Modelos matriciales: cadenas de Markov

Problemas para la ciencia de datos

PID_00262437

Francesc Pozo Montero
Jordi Ripoll Missé

Universitat Oberta de Catalunya

Francesc Pozo Montero

Licenciado en Matemáticas por la Universidad de Barcelona (2000) y doctor en Matemática Aplicada por la Universidad Politécnica de Cataluña (2005). Ha sido profesor asociado de la Universidad Autónoma de Barcelona y profesor asociado, colaborador y actualmente profesor agregado en la Universidad Politécnica de Cataluña. Además, es cofundador del Grupo de Innovación Matemática E-learning (GIMEL), responsable de varios proyectos de innovación docente y autor de varias publicaciones. Como miembro del grupo de investigación consolidado CoDAlab, centra su investigación en la teoría de control y las aplicaciones en ingeniería mecánica y civil, así como en el uso de la ciencia de datos para la monitorización de la integridad estructural y para la monitorización de la condición, sobre todo en turbinas eólicas.

Jordi Ripoll Missé

Licenciado en Matemáticas y doctor en Ciencias Matemáticas por la Universidad de Barcelona (2005). Profesor colaborador de la Universitat Oberta de Catalunya desde 2011 y profesor del Departamento de Informática, Matemática Aplicada y Estadística de la Universidad de Girona (UdG) desde 1996, donde actualmente es profesor agregado y desarrolla tareas de investigación en el ámbito de la biología matemática (modelos con ecuaciones en derivadas parciales y dinámica evolutiva). También ha sido profesor y tutor de la UNED en dos etapas, primero en el centro asociado de Terrassa y actualmente en el de Girona. Ha participado en numerosos proyectos de innovación docente, especialmente en cuanto al aprendizaje de las matemáticas en línea.

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por la profesora: Cristina Cano Bastidas (2019)

Primera edición: febrero 2019

Av. Tibidabo, 39-43, 08035 Barcelona
Diseño: Manel Andreu
Realización editorial: Oberta UOC Publishing, SL

Índice

1. Problemas

Modelos matriciales: cadenas de Markov

Temas: *Introducción a los modelos matriciales en tiempo discreto. Concepto de cadenas de Markov en tiempo discreto. Diagrama de estados y probabilidades de transición. Evolución en el tiempo de una cadena de Markov. Matrices positivas y valor propio dominante. Distribuciones de estado estacionarias. Aplicaciones.*

Una cadena de Markov es un tipo de proceso estocástico en el que el sistema no tiene memoria, es decir, la evolución futura solo depende del presente y no del estado del sistema en el pasado. Concretamente, en las cadenas de Markov la distribución de las probabilidades condicionadas de los estados futuros (probabilidades de transición) dependen del estado presente del proceso y no de la secuencia de acontecimientos que han sucedido en el pasado.

En este módulo nos centraremos en las llamadas *cadenas de Markov homogéneas*, en las que las probabilidades de transición son independientes del tiempo (lo que se conoce en otros contextos como *sistemas dinámicos autónomos*). Si trabajamos a tiempo discreto $t \geq 0$, tendremos un modelo matricial lineal, donde la matriz P del sistema se puede describir por columnas o por filas. Habitualmente se construye la matriz (positiva) P mediante un diagrama de estados (*nodos*) y de las probabilidades de transición entre estados (*flechas entre nodos*).

La evolución en el tiempo de la cadena de Markov se obtiene por iteración de la fórmula $\vec{x}_{t+1} = P\,\vec{x}_t$, $t \geq 0$, es decir, multiplicando la matriz del sistema por el vector (columna) de los estados:

$$\begin{pmatrix} x^{(1)}_{t+1} \\ \vdots \\ x^{(n)}_{t+1} \end{pmatrix} = \begin{pmatrix} p_{11} & \dots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \dots & p_{nn} \end{pmatrix} \begin{pmatrix} x^{(1)}_{t} \\ \vdots \\ x^{(n)}_{t} \end{pmatrix}, \quad t = 0,1,2,\dots$$

Bajo ciertas hipótesis sobre la matriz P, la distribución límite del sistema coincide con el estado estacionario dado por el vector propio de valor propio 1 de la matriz P. Pero es necesario que este vector propio esté normalizado, de forma que la suma de sus componentes sea igual a 1: $\vec{x} = P\,\vec{x}$ con $x^{(1)} + \dots + x^{(n)} = 1$. Si el vector propio no está normalizado, solo hay que dividirlo por la suma de sus componentes: $\frac{1}{v_1 + \ldots + v_n}(v_1, \ldots, v_n)$.

A partir de esta teoría podemos resolver los siguientes problemas y ejercicios de autoevaluación:

1. Se ha hecho un estudio sobre el movimiento de los trabajadores de una empresa multinacional, que tiene la sede principal en Londres (LON) y tres filiales repartidas por las ciudades de París (PAR), Milán (MIL) y Barcelona (BCN). Se ha estimado que, cada año, el 70 % de los trabajadores localizados en cada una de las ciudades no cambia de puesto de trabajo de un año a otro. Según los datos recopilados, el 30 % restante de los trabajadores se distribuye entre cada una de las localizaciones de la manera siguiente:

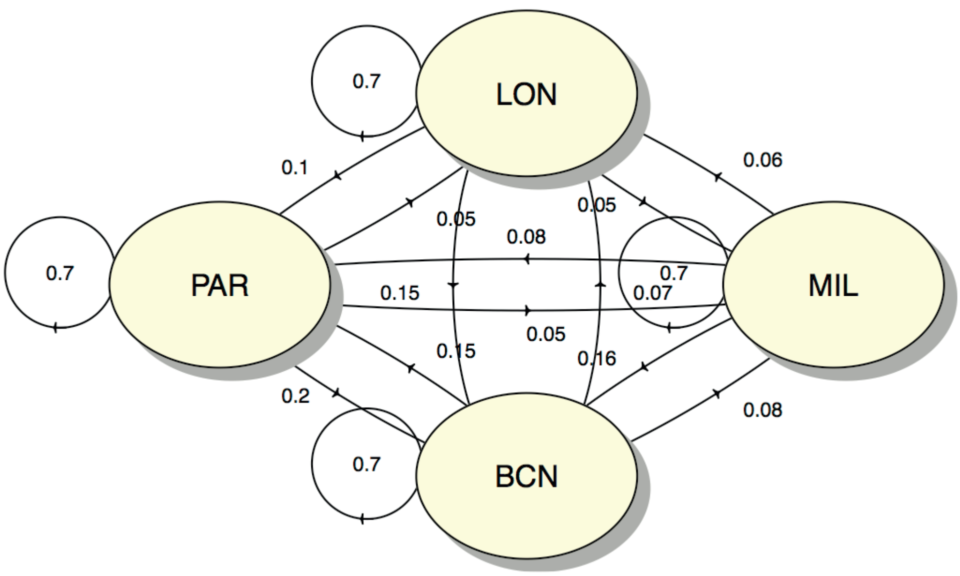
- Anualmente, el 10 % de los trabajadores de LON se va a PAR; el 5 %, a MIL, y el resto, a BCN.

- Cada año, el 20 % de los trabajadores de PAR se va a BCN y el resto se divide en partes iguales entre LON y MIL.

- Cada año, el 6 % de los trabajadores de MIL se va a LON; el 8 %, a PAR, y el resto, a BCN.

- Finalmente, cada año, el 7 % de los trabajadores de BCN se va a LON; el 8 %, a MIL, y el resto, a PAR.

Diagrama de estados y probabilidades de transición: movimiento de trabajadores



a) Escribid la matriz de esta cadena de Markov a partir del diagrama de estados (4 nodos: LON, PAR, MIL y BCN) y de las probabilidades de transición. Recordad que la suma de las columnas de esta matriz tiene que ser igual a 1.

b) Disponemos de los datos para el año 2019. La plantilla total de la empresa es de N = 236 trabajadores repartidos así: 46 en LON, 58 en PAR, 70 en MIL y 62 en BCN. Calculad cuántos trabajadores habrá el año siguiente en cada ciudad.

c) Considerando la distribución geográfica de trabajadores del apartado anterior para el año 2019, ¿cuántos trabajadores había el 2018 en Barcelona?

d) Según este modelo, ¿cuántos trabajadores habrá en 2025 trabajando en cada ciudad?

e) ¿Qué porcentaje de trabajadores se espera que haya en cada una de las sedes a largo plazo?

2. Las encuestas de satisfacción de los estudiantes de una universidad sobre la docencia recibida clasifican a los estudiantes en tres categorías:

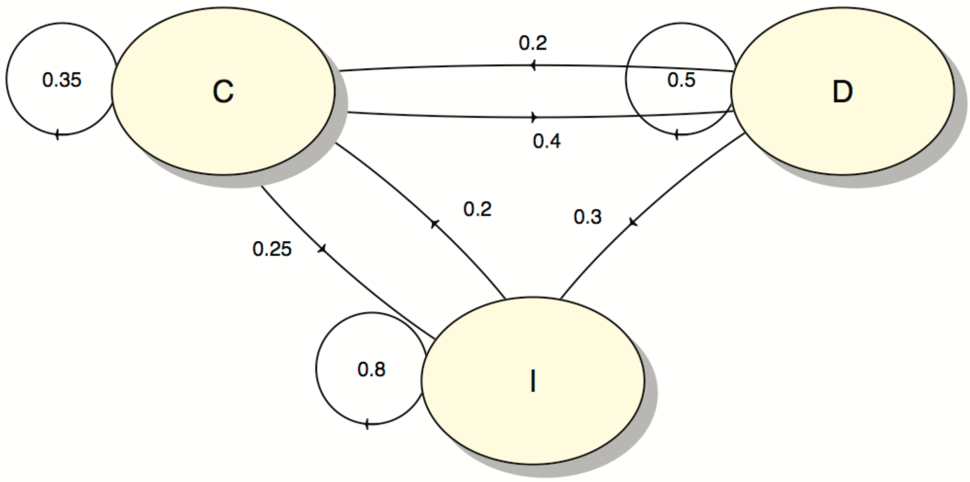
contentos $\overset{\circ\circ}{\smile}$, descontentos $\overset{\circ\circ}{\frown}$ e indiferentes $\overset{\circ\circ}{-}$.

A partir de los datos recopilados durante los últimos años, se han podido estimar las probabilidades de cambiar de opinión que tienen los estudiantes sobre la docencia que han recibido. Los datos, en forma de matriz, son los siguientes:

$$P = \begin{array}{c} \begin{array}{ccc} C & D & I \end{array} \\ \begin{pmatrix} 35\,\% & 20\,\% & 20\,\% \\ 40\,\% & 50\,\% & 0 \\ 25\,\% & 30\,\% & 80\,\% \end{pmatrix} \end{array} \begin{array}{c} \\ C \\ D \\ I \end{array}$$

donde cada elemento de la matriz, p_{ij}, $1 \leq i,j \leq 3$, es la probabilidad de transición del estado j al estado i en una unidad de tiempo. El diagrama de los tres estados con las probabilidades de transición correspondientes es:

Diagrama de estados y probabilidades de transición: satisfacción estudiantes



a) Comprobad que la matriz transpuesta P^T tiene el vector propio $\vec{v} = (1,1,1)^T$ de valor propio $\lambda_1 = 1$. Recordad que los valores propios de una matriz y los de su transpuesta son los mismos, pero los vectores propios pueden ser diferentes.

b) Calculad los otros dos valores propios de P sabiendo que el determinante y la traza de la matriz cumplen que $\det(P) = 1 \cdot \lambda_2 \cdot \lambda_3$ y $\mathrm{tr}(P) = 1+\lambda_2+\lambda_3$.

c) ¿Cuál será la proporción entre *contentos : descontentos : indiferentes* a largo plazo? Expresad los porcentajes (%) de cada categoría.

2. Soluciones de los problemas

1. a) La matriz de las probabilidades de cambiar de sede, es decir, de los movimientos de los trabajadores de la empresa multinacional es:

$$P = \begin{array}{c} \begin{array}{cccc} LON & PAR & MIL & BCN \end{array} \\ \begin{pmatrix} 0.70 & 0.05 & 0.06 & 0.07 \\ 0.10 & 0.70 & 0.08 & 0.15 \\ 0.05 & 0.05 & 0.70 & 0.08 \\ 0.15 & 0.20 & 0.16 & 0.70 \end{pmatrix} \end{array} \begin{array}{c} \\ LON \\ PAR \\ MIL \\ BCN \end{array}$$

b) Si multiplicamos la matriz P por los datos del problema, es decir, por el vector del número de trabajadores en cada sede en 2019, obtendremos la predicción, según este modelo, del número de trabajadores que habrá en cada sede en 2020:

$$\begin{array}{c} \begin{array}{cccc} LON & PAR & MIL & BCN \end{array} \\ \begin{pmatrix} 0.70 & 0.05 & 0.06 & 0.07 \\ 0.10 & 0.70 & 0.08 & 0.15 \\ 0.05 & 0.05 & 0.70 & 0.08 \\ 0.15 & 0.20 & 0.16 & 0.70 \end{pmatrix} \end{array} \begin{array}{c} 2019 \\ \begin{pmatrix} 46 \\ 58 \\ 70 \\ 62 \end{pmatrix} \end{array} = \begin{array}{c} 2020 \\ \begin{pmatrix} 43.64 \\ 60.10 \\ 59.16 \\ 73.10 \end{pmatrix} \end{array} \begin{array}{c} \\ LON \\ PAR \\ MIL \\ BCN \end{array}$$

Nota

El resultado final se puede redondear al entero más cercano. Dado que son predicciones medias, se obtienen decimales.

c) Según el modelo, para saber el número de trabajadores que había en cada sede en 2018 tenemos que resolver el sistema lineal siguiente:

$$\begin{array}{c} \begin{array}{cccc} LON & PAR & MIL & BCN \end{array} \\ \begin{pmatrix} 0.70 & 0.05 & 0.06 & 0.07 \\ 0.10 & 0.70 & 0.08 & 0.15 \\ 0.05 & 0.05 & 0.70 & 0.08 \\ 0.15 & 0.20 & 0.16 & 0.70 \end{pmatrix} \end{array} \begin{array}{c} 2018 \\ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \end{array} = \begin{array}{c} 2019 \\ \begin{pmatrix} 46 \\ 58 \\ 70 \\ 62 \end{pmatrix} \end{array} \begin{array}{c} \\ LON \\ PAR \\ MIL \\ BCN \end{array}$$

y obtenemos $x_1 = 49.98716$, $x_2 = 56.78659$, $x_3 = 87.61808$ y $x_4 = 41.60816$. Por lo tanto, en Barcelona había $41.60816 \simeq 42$ trabajadores de la empresa.

d) Para saber los trabajadores que habrá en 2025 a partir de los datos del 2019, tenemos que iterar la cadena de Markov seis veces:

$$P \cdot x_{2019} = x_{2020}\ , \quad P \cdot x_{2020} = x_{2021}\ , \quad P \cdot x_{2021} = x_{2022}$$

$$P \cdot x_{2022} = x_{2023}\ , \quad P \cdot x_{2023} = x_{2024}\ , \quad P \cdot x_{2024} = x_{2025},$$

es decir, para pasar del año 2019 al 2025 tenemos que multiplicar por la sexta potencia P^6 de la matriz:

$$\begin{array}{cccc} \mathit{LON} & \mathit{PAR} & \mathit{MIL} & \mathit{BCN} \end{array} \qquad 2019 \qquad 2025$$

$$\begin{pmatrix} 0.70 & 0.05 & 0.06 & 0.07 \\ 0.10 & 0.70 & 0.08 & 0.15 \\ 0.05 & 0.05 & 0.70 & 0.08 \\ 0.15 & 0.20 & 0.16 & 0.70 \end{pmatrix}^6 \begin{pmatrix} 46 \\ 58 \\ 70 \\ 62 \end{pmatrix} = \begin{pmatrix} 40.26034 \\ 66.55989 \\ 42.97274 \\ 86.20703 \end{pmatrix} \begin{array}{c} \mathit{LON} \\ \mathit{PAR} \\ \mathit{MIL} \\ \mathit{BCN} \end{array}$$

e) Para calcular la distribución límite de trabajadores en cada sede, tenemos que calcular el vector propio $\vec{x}$ de valor propio $\lambda = 1$ de la matriz P:

$$\begin{array}{cccc} \mathit{LON} & \mathit{PAR} & \mathit{MIL} & \mathit{BCN} \end{array}$$

$$\begin{pmatrix} 0.70 & 0.05 & 0.06 & 0.07 \\ 0.10 & 0.70 & 0.08 & 0.15 \\ 0.05 & 0.05 & 0.70 & 0.08 \\ 0.15 & 0.20 & 0.16 & 0.70 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \begin{array}{c} \mathit{LON} \\ \mathit{PAR} \\ \mathit{MIL} \\ \mathit{BCN} \end{array}$$

Este sistema tiene infinitas soluciones. Si escogemos una que no sea nula y expresamos las componentes en porcentajes, obtendremos:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 16.9\% \\ 28.7\% \\ 17.5\% \\ 36.9\% \end{pmatrix} \begin{array}{c} \mathit{LON} \\ \mathit{PAR} \\ \mathit{MIL} \\ \mathit{BCN} \end{array}$$

Recordad que cualquier múltiple no nulo de un vector propio también es vector propio y, por este motivo, podemos normalizarlo (dividir el vector por la suma de sus componentes) y expresarlo en porcentajes.
El código R necesario para resolver todo el ejercicio es el siguiente:

```
> # 1. Movimiento de los trabajadores de una empresa multinacional
> Etiquetas <- c(" LON", "PAR", "MIL", "BCN")
> MatrizP <- matrix(c(0.7,0.05,0.06,0.07,0.1,0.7,0.08,0.15,
```

```
+                     0.05,0.05,0.7,0.08,0.15,0.2,0.16,0.7), 4,4,
+                     byrow=T, dimnames = list(Etiquetas, Etiquetas))
> # install.packages("markovchain")
> library(markovchain)
> # Creamos un objeto nuevo que es una cadena de Markov (por columnas)
> CM_trabajadores <- new("markovchain", states = Etiquetas, byrow = F,
+                    transitionMatrix = MatrizP,
+                    name = "Movimiento de los trabajadores entre las sedes de la empresa")
> CM_trabajadores

Movimiento de los trabajadores entre las sedes de la empresa
 A  4 - dimensional discrete Markov Chain defined by the following states:
 LON, PAR, MIL, BCN
 The transition matrix  (by cols)  is defined as follows:
     LON  PAR  MIL  BCN
LON 0.70 0.05 0.06 0.07
PAR 0.10 0.70 0.08 0.15
MIL 0.05 0.05 0.70 0.08
BCN 0.15 0.20 0.16 0.70

> # install.packages("diagram")
> library(diagram)
> # (a) Dibujo del diagrama de estados y probabilidades de transición
> plotmat(MatrizP, pos = c(1,2,1), lwd = 1, box.lwd = 1,cex.txt = 0.7,
+   box.size = 0.09, box.type = "circle", box.prop = 0.75,
+   box.col = " light yellow", arr.length=.05, arr.width=.1,
+   self.cex = .5, self.shifty = .01, self.shiftx = -.13,
+   main = "Diagrama de estados y probabilidades de transición: movimiento de trabajadores")
> # (b) Datos de 2019:
> library(expm);
> x2019 <- matrix(c(46,58,70,62), 4, 1, dimnames = list(Etiquetas))
> x2019

    [,1]
LON   46
PAR   58
MIL   70
BCN   62

> # Distribución de trabajadores en 2019
> x2020 <- MatrizP %*% x2019 # multiplicamos por la matriz para avanzar un año
> x2020

     [,1]
LON 43.64
PAR 60.10
MIL 59.16
BCN 73.10
```

```
> # Distribución de trabajadores en 2020
> # (c) Datos de 2018
> x2018 <- solve(MatrizP,x2019)  # solucionamos el sistema para retroceder un año
> x2018

        [,1]
LON 49.98716
PAR 56.78659
MIL 87.61808
BCN 41.60816

> # Distribución de trabajadores en 2018
> # (d) Datos del 2025
> CM_trabajadores^6  # cadena de Markov seis años más tarde (2025-2019= 6)

Movimiento de los trabajadores entre las sedes de la empresa^6
 A  4 - dimensional discrete Markov Chain defined by the following states:
 LON, PAR, MIL, BCN
 The transition matrix  (by cols)  is defined as follows:
          LON       PAR       MIL       BCN
LON 0.2258980 0.1525205 0.1578021 0.1609144
PAR 0.2678989 0.3188126 0.2571385 0.2862213
MIL 0.1549842 0.1579423 0.2309254 0.1696457
BCN 0.3512189 0.3707246 0.3541340 0.3832185

> x2025 <- (MatrizP %^% 6) %*% x2019
> x2025

        [,1]
LON 40.26034
PAR 66.55989
MIL 42.97274
BCN 86.20703

> # (e) Estado estacionario = distribución límite de estados
> V <- eigen(MatrizP)
> V # valores y vectores propios de la matriz del sistema

eigen() decomposition
$values
[1] 1.0000000+0.0000000i 0.6422344+0.0071586i 0.6422344-0.0071586i
[4] 0.5155311+0.0000000i

$vectors
             [,1]                  [,2]                  [,3]          [,4]
[1,] -0.3204833+0i  0.0198523+0.4321114i  0.0198523-0.4321114i  0.1085121+0i
[2,] -0.5452498+0i  0.4791925-0.2714381i  0.4791925+0.2714381i  0.5294608+0i
[3,] -0.3310336+0i -0.6733400+0.0000000i -0.6733400+0.0000000i  0.1832046+0i
[4,] -0.7002927+0i  0.1742952-0.1606734i  0.1742952+0.1606734i -0.8211775+0i
```

```
> vaps <- V$values
> abs(vaps) # valores propios en valor absoluto

[1] 1.0000000 0.6422743 0.6422743 0.5155311

> veps <- V$vectors
> steadystate <- abs(veps[,1])
> (steadystate <- 100*steadystate/sum(steadystate))

[1] 16.89369 28.74184 17.44983 36.91464
```

2. a) Comprobamos que $P^T \cdot \vec{v} = \vec{v}$ con $\vec{v} = (1,1,1)^T$:

$$\begin{array}{c} \begin{array}{ccc} C & D & I \end{array} \\ \begin{pmatrix} 0.35 & 0.40 & 0.25 \\ 0.20 & 0.50 & 0.30 \\ 0.20 & 0 & 0.80 \end{pmatrix} \end{array} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} & \begin{array}{c} C \\ D \\ I \end{array} \end{array}$$

Por lo tanto, la matriz transpuesta P^T tiene el valor propio $\lambda_1 = 1$ y, en consecuencia, la matriz del sistema P también tiene el valor propio $\lambda_1 = 1$.

b) Calculamos primero la traza y el determinante de la matriz: $\mathrm{tr}(P) = 1.65$ y $\det(P) = 0.075$. Para encontrar los otros dos valores propios λ_2, λ_3 podemos resolver las ecuaciones de la siguiente manera. De la segunda ecuación tenemos que $\lambda_2 = 0.65 - \lambda_3$ y entonces la primera ecuación pasa a ser $0.075 = (0.65 - \lambda_3)\lambda_3$, que es una ecuación de segundo grado con dos soluciones diferentes: $\lambda = 0.5$ o $\lambda = 0.15$.

c) Para determinar la distribución límite de estudiantes en cada categoría, tenemos que calcular el vector propio de valor propio $\lambda = 1$, $P \cdot \vec{x} = \vec{x}$:

$$\begin{array}{c} \begin{array}{ccc} C & D & I \end{array} \\ \begin{pmatrix} 0.35 & 0.20 & 0.20 \\ 0.40 & 0.50 & 0 \\ 0.25 & 0.30 & 0.80 \end{pmatrix} \end{array} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} & \begin{array}{c} C \\ D \\ I \end{array} \end{array}$$

De las infinitas soluciones de este sistema, elegimos una que no sea el vector cero y lo expresamos en porcentajes:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{array}{cc} \begin{pmatrix} 23.5\,\% \\ 18.8\,\% \\ 57.7\,\% \end{pmatrix} & \begin{array}{c} C \\ D \\ I \end{array} \end{array}$$

El código R necesario para resolver todo el ejercicio es el siguiente:

```
> # 2. Encuestas de satisfacción de los estudiantes
> Etiquetas <- c(" C", "D", " I")
> MatrizP <- matrix(c(0.35,0.20,0.20,0.40,0.50,0,0.25,0.30,0.80),
+                    3,3, byrow=T, dimnames = list(Etiquetas, Etiquetas))
> # Creamos un objeto nuevo que es una cadena de Markov (por columnas)
> CM_estudiantes <- new("markovchain", states = Etiquetas, byrow = F,
+                       transitionMatrix = MatrizP,
+                       name = "Encuestas de satisfacción de los estudiantes")
> CM_estudiantes

Encuestas de satisfacción de los estudiantes
 A  3 - dimensional discrete Markov Chain defined by the following states:
 C, D, I
 The transition matrix  (by cols)  is defined as follows:
     C   D   I
C 0.35 0.2 0.2
D 0.40 0.5 0.0
I 0.25 0.3 0.8

> # Dibujo del diagrama de estados y probabilidades de transición
> plotmat(MatrizP, pos = c(2,1), lwd = 1, box.lwd = 1,cex.txt = 0.7,
+         box.size = 0.09, box.type = "circle", box.prop = 0.75,
+         box.col = " light yellow", arr.length=.05, arr.width=.1,
+         self.cex = .5, self.shifty = .01, self.shiftx = -.13,
+ main = "Diagrama de estados y probabilidades de transición: satisfacción estudiantes")
> # (a) La matriz P y su transpuesta tienen valor propio = 1
> vep <- matrix(c(1,1,1),3,1)
> t(MatrizP) %*% vep # la matriz transpuesta tiene valor propio (1,1,1)

  [,1]
C    1
D    1
I    1

> # (b) Resto de valores propios:
> V <- eigen(MatrizP)
> V # valores y vectores propios de la matriz del sistema

eigen() decomposition
$values
[1] 1.00 0.50 0.15

$vectors
          [,1]          [,2]        [,3]
[1,] 0.3617281 -2.386516e-16  0.65561007
[2,] 0.2893825 -7.071068e-01 -0.74926865
[3,] 0.8862339  7.071068e-01  0.09365858
```

```
> VT <- eigen(t(MatrizP))
> VT # valores y vectores propios de la matriz transpuesta

eigen() decomposition
$values
[1] 1.00 0.50 0.15

$vectors
          [,1]       [,2]       [,3]
[1,] 0.5773503 -0.6948499 -0.9169493
[2,] 0.5773503 -0.5500895  0.2821382
[3,] 0.5773503  0.4632333  0.2821382

> # (c) Proporción límite en porcentajes
> vaps <- V$values
> abs(vaps) # valores propios en valor absoluto

[1] 1.00 0.50 0.15

> veps <- V$vectors
> steadystate <- abs(veps[,1])
> (steadystate <- 100*steadystate/sum(steadystate))

[1] 23.52941 18.82353 57.64706
```

3. Ejercicios de autoevaluación

1. La dinámica de una cadena de Markov para una población dividida en dos grupos (individuos sanos e individuos enfermos) se refleja en la matriz

$$P = \begin{pmatrix} 0.25 & r \\ 0.75 & 1-r \end{pmatrix},$$

donde r es la probabilidad de cambiar del segundo grupo al primero en cada periodo de tiempo. ¿Para qué valor de r la proporción entre los individuos sanos y los individuos enfermos será, a la larga, de 2 : 4?

○ $r = 0.625$

○ $r = 0.375$

○ $r = 0.125$

○ Ninguna de las respuestas anteriores es correcta.

2. Una población en la que los individuos están clasificados en contentos ☺ o enfadados ☹ evoluciona según una cadena de Markov. En promedio, cada año el 30 % de los contentos se indigna y, en cambio, el 20 % de los enfadados recupera el buen humor. A largo plazo, ¿cuál de las siguientes afirmaciones es cierta?

○ La proporción de contentos y enfadados será de 30 % : 70 %.

○ Habrá el mismo número de contentos que de enfadados.

○ La proporción de contentos y enfadados será de 40 % : 60 %.

○ La proporción entre contentos y enfadados dependerá de los datos iniciales del problema.

3. Mensualmente se recogen datos de la opinión que tienen los habitantes de un país sobre dos candidatas a la presidencia del gobierno: la señora Mariana y la señora Petra. Consideramos que algunos son partidarios de Mariana y otros de Petra y que el porcentaje de los que continúan creyendo en su candidata de un mes al siguiente es del 90 % y del 75 %, respectivamente. A la larga, ¿cuál será el porcentaje de individuos a favor de cada candidata?

○ Mariana : Petra = 88.2 % : 11.8 %

○ Mariana : Petra = 71.4 % : 28.6 %

○ Mariana : Petra = 45.4 % : 54.6 %

○ Mariana : Petra = 21.7 % : 78.3 %

4. Los habitantes de una región metropolitana se clasifican en dos grupos: residentes en la ciudad (centro) y residentes en los suburbios (periferia). Cada año, el 5 % de la población de la ciudad se va hacia los suburbios y el 3 % de la población de los suburbios se muda a la ciudad. Escribid la matriz de Markov del modelo y calculad la distribución de población que habrá a la larga.

◯ Ciudad : Suburbio = 62.5 % : 37.5 %

◯ Ciudad : Suburbio = 37.5 % : 62.5 %

◯ Ciudad : Suburbio = 5 % : 95 %

◯ Ciudad : Suburbio = 97 % : 3 %

5. Considerad una cadena de Markov para la distribución geográfica de una especie de ave migratoria distribuida en tres hábitats diferentes. Se han recopilado datos sobre las probabilidades (medias) de transición: el 23 % de las aves del primer hábitat se queda y el 55 % emigra al segundo; el 29 % de las aves del segundo hábitat se queda y el 65 % emigra al primero. Finalmente, el 48 % y el 50 % de las aves del tercer hábitat emigra hacia el primero y el segundo, respectivamente. Según este modelo, ¿cuál de los tres hábitats será el más poblado a largo plazo?

◯ El primero.

◯ El segundo.

◯ El tercero.

◯ No se puede saber qué hábitat será el más poblado.

6. Una cadena de Markov para los consumidores de un producto comercializado por dos empresas (Nestlé y Unilever) se da por esta matriz:

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

donde $0 \leq p,q \leq 1$ son las probabilidades de continuar comprando semanalmente el mismo producto de cada una de las empresas, respectivamente. A partir de la fórmula para las potencias de la matriz $P^t = V \cdot D^t \cdot V^{-1}$, $t \geq 0$, donde la matriz V son los vectores propios por columnas y la matriz diagonal D son los valores propios de P, calculad la expresión de la matriz de transición al cabo de 52 semanas:

$$P^{52} = \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}^{52}$$

en función de las probabilidades de transición p y q que hagan que se cumpla que $p + q \neq 2$.

○ $P^{52} = \frac{1}{2-p-q} \begin{pmatrix} 1-q & 1 \\ 1-p & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (p+q-1)^{52} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1-p & q-1 \end{pmatrix}$

○ $P^{52} = \begin{pmatrix} p^{52} & 0 \\ 0 & q^{52} \end{pmatrix}$

○ $P^{52} = \frac{1}{p+q-2} \begin{pmatrix} 1-p & 1 \\ 1-q & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (p+q-1)^{52} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1-p & 1-q \end{pmatrix}$

○ $P^{52} = \begin{pmatrix} (1-p)^{52} & 0 \\ 0 & (1-q)^{52} \end{pmatrix}$

7. Considerad la matriz P del ejercicio anterior para la cadena de Markov de los consumidores de un producto comercializado por dos empresas (Nestlé y Unilever). Encontrad para qué caso se cumple el límite de la cadena de Markov:

$$\lim_{t\to\infty} \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}^t = \frac{1}{2-(p+q)} \begin{pmatrix} 1-q & 1-q \\ 1-p & 1-p \end{pmatrix}$$

en función de las probabilidades de transición p y q.

○ Si $p+q \neq 2$ y $p+q \neq 1$.

○ Si $p+q \neq 1$ y $p+q \neq 0$.

○ Si $p+q \neq 2$ y $p+q \neq 0$.

○ Ninguna de las respuestas anteriores es correcta.

Nota del ejercicio

Observad que la matriz límite tiene las dos columnas iguales y que los elementos de cada columna suman 1.

4. Soluciones de los ejercicios de autoevaluación

1. La probabilidad r de recuperarse de la enfermedad se calcula resolviendo la siguiente ecuación:

$$\begin{array}{r} \textbf{sanos} \\ \textbf{enfermos} \end{array} \begin{pmatrix} 0.25 & r \\ 0.75 & 1-r \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

y el valor de r es:

- ○ $r = 0.625$
- ⊗ $r = 0.375$
- ○ $r = 0.125$
- ○ Ninguna de las respuestas anteriores es correcta.

2. Para encontrar el porcentaje de contentos y enfadados que habrá según el modelo, tenemos que resolver este sistema:

$$\begin{array}{r} \textbf{contentos} \\ \textbf{enfadados} \end{array} \begin{pmatrix} 70\,\% & 20\,\% \\ 30\,\% & 80\,\% \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Una de las soluciones del sistema es $(x_1,x_2) = (20,30) \equiv (20,30)/50$ y, por lo tanto, la afirmación correcta es:

- ○ La proporción de contentos y enfadados será de 30 % : 70 %.
- ○ Habrá el mismo número de contentos que de enfadados.
- ⊗ La proporción de contentos y enfadados será de 40 % : 60 %.
- ○ La proporción entre contentos y enfadados dependerá de los datos iniciales del problema.

3. Para saber el porcentaje de individuos a favor de cada candidata que habrá según el modelo, tenemos que resolver este sistema:

$$\begin{array}{r} \textbf{Mariana} \\ \textbf{Petra} \end{array} \begin{pmatrix} 90\,\% & 25\,\% \\ 10\,\% & 75\,\% \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Una de las soluciones del sistema es $(x_1,x_2) = (25,10) \equiv (25,10)/35$ y, por lo tanto, la proporción será:

- ○ Mariana : Petra = 88.2 % : 11.8 %
- ⊗ Mariana : Petra = 71.4 % : 28.6 %
- ○ Mariana : Petra = 45.4 % : 54.6 %
- ○ Mariana : Petra = 21.7 % : 78.3 %

4. A partir de los datos del modelo, la distribución geográfica de población que habrá entre el centro de la ciudad y los suburbios se obtiene resolviendo:

$$\begin{array}{c} \textbf{ciudad} \\ \textbf{suburbio} \end{array} \begin{pmatrix} 95\,\% & 3\,\% \\ 5\,\% & 97\,\% \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Una de las soluciones del sistema es $(x_1,x_2) = (3,5) \equiv (3,5)/8$ y, por lo tanto, la proporción será:

○ Ciudad : Suburbio = 62.5 % : 37.5 %
⊗ Ciudad : Suburbio = 37.5 % : 62.5 %
○ Ciudad : Suburbio = 5 % : 95 %
○ Ciudad : Suburbio = 97 % : 3 %

5. Con los datos recopilados podemos saber la distribución geográfica de esta especie en cada hábitat. La distribución límite se obtiene resolviendo el sistema:

$$\begin{array}{c} \textbf{primer hábitat} \\ \textbf{segundo hábitat} \\ \textbf{tercer hábitat} \end{array} \begin{pmatrix} 23\,\% & 65\,\% & 48\,\% \\ 55\,\% & 29\,\% & 50\,\% \\ 22\,\% & 6\,\% & 2\,\% \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

o agrupando las variables, en las que obtenemos un sistema lineal homogéneo:

$$\begin{array}{c} \textbf{primer hábitat} \\ \textbf{segundo hábitat} \\ \textbf{tercer hábitat} \end{array} \begin{pmatrix} -77\,\% & 65\,\% & 48\,\% \\ 55\,\% & -71\,\% & 50\,\% \\ 22\,\% & 6\,\% & -98\,\% \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Podéis resolver el sistema lineal por el método de Gauss, por ejemplo. Si normalizamos una de las soluciones obtenidas, obtendremos que la proporción en cada hábitat será: 44.28 % (primero), 43.14 % (segundo) y 12.58 % (tercero). La respuesta correcta es:

⊗ El primero.
○ El segundo.
○ El tercero.
○ No se puede saber qué hábitat será el más poblado.

6. Para calcular la potencia $t = 52$ (semanas) de una cadena de Markov con dos estados:

$$\begin{array}{c} \textbf{Nestlé} \\ \textbf{Unilever} \end{array} \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$$

tenemos que calcular los valores propios y los vectores propios de la matriz. Los valores propios son $\lambda_1 = 1$ y $\lambda_2 = p+q-1$ y los vectores propios son $\vec{v}_1 = (1-q,1-p)$ y $\vec{v}_2 = (1,-1)$. Entonces, usando la fórmula por las potencias:

$$\begin{array}{c} \textbf{Nestlé} \\ \textbf{Unilever} \end{array} \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}^t = \begin{pmatrix} 1-q & 1 \\ 1-p & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p+q-1 \end{pmatrix}^t \begin{pmatrix} 1-q & 1 \\ 1-p & -1 \end{pmatrix}^{-1}$$

y calculando la inversa, obtendremos la solución:

⊗ $P^{52} = \frac{1}{2-p-q} \begin{pmatrix} 1-q & 1 \\ 1-p & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (p+q-1)^{52} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1-p & q-1 \end{pmatrix}$

○ $P^{52} = \begin{pmatrix} p^{52} & 0 \\ 0 & q^{52} \end{pmatrix}$

○ $P^{52} = \frac{1}{p+q-2} \begin{pmatrix} 1-p & 1 \\ 1-q & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (p+q-1)^{52} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1-p & 1-q \end{pmatrix}$

○ $P^{52} = \begin{pmatrix} (1-p)^{52} & 0 \\ 0 & (1-q)^{52} \end{pmatrix}$

7. Para obtener el límite propuesto cuando $t \to \infty$ de la cadena de Markov con dos estados:

$$\begin{array}{c} \textbf{Nestlé} \\ \textbf{Unilever} \end{array} \begin{pmatrix} p & 1-q \\ 1-p & q \end{pmatrix}$$

tenemos que asegurarnos de que la matriz del sistema tiene valor propio dominante, es decir, $|\lambda_1| > |\lambda_2|$. Como los valores propios son $\lambda_1 = 1$ y , $\lambda_2 = p + q - 1$, la condición se cumple para:

$$|p+q-1| < 1 , \quad -1 < p+q-1 < 1 , \quad 0 < p+q < 2.$$

Dicho de otro modo, las dos únicas cadenas de Markov que no cumplen el límite son:

$$\begin{array}{c} \textbf{Nestlé} \\ \textbf{Unilever} \end{array} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

correspondientes a dos casos extremos. En el primer caso ($p = q = 1$), los consumidores siempre compran el producto de la misma marca; en el segundo caso ($p = q = 0$), cada semana todos los consumidores cambian de marca del producto que compraron la semana pasada. Por lo tanto, la respuesta correcta es:

○ Si $p + q \neq 2$ y $p + q \neq 1$.
○ Si $p + q \neq 1$ y $p + q \neq 0$.
⊗ Si $p + q \neq 2$ y $p + q \neq 0$.
○ Ninguna de las respuestas anteriores es correcta.