matriz irreducible, lo que es una contradicción. Así, $P^{N+1} > 0$.

- Hipótesis de inducción: Supongamos que $P^{N+m-1} > 0$.

- $P^{N+m} = P^{N+m-1} P$. Razonando de la misma manera que en el caso $m = 1$, se consigue que $P^{N+m} > 0$.

Como $P^{N+m} > 0$ para cualquier $m \geq 1$, tenemos que para todo $i, j \in S$, $p_{ij}^{(N+m)} > 0$, para cualquier $m \geq 1$. En particular, para todo $i \in S$, $p_{ii}^{(N+m)} > 0$ para cualquier $m \in \mathbb{N}$. Entonces, para todo $i \in S$, $d(i) = 1$. Luego P es aperiódica. □

Lema 2.3.4. *Sea $P = [p_{ij}]_{i,j \in S}$ la matriz de transición de una cadena de Markov. Si $\lim_{m \to +\infty} P^m = Q$ existe, Q es una matriz estocástica.*

Demostración. Supongamos que $\lim\limits_{m\to+\infty} P^m = Q$ existe. Por el Teorema 2.1.3 $p_{ij}^{(m)}$ es el (i,j)-ésimo elemento de la matriz P^m. Por la Observación 2.1.2 sabemos que para cada $m \in \mathbb{N}$ P^m es una matriz estocástica. Luego $\sum_{j=1}^{n} p_{ij}^{(m)} = 1$ para todo $i \in S$. Como $\lim\limits_{m\to+\infty} p_{ij}^{(m)} = q_{ij}$ para todo $i,j \in S$ tenemos que $\sum_{j\in S} q_{ij} = \sum_{j\in S} \lim\limits_{m\to+\infty} p_{ij}^{(m)} = \lim\limits_{m\to+\infty} \sum\limits_{j\in S} p_{ij}^{(m)} = \lim\limits_{m\to+\infty} 1 = 1$ para cada $i \in S$. Así Q es una matriz estocástica. □

Definición 2.3.1. Un *vector de probabilidad* es un vector cuyas componentes son no negativas y suman 1.

Definición 2.3.2. Dados $A \in \mathbb{R}^{n\times n}$ y λ un valor propio de A, se dice que un vector no nulo $y \in \mathbb{R}^n$ es un *vector propio por la izquierda* de A asociado a λ si $y^T A = y^T \lambda$.

Teorema 2.3.5. *Sea P la matriz de transición de una cadena de Markov irreducible y aperiódica. Entonces, $\lim\limits_{k\to+\infty} P^k = Q$ existe. Además, $PQ = QP = Q = Q^2$. Todas las filas de la matriz Q son iguales. Cualquier fila de Q viene dada por el único vector de probabilidad u tal que $uP = u$.*

Demostración. Por el Teorema 2.3.3 sabemos que existe $N \in \mathbb{N}$ tal que $P^N > 0$. Además, como P es una matriz estocástica, por el Teorema 1.2.1 sabemos que 1 es un valor propio de P y que el radio espectral de P es 1. Entonces, por el Teorema 1.1.15 $\lim\limits_{k\to+\infty} P^k = Q$ existe. Veamos que $PQ = Q = QP$:

$$PQ = P \lim_{k\to+\infty} P^k = \lim_{k\to+\infty} (PP^k) = \lim_{k\to+\infty} P^{k+1} = Q,$$

$$QP = (\lim_{k\to+\infty} P^k)P = \lim_{k\to+\infty} (P^k P) = \lim_{k\to+\infty} P^{k+1} = Q.$$

Veamos que todas las filas de Q son iguales. Por el Lema 2.3.4 sabemos que Q es una matriz estocástica. Sean q_i y q_j la i-ésima y j-ésima filas de Q respectivamente. Entonces $q_i, q_j \neq 0$ por ser Q una matriz estocástica y como $QP = Q$, $q_i P = q_i$ y $q_j P = q_j$. Como P es una matriz no negativa e irreducible y su radio espectral es 1, por el Teorema 1.1.14 sabemos que la multiplicidad geométrica del valor propio 1 es 1. Por tanto, $q_i = \alpha q_j$ para algún $\alpha \in \mathbb{R}$. Como Q es estocástica, $1 = \sum_{k=1}^{n} q_{ik} = \sum_{k=1}^{n} \alpha q_{jk} = \alpha \sum_{k=1}^{n} q_{jk} = \alpha$. Entonces, $q_i = q_j$ y queda demostrado que todas las filas de Q son iguales.

Como Q es una matriz estocástica y todas sus filas son iguales, tenemos que:

$$Q^2 = \begin{bmatrix} a_1 & \cdots & a_n \\ \vdots & \ddots & \vdots \\ a_1 & \cdots & a_n \end{bmatrix}^2 = \begin{bmatrix} a_1 \sum_{k=1}^{n} a_k & \cdots & a_n \sum_{k=1}^{n} a_k \\ \vdots & \ddots & \vdots \\ a_1 \sum_{k=1}^{n} a_k & \cdots & a_n \sum_{k=1}^{n} a_k \end{bmatrix} = \sum_{k=1}^{n} a_k \begin{bmatrix} a_1 & \cdots & a_n \\ \vdots & \ddots & \vdots \\ a_1 & \cdots & a_n \end{bmatrix} = Q$$

Finalmente, veamos que cualquier fila de Q viene dada por el único vector de probabilidad u tal que $uP = u$. Supongamos por reducción al absurdo que existen dos vectores de probabilidad u_1 y u_2 tales que $u_1 \neq u_2$ y $u_1 P = u_1$ y $u_2 P = u_2$. Entonces, como la multiplicidad geométrica del valor propio 1 es 1, tenemos que $u_1 = \alpha u_2$ para algún $\alpha \in \mathbb{R}$. Como son vectores de probabilidad, $1 = \sum_{k=1}^{n} u_{1k} = \sum_{k=1}^{n} \alpha u_{2k} = \alpha \sum_{k=1}^{n} u_{2k} = \alpha$. Pero esto es imposible ya que $u_1 \neq u_2$. Por tanto, $u_1 = u_2$ y queda demostrado que cualquier fila de Q viene dada por el único vector de probabilidad u tal que $uP = u$. □

Una consecuencia del resultado anterior es que si se tiene una cadena de Markov finita, irreducible y aperiódica, la probabilidad de que a largo plazo la cadena se halle en el estado k no depende del punto de partida de la cadena.

Corolario 2.3.6. *Para una cadena de Markov irreducible y aperiódica siempre existe* $\lim\limits_{m \to +\infty} p_{ij}^{(m)}$ *y no depende de* i*, es decir,*

$$\lim_{m \to +\infty} p_{ij}^{(m)} = u_j \text{ para todo } i, j \in S.$$

Además, $\{u_j\}_{j \in S}$ *es la única familia de soluciones no negativas de*

$$\begin{cases} u_j = \sum_{i \in S} u_i p_{ij} \\ \sum_{j \in S} u_j = 1 \end{cases}.$$

A $\{u_j\}_{j \in S}$ se le llama *distribución de probabilidad límite, de estado estable o de equilibrio* de la cadena.

Ejemplo 2.3.1. Con los datos del Ejemplo 2.1.1 vamos a calcular $\lim\limits_{k \to +\infty} P^k$.

Anteriormente hemos visto que en este caso la cadena de Markov es irreducible y aperiódica. Por lo tanto, por el Teorema 2.3.5 $\lim\limits_{k \to +\infty} P^k = Q$ existe y cualquier fila de Q viene dada por el único vector de probabilidad u tal que $uP = u$. Vamos a calcular $u = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$:

$$\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}4 & 0{,}4 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$$

$$\Leftrightarrow \begin{cases} 0{,}7u_1 + 0{,}3u_2 + 0{,}2u_3 = u_1 \\ 0{,}2u_1 + 0{,}5u_2 + 0{,}4u_3 = u_2 \\ 0{,}1u_1 + 0{,}2u_2 + 0{,}4u_3 = u_3 \end{cases}$$

$$\Leftrightarrow \begin{cases} -0{,}3u_1 + 0{,}3u_2 + 0{,}2u_3 = 0 & (2.4) \\ 0{,}2u_1 - 0{,}5u_2 + 0{,}4u_3 = 0 & (2.5) \\ 0{,}1u_1 + 0{,}2u_2 - 0{,}6u_3 = 0 & (2.6) \end{cases}$$

Como la ecuación (2.6) es una combinación lineal de las dos primeras ((2.6) = −(2.4) − (2.5)),

$$\begin{cases} -0{,}3u_1 + 0{,}3u_2 + 0{,}2u_3 = 0 \\ 0{,}2u_1 - 0{,}5u_2 + 0{,}4u_3 = 0 \end{cases}.$$

Y resolviendo el sistema se consigue que $u_1 = \frac{22}{9}u_3$ y $u_2 = \frac{16}{9}u_3$ para cualquier u_3. Pero como $\sum_{j=1}^{3} u_j = 1$, tenemos que $\frac{22}{9}u_3 + \frac{16}{9}u_3 + u_3 = 1$ y se obtiene $u_3 = \frac{9}{47}$. Entonces $u = \begin{bmatrix} \frac{22}{47} & \frac{16}{47} & \frac{9}{47} \end{bmatrix}$. Por tanto

$$\lim_{k\to+\infty} P^k = Q = \begin{bmatrix} \frac{22}{47} & \frac{16}{47} & \frac{9}{47} \\ \frac{22}{47} & \frac{16}{47} & \frac{9}{47} \\ \frac{22}{47} & \frac{16}{47} & \frac{9}{47} \end{bmatrix}.$$

Esto significa que la probabilidad de que al cabo de muchos años una familia sea de clase social baja es $\frac{22}{47}$, indistintamente de cúal haya sido su estatus social en un primer momento. La probabilidad de que al cabo de muchos años una familia sea de clase social media es $\frac{16}{47}$. Y por último la probabilidad de que al cabo de muchos años una familia sea de clase social alta es $\frac{9}{47}$.

Capítulo 3

Una aplicación de las cadenas de Markov: PageRank

En Internet hay millones de páginas web, por lo que es fundamental que haya buscadores rápidos y eficientes que proporcionen al usuario las páginas que contienen la información que necesita. En este capítulo vamos a hablar sobre el *PageRank*, el método que utiliza Google para clasificar las páginas web y así poder proporcionar al usuario aquellas páginas que serán de su interés.

3.1. Introducción

Hoy en día Google es el buscador que más se utiliza en la web. Empezó a funcionar en 1998 y poco después de su puesta en marcha se convirtió en uno de los buscadores más eficientes. Cada segundo está proporionando sevicios a millones de consultas recibidas por usuarios de Internet. Google organiza una gran cantidad de información y la hace universalmente accesible. El algoritmo PageRank es la técnica que utiliza Google para clasificar las páginas web de acuerdo a su importancia. PageRank fue desarrollado por los informáticos Larry Page y Sergey Brin durante sus doctorados en la Universidad de Standford. Juntos escribieron el artículo “The Anatomy of a Large-Scale Hypertextual Web Search Engine”[3].

El artículo [3] dice así, “Google es un prototipo de buscador a gran escala que hace uso de la estructura presente en el hipertexto*. Está diseñado para rastrear e indexar la Web de manera eficiente y producir resultados de búsqueda mucho más satisfactorios que los sistemas existentes”.

* Hipertexto: Herramienta con estructura no secuencial que permite crear, agregar, enlazar y compartir información de diversas fuentes por medio de enlaces asociativos.

PageRank es un algoritmo *independiente de consulta* y *contenido*. Que sea independiente de consulta significa que el algoritmo clasifica todas las páginas sin conexión después de que el rastreador (programa que inspecciona las páginas del World Wild Web de forma metódica y automatizada) haya descargado e indexado las páginas. Que sea independiente de contenido significa que el algoritmo PageRank no incluye el contenido de una página web a la hora de clasificarla sino que utiliza la estructura de enlaces de la web para calcular su rango (PageRank). Cuando un usuario escribe un término en el buscador, el algoritmo PageRank solo encuentra las páginas en la web que coinciden con el término que se quiere consultar y las presenta al usuario en orden de su PageRank.

3.2. Uso de las cadenas de Markov en el algoritmo PageRank

El algoritmo PageRank considera la Web como un grafo dirigido cuyos nodos son las páginas web y los arcos son los enlaces que hay entre ellas. El número de páginas web en la actualidad es muy grande, pero finito. Supongamos que hay N páginas web. El PageRank de la i-ésima página ($i = 1, \ldots, N$) lo denotaremos mediante $PR(i)$ y viene dado [3] por la siguiente ecuación:

$$PR(i) = (1 - d) + \sum_{j\,:\,j \to i} PR(j) \frac{d}{C(j)}, \quad i = 1, \ldots, N \tag{3.1}$$

donde $C(j)$ es el número de enlaces que salen desde la página j, d es un parámetro de amortiguación que está entre cero y uno y $j \to i$ significa que hay un enlace desde la página j a la página i, es decir, en terminología del capítulo anterior, j conduce a i. En [3] se dice que normalmente d se establece a 0.85.

Para relacionar el PageRank con las cadenas de Markov, vamos a suponer que cada página web es un estado de la cadena. Llegaremos a la conclusión de que el PageRank de una página web es proporcinal a la probabilidad límite de dicho estado en la cadena de Markov. Entonces, cuanto mayor sea la probabilidad, mayor será el PageRank de la página web.

Supondremos que un usuario que está en una página web continúa clicando en los enlaces sin volver nunca hacia atrás. Pero puede ocurrir que el usuario se aburra y que reinicie su búsqueda sin seguir los enlaces de la página web en la que se encuentra. Para representar este fenómeno, vamos a definir un nuevo estado, la página de reinicio, y vamos a nombrarlo como el estado $i = 0$. Denotaremos mediante p_{ji} la probabilidad de ir del estado j al estado i. Asumimos que:

(i) Para cada $j = 0, \ldots, N$ la probabilidad de ir del estado j al estado 0 es $p_{j0} = 1 - d$. Es decir, la probabilidad de llegar al estado 0 no depende del estado actual.

(ii) La probabilidad de ir del estado 0 al estado j es $p_{0j} = \frac{d}{N}$ para cualquier $j = 1, \ldots, N$.

(iii) La probabilidad de ir del estado j al estado i tal que $i, j \neq 0$ es

$$p_{ji} = \begin{cases} 0 & \text{si } j \nrightarrow i \\ \frac{d}{C(j)} & \text{si } j \rightarrow i \end{cases} .$$

(iv) Nótese que $\sum_{i=0}^{N} p_{0i} = p_{00} + \sum_{i=1}^{N} p_{0i} = (1-d) + \sum_{i=1}^{N} \frac{d}{N} = 1 - d + d = 1$.

(v) Nótese que $\sum_{i=0}^{N} p_{ji} = p_{j0} + \sum_{i=1}^{N} p_{ji} = (1-d) + \sum_{i\,:\,j\rightarrow i} \frac{d}{C(j)} = 1 - d + d\frac{C(j)}{C(j)} = 1$ para todo $j = 1, \ldots, N$.

Sea u_i la probabilidad límite del estado i. Suponiendo que la cadena de Markov es irreducible y aperiódica, el Corolario 2.3.6 nos asegura que se cumplen estas ecuaciones:

$$\begin{cases} u_i = \sum_{j=0}^{N} u_j p_{ji} \\ \sum_{i=0}^{N} u_i = 1 \end{cases} .$$

Luego

$$u_0 = \sum_{j=0}^{N} u_j p_{j0} = \sum_{j=0}^{N} u_j (1-d) = (1-d) \sum_{j=0}^{N} u_j = 1 - d,$$

$$u_i = \sum_{j=0}^{N} u_j p_{ji} = u_0 p_{0i} + \sum_{j=1}^{N} u_j p_{ji} = (1-d)\frac{d}{N} + \sum_{j\,:\,j\rightarrow i} u_j \frac{d}{C(j)}, \quad i = 1, \ldots, N. \tag{3.2}$$

Multiplicando por $\frac{N}{d}$ la ecuación (3.2)

$$\frac{N}{d} \cdot u_i = (1-d) + \sum_{j\,:\,j\rightarrow i} \left(\frac{N}{d} \cdot u_j\right) \frac{d}{C(j)}.$$

Si definimos $PR(i) = \frac{N}{d} \cdot u_i$, se consigue la ecuación (3.1):

$$PR(i) = (1-d) + \sum_{j\,:\,j\rightarrow i} PR(j) \frac{d}{C(j)}, \quad i = 1, \ldots, N.$$

Entonces cuando el usuario escribe un término en el buscador Google, se le proporcionan las páginas web que coinciden con dicho término ordenadas por el PageRank. Cuanto mayor sea el Page Rank de una página web antes aparecerá en la lista de páginas que presenta Google.

Bibliografía

[1] R. B. Bapat y T. E. S. Raghavan, *Nonnegative matrices and applications*, Cambridge University Press, 1997.

[2] A. Berman y R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Society for Industrial and Applied Mathematics, 1994.

[3] S. Brin y L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Computer Network and ISDN Systems, Vol. 30, Issue 1-7, p. 107-117, 1998.

[4] C. R. Cassady y J. A. Nachlas, *Probability Models in Operations Research*, CRC Press, 2009.

[5] R. Durrett, *Essentials of Stochastic Processes*, Springer, 2012.

[6] R. A. Horn y C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1985.

[7] P. Lancaster y M. Tismenetsky, *The theory of matrices with applications*, Academic Press, 1985.

[8] H. Minc, *Nonnegarive Matrices*, John Wiley & Sons, 1988.

[9] P. Rai y A. Lal, *Google PageRank Algorithm: Markov Chain Model and Hidden Markov Model*, International Journal of Computer Applications, Vol. 138, No. 9, 2016.

[10] P. Ravi Kumar, A. Kumar Singh y K. Leng Goh, *Application of Markov Chain in the PageRank Algorithm*, Pertanika Journal of Science & Technology, Vol. 21, Issue 2, p. 541-553, 2013.

[11] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer-Verlag New York Heidelberg Berlin, 1981.