
Caso de uso y guía de resolución en R

Cadenas de Markov: estudio del viento

PID_00262439

Francesc Pozo Montero
Jordi Ripoll Missé

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
© Francesc Pozo Montero, Jordi Ripoll Missé
Todos los derechos reservados
© de esta edición, FUOC, 2019
Av. Tibidabo, 39-43, 08035 Barcelona
Diseño: Manel Andreu
Realización editorial: Oberta UOC Publishing, SL

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Índice

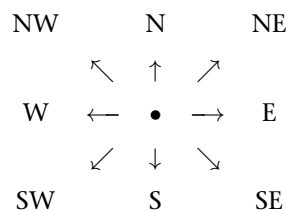
1. Caso de uso: estudio de los cambios de dirección del viento	5
2. Soluciones a las cuestiones	7
3. Guía de resolución en R. Álgebra para ciencia de datos	11

1. Caso de uso: estudio de los cambios de dirección del viento

En este caso de uso ejemplificaremos, a partir de datos reales (medidas empíricas) de una cadena de Markov para las direcciones del viento en una localización, cómo se puede deducir la distribución de probabilidad de la dirección del viento en una localización geográfica.

La dirección en la que sopla el viento en un lugar determinado se puede considerar un proceso estocástico en el cual nos basta con saber cuáles son las probabilidades de que el viento cambie de dirección en cada periodo de tiempo.

Según los datos recopilados en la tesis doctoral de Palafox Duarte y la inferencia estadística realizada en aquel trabajo, podemos considerar una cadena de Markov homogénea para la dirección del viento en la ciudad de Guanajuato (México, latitud 21° 00' 59" N), para un tiempo discreto medido en horas. La dirección del viento, es decir, la orientación del vector del viento, se define como la dirección y el sentido desde el cual sopla el viento. Las direcciones del viento se han clasificado dividiendo la circunferencia en ocho sectores iguales de 45° y asociando a cada sector uno de los puntos cardinales principales y secundarios:



Por lo tanto, este modelo tiene ocho estados posibles etiquetados por los puntos cardinales:

$$\{N, NE, E, SE, S, SW, W, NW\}$$

Las probabilidades de cambiar cada hora de una dirección a otra son las que da la matriz siguiente:

$$P = \begin{pmatrix} 0.541 & 0.3604 & 0.261 & 0.18 & 0.12 & 0.16 & 0.37 & 0.36 \\ 0.311 & 0.44 & 0.281 & 0.11 & 0.070 & 0.057 & 0.13 & 0.206 \\ 0.061 & 0.11 & 0.27 & 0.102 & 0.031 & 0.023 & 0.06 & 0.033 \\ 0.034 & 0.041 & 0.083 & 0.311 & 0.22 & 0.034 & 0.02 & 0.236 \\ 0.021 & 0.0264 & 0.06 & 0.181 & 0.323 & 0.256 & 0.1 & 0.066 \\ 0.011 & 0.007 & 0.03 & 0.081 & 0.16 & 0.27 & 0.12 & 0 \\ 0.008 & 0.0062 & 0.01 & 0.035 & 0.07 & 0.12 & 0.14 & 0.033 \\ 0.013 & 0.009 & 0.005 & 0 & 0.006 & 0.08 & 0.06 & 0.066 \end{pmatrix}$$

donde cada elemento P_{ij} representa la probabilidad de cambiar de la dirección j a la dirección i .

Bibliografía recomendada

Martha Cecilia Palafox Duarte (2009). *Inferencia estadística para cadenas de Markov*.

Cabe destacar que el análisis que hemos hecho aquí sobre los cambios en la dirección del viento en una ciudad específica, es decir, la recopilación de datos (cadena de cambios del viento por horas), la inferencia estadística para estimar la matriz de probabilidades de transición y la predicción posterior de la distribución del viento, puede ser útil para llevar a cabo una planificación urbanística (instalaciones deportivas al aire libre, zonas verdes...), para construir parques eólicos, etc.

Cuando sea necesario, mediante el lenguaje de programación R resolved estas cuestiones:

1. ¿ $P \geq 0$ es una matriz simétrica? ¿Qué elementos de la matriz P son cero?
2. ¿En qué dirección existe una mayor probabilidad de que el viento siga soplando en la misma dirección? ¿Y en cuál existe una mayor probabilidad de que cambie de dirección?
3. Calculad la matriz P^2 y comprobad que $P^2 > 0$. Según este modelo, si en un determinado instante de tiempo observamos que el viento presenta una dirección sudeste (SE), ¿cuál es la probabilidad de que el viento cambie en dirección noroeste (NW) después de dos periodos de tiempos? ¿Y de NW a SW?
4. Hallad los valores propios λ_i , $i = 1 \dots 8$, de la matriz P . Ordenadlos de mayor a menor según su módulo $|\lambda_i|$. Comprobad que se cumpla lo siguiente:

$$\lambda_1 = 1 > |\lambda_2| \geq \dots \geq |\lambda_8|$$

5. Calculad el vector propio \vec{v} del valor propio $\lambda_1 = 1$ de la matriz P . ¿Cuál es la distribución de probabilidad estacionaria de este modelo? Normalizad el vector propio $\frac{1}{v_1 + \dots + v_n}(v_1, \dots, v_n)$.
6. Calculad la matriz P^{40} . ¿Qué observáis? ¿Qué relación guarda con el vector propio hallado en el apartado anterior?
7. Finalmente, calculad la media de cada una de las filas de la matriz P y escribid el resultado en forma de vector. ¿El vector obtenido es muy diferente del vector propio (normalizado) hallado antes? En este caso, podéis usar `rowMeans(P)`.

2. Soluciones a las cuestiones

1. Definimos la matriz P , de dimensión 8×8 , de las probabilidades de cambio de la dirección del viento y observamos que todos los elementos son estrictamente positivos, excepto $P_{84} = 0$ y $P_{68} = 0$. Esto significa que cuando el viento es de sudeste no cambia directamente a noroeste durante la próxima hora y que cuando es de noroeste tampoco cambia directamente a suroeste, respectivamente.

```
> Etiquetas <- c("N", "NE", "E", "SE", "S", "SW", "W", "NW")
> P <- matrix(c(0.541 , 0.3604 , 0.261 , 0.18 , 0.12 , 0.16 , 0.37 , 0.36,
+             0.311 , 0.44 , 0.281 , 0.11 , 0.070 , 0.057 , 0.13 , 0.206,
+             0.061 , 0.11 , 0.27 , 0.102 , 0.031 , 0.023 , 0.06 , 0.033,
+             0.034 , 0.041 , 0.083 , 0.311 , 0.22 , 0.034 , 0.02 , 0.236,
+             0.021 , 0.0264 , 0.06 , 0.181 , 0.323 , 0.256 , 0.1 , 0.066,
+             0.011 , 0.007 , 0.03 , 0.081 , 0.16 , 0.27 , 0.12 , 0,
+             0.008 , 0.0062 , 0.01 , 0.035 , 0.07 , 0.12 , 0.14 , 0.033,
+             0.013 , 0.009 , 0.005 , 0 , 0.006 , 0.08 , 0.06 , 0.066), 8, 8,
+             byrow=T, dimnames = list(Etiquetas, Etiquetas))
> # install.packages("markovchain")
> library(markovchain)
> # Creamos un nuevo objeto, que es una cadena de Markov (por columnas)
> CM <- new("markovchain", states = Etiquetas, byrow = F,
+          transitionMatrix = P, name = "Cambios en la dirección del viento")
> CM
```

Cambios en la dirección del viento

A 8 - dimensional discrete Markov Chain defined by the following states:

N, NE, E, SE, S, SW, W, NW

The transition matrix (by cols) is defined as follows:

	N	NE	E	SE	S	SW	W	NW
N	0.541	0.3604	0.261	0.180	0.120	0.160	0.37	0.360
NE	0.311	0.4400	0.281	0.110	0.070	0.057	0.13	0.206
E	0.061	0.1100	0.270	0.102	0.031	0.023	0.06	0.033
SE	0.034	0.0410	0.083	0.311	0.220	0.034	0.02	0.236
S	0.021	0.0264	0.060	0.181	0.323	0.256	0.10	0.066
SW	0.011	0.0070	0.030	0.081	0.160	0.270	0.12	0.000
W	0.008	0.0062	0.010	0.035	0.070	0.120	0.14	0.033
NW	0.013	0.0090	0.005	0.000	0.006	0.080	0.06	0.066

2. Las probabilidades de que el viento continúe soplando en la misma dirección son los elementos de la diagonal de P . Por lo tanto, la mayor probabilidad de que el viento siga soplando en la misma dirección (norte, N) es

$P_{11} = 54.1$ %. Por otro lado, las probabilidades de que el viento cambie de dirección son los elementos externos de la diagonal de P . Por lo tanto, la mayor probabilidad de que el viento cambie de dirección (de W a N) es $P_{17} = 37$ %.

3. Calculamos P^2 y observamos que, en efecto, es una matriz con todos los elementos estrictamente positivos. Esto implica que, a partir de una dirección, podemos cambiar a cualquier otra durante un intervalo de tiempo de dos horas (cadena o matriz irreducible). Si el viento tiene dirección sudeste, la probabilidad de que el viento cambie a dirección noroeste después de dos horas es $P_{84}^2 = 1.35$ %. En cambio, la probabilidad de cambiar de noroeste a suroeste después de dos horas es $P_{68}^2 = 4$ %.

```
> CM^2 # cadena de Markov después de dos periodos de tiempo (2 h)
```

```
Cambios de dirección del viento^2
```

```
A 8 - dimensional discrete Markov Chain defined by the following states:
```

```
N, NE, E, SE, S, SW, W, NW
```

```
The transition matrix (by cols) is defined as follows:
```

	N	NE	E	SE	S	SW	W	NW
N	0.4387264	0.3994644	0.3453834	0.267256	0.230259	0.2663458	0.370882	0.3639854
NE	0.3317870	0.3460114	0.2980510	0.189089	0.143097	0.1504330	0.235730	0.2603390
E	0.0889620	0.1059148	0.1315120	0.091916	0.063921	0.0496940	0.071350	0.0838060
SE	0.0550040	0.0604686	0.0842180	0.159091	0.157259	0.1070400	0.072150	0.1275770
S	0.0406424	0.0447386	0.0725124	0.151794	0.198733	0.1814868	0.099402	0.0866684
SW	0.0200020	0.0205234	0.0385610	0.086031	0.123840	0.1338630	0.073600	0.0400280
W	0.0123952	0.0119992	0.0188002	0.041317	0.061212	0.0728134	0.048046	0.0241652
NW	0.0124810	0.0108796	0.0109620	0.013506	0.021679	0.0383240	0.028840	0.0134310

4. Debido a que $P \geq 0$ es una matriz positiva, siempre tiene como mínimo un valor propio positivo $\lambda \geq 0$. Además, al tratarse de una matriz estocástica (cadena de Markov), el valor propio dominante de P siempre es $\lambda_1 = 1$. Podemos calcular, también, todos los valores propios de la matriz P , ordenados de mayor a menor según su módulo:

$$\lambda_1 = 1 > |\lambda_2| = 0.58 > |\lambda_3| = 0.26 > |\lambda_4| = 0.17 > |\lambda_5| = 0.13 > |\lambda_6| = 0.11 > |\lambda_7| = 0.05 = |\lambda_8| = 0.05$$

Si se cumple esta última propiedad, afirmaremos que $\lambda_1 = 1$ es el valor propio estrictamente dominante de la matriz.

```
> V <- eigen(P)
```

```
> vaps <- V$values # valores propios de la matriz
```

```
> abs(vaps) # valores propios en módulo
```

```
[1] 1.00000000 0.58015192 0.26267674 0.16790667 0.13261832 0.11360841 0.05325669
[8] 0.05325669
```

```
> abs(vaps)[1] > abs(vaps)[2] # comprobación
```

```
[1] TRUE
```


5. En este ejemplo podemos ver que existe una distribución límite de estados (direcciones del viento) y que coincide con la distribución estacionaria. El estado estacionario es el vector propio $\vec{v} \geq 0$ de valor propio $\lambda_1 = 1$ de la matriz P . Si calculamos este vector y lo normalizamos de manera que la suma de las componentes sea 1, obtendremos esta distribución de probabilidad para la dirección del viento:

N	NE	E	SE	S	SW	W	NW
38 %	29.7 %	9.4 %	7.9 %	7.3 %	4.1 %	2.2 %	1.4 %

Esta distribución indica que, en esta localización, el viento soplará mayormente de componente norte o nordeste.

```
> # Estado estacionario = distribución límite de estados
> veps <- V$vectores # vectores propios de la matriz
> steadystate <- abs(veps[,1])
> (steadystate <- steadystate/sum(steadystate))

[1] 0.38002783 0.29737390 0.09436523 0.07852117 0.07278991 0.04075919 0.02212273
[8] 0.01404003
```

6. Calculamos la potencia 40 de la matriz P y observamos que todas las columnas de esta nueva matriz coinciden (aproximadamente) con el vector propio normalizado \vec{v} . De hecho, las sucesivas potencias, P^2, P^3, \dots , tienden hacia el mismo resultado.

```
> CM^40 # cadena de Markov después de cuarenta periodos de tiempo (40 h)

Cambios de dirección del viento^40
A 8 - dimensional discrete Markov Chain defined by the following states:
N, NE, E, SE, S, SW, W, NW
The transition matrix (by cols) is defined as follows:
      N      NE      E      SE      S      SW      W
N 0.38002783 0.38002783 0.38002783 0.38002783 0.38002783 0.38002783 0.38002783
NE 0.29737390 0.29737390 0.29737390 0.29737390 0.29737390 0.29737390 0.29737390
E 0.09436523 0.09436523 0.09436523 0.09436523 0.09436523 0.09436523 0.09436523
SE 0.07852117 0.07852117 0.07852117 0.07852117 0.07852117 0.07852117 0.07852117
S 0.07278991 0.07278991 0.07278991 0.07278991 0.07278991 0.07278991 0.07278991
SW 0.04075919 0.04075919 0.04075919 0.04075919 0.04075919 0.04075919 0.04075919
W 0.02212273 0.02212273 0.02212273 0.02212273 0.02212273 0.02212273 0.02212273
NW 0.01404003 0.01404003 0.01404003 0.01404003 0.01404003 0.01404003 0.01404003

      NW
N 0.38002783
NE 0.29737390
E 0.09436523
SE 0.07852117
S 0.07278991
SW 0.04075919
W 0.02212273
NW 0.01404003
```

7. Si calculamos la media por filas de la matriz P , obtendremos una primera aproximación a la distribución estacionaria del modelo:

	N	NE	E	SE	S	SW	W	NW
VEP	38 %	29.7 %	9.4 %	7.9 %	7.3 %	4.1 %	2.2 %	1.4 %
media filas	29.4 %	20.1 %	8.6 %	12.2 %	12.9 %	8.5 %	5.3 %	3.0 %

Podemos observar que estos dos vectores no son muy diferentes.

```
> rowMeans(P) # media por filas
```

```
      N      NE      E      SE      S      SW      W      NW
0.294050 0.200625 0.086250 0.122375 0.129175 0.084875 0.052775 0.029875
```

3. Guía de resolución en R. Álgebra para ciencia de datos

1) Definición y operaciones de una matriz A de dimensiones $n \times m$:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}$$

```
> # A <-matrix(c(...), n,m,byrow= T o F)
> # help("matrix"); example("matrix")
> A <-matrix(c(3,2,1,3), 2,2,byrow= T)
> A

      [,1] [,2]
[1,]    3    2
[2,]    1    3

> det(A) # determinante de la matriz

[1] 7

> t(A) # matriz transpuesta

      [,1] [,2]
[1,]    3    1
[2,]    2    3

> solve(A) # matriz inversa A^(-1)

      [,1] [,2]
[1,] 0.4285714 -0.2857143
[2,] -0.1428571 0.4285714

> library(expm) # librería
> A %^% 2 # matriz al cuadrado

      [,1] [,2]
[1,]   11   12
[2,]    6   11

> rankMatrix(A)[1] # rango de la matriz

[1] 2
```

Nota

El símbolo # sirve para escribir comentarios. Los símbolos de verdadero o falso son T o F y también TRUE o FALSE.

2) Valores y vectores propios de una matriz cuadrada A : $A\vec{v} = \lambda\vec{v}$, donde λ es el valor propio (VAP) y $\vec{v} = (v_1, \dots, v_n)$ es el vector propio (VEP):

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

```
> # V <- eigen(A)
> # help("eigen"); example("eigen")
> V <- eigen(A)
> V # valores y vectores propios de la matriz

eigen() decomposition
$values
[1] 4.414214 1.585786

$vectors
      [,1]      [,2]
[1,] 0.8164966 -0.8164966
[2,] 0.5773503  0.5773503

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

[1] 4.414214 1.585786

> veps <- V$vectors
> # A= veps * diag(vaps) * veps^(-1) descomposición
> diag(vaps); solve(veps)
      [,1]      [,2]
[1,] 4.414214 0.000000
[2,] 0.000000 1.585786

      [,1]      [,2]
[1,] 0.6123724 0.8660254
[2,] -0.6123724 0.8660254

> veps %*% diag(vaps) %*% solve(veps)
      [,1] [,2]
[1,]    3    2
[2,]    1    3

> sum(diag(A)) == sum(vaps) # traza de la matriz = suma de los vaps

[1] TRUE
```

3) Ajuste lineal. Recta de regresión $y = ax + b$, donde a y b son coeficientes reales. Método de los mínimos cuadrados (*lm* = *linear model*):

Datos	X	$x_1, x_2, \dots, x_n.$
	Y	$y_1, y_2, \dots, y_n.$

```
> # ajuste <- lm(y ~ x) # y=a*x+b
> # summary(ajuste) # resumen
> # help("lm"); example("lm")
> x <- c(3,2,1); y <- c(1,3,-5);
> ajuste <- lm(y ~ x)
> coef <- coefficients(ajuste); # y=a*x+b, a=coef[2], b=coef[1]
> coef

(Intercept)          x
-6.333333      3.000000

> yy <- fitted(ajuste) # valores de las predicciones
> yy

      1      2      3
2.666667 -0.333333 -3.333333

> plot(x,y,col= "blue") # datos (círculos)
> points(x,yy, pch=2) # predicciones (triángulos)
> abline(ajust,col = "red") # recta
> title("Recta de regresión")
```

4) Resolución de sistemas de ecuaciones lineales $A\vec{x} = \vec{b}$:

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

```
> # help("solve"); example("solve")
> b <- c(3,5)
> cbind(A,b) # matriz ampliada del sistema

      b
[1,] 3 2 3
[2,] 1 3 5

> x <- solve(A,b) # solución del sistema lineal
> x

[1] -0.1428571  1.7142857
```

5) Cadenas de Markov $\vec{x}_{t+1} = P\vec{x}_t$, $t \geq 0$, donde P es la matriz de las probabilidades de transición (por columnas) y \vec{x}_t es el vector (columna) de los n estados del sistema en el tiempo t :

$$\begin{pmatrix} x_{t+1}^{(1)} \\ \vdots \\ x_{t+1}^{(n)} \end{pmatrix} = \begin{pmatrix} P_{11} & \dots & P_{1n} \\ \vdots & & \vdots \\ P_{n1} & \dots & P_{nm} \end{pmatrix} \begin{pmatrix} x_t^{(1)} \\ \vdots \\ x_t^{(n)} \end{pmatrix}$$

```
> # Etiquetas <- c("Estado1", "Estado2", ...) # n estados
> # P <- matrix(c(...), n,n, byrow=T o F, dimnames= list(Etiquetas, Etiquetas))
> # matriz transición
> # CM <- new("markovchain", states= Etiquetas, byrow= T o F, transitionMatrix= P,
> #         name= "Título del problema")
> # plotmat(P, pos= c(n1,n2,n3 ...),
> #         main = "Diagrama de estados y probabilidades de transición: problema")
> # CM^t # cadena de Markov después de t unidades de tiempo
> # Estado estacionario = distribución límite de estados (%), (columnas o filas)
> # 100*steadyStates(t(CM) o CM)
> library(markovchain); library(diagram) # librerías
> Etiquetas <- c("NYC", "BCN") # 2 estados
> # matriz de transición (por columnas)
>
> P <- matrix(c(0.7,0.3,0.4,0.6),2,2,dimnames= list(Etiquetas, Etiquetas))
> CM <- new("markovchain",states= Etiquetas,byrow= F,
>         transitionMatrix= P,name= "Movimiento empleados")
> CM # Cadena de Markov
```

Movimiento empleados

A 2 - dimensional discrete Markov Chain defined by the following states:
NYC, BCN

The transition matrix (by cols) is defined as follows:

NYC BCN

NYC 0.7 0.4

BCN 0.3 0.6

```
> plotmat(P, pos= c(2), main =
>         "Diagrama de estados y probabilidades de transición: movimiento empleados")
> CM^20 # cada columna es la distribución límite de estados
```

Movimiento empleados^20

A 2 - dimensional discrete Markov Chain defined by the following states:
NYC, BCN

The transition matrix (by cols) is defined as follows:

NYC BCN

NYC 0.5714286 0.5714286

BCN 0.4285714 0.4285714

```
> 100*steadyStates(t(CM)) # estado estacionario (%) = distribución límite de estados (%)
      NYC      BCN
[1,] 57.14286 42.85714

> V <- eigen(P); vaps <- V$values; vaps[2]/vaps[1]

[1] 0.3
```

