

Sesión 3

Cadenas de Markov.

En la teoría de la probabilidad, se conoce como cadena de Markov a un tipo especial de proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende del evento inmediatamente anterior. Esta dependencia del evento anterior distingue a las cadenas de Markov de las series de eventos independientes, como tirar una moneda al aire o un dado.

Una secuencia $\{X_n\}_n$ de variables aleatorias con valores en un espacio finito Ω se denomina *cadena de Markov en tiempo discreto* cuando

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n),$$

donde x_i se denomina estado del sistema en el instante i .

Para simplificar la notación, supondremos que el espacio finito de estados es $\Omega = \{1, \dots, n\}$.

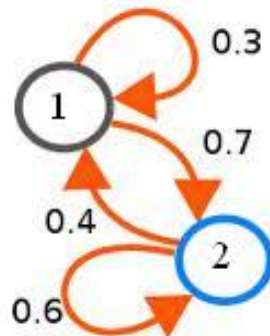
En esta práctica consideraremos cadenas de Markov *homogéneas*, en las que la probabilidad de pasar del estado i al j no depende del instante de tiempo en que nos encontramos:

$$P(X_{n+1} = j | X_n = i) = P(X_2 = j | X_1 = i).$$

En ese caso, las probabilidades de pasar de un estado i a otro estado j pueden resumirse mediante la llamada *matriz de transición* T , donde $T_{ij} = P(X_{n+1} = j | X_n = i)$.

3.1. Cálculos probabilísticos con RCommander

Consideremos una cadena de Markov sencilla, como la representada en la siguiente figura:



Esto quiere decir que la matriz de transición asociada es:

$$T = \begin{pmatrix} 0,3 & 0,7 \\ 0,4 & 0,6 \end{pmatrix}$$

De esta forma, si inicialmente estamos en el estado 1 con probabilidad 0.5 y en el estado 2 con probabilidad 0.5, al cabo de una iteración las probabilidades serán:

$$\begin{pmatrix} 0,5 & 0,5 \end{pmatrix} \cdot \begin{pmatrix} 0,3 & 0,7 \\ 0,4 & 0,6 \end{pmatrix} = \begin{pmatrix} 0,35 & 0,65 \end{pmatrix}$$

De la misma forma, podríamos calcular las probabilidades de estar en los estados 1 y 2 al cabo de n iteraciones: sería

$$\begin{pmatrix} 0,5 & 0,5 \end{pmatrix} \cdot \begin{pmatrix} 0,3 & 0,7 \\ 0,4 & 0,6 \end{pmatrix}^n$$

Para realizar estos cálculos con RCommander, basta con crear una función que realice la multiplicación de matrices.

Para crear una matriz, usamos la instrucción:

```
matrix(data, nrow, ncol, byrow)
```

La instrucción `byrow` indica si los datos se van rellenando por filas o columnas.

Para multiplicar dos matrices, basta con usar la instrucción `%*%`: nótese que si usamos solamente la multiplicación se realiza elemento a elemento.

```
inic=matrix(c(0.5,0.5), 1, 2, byrow=T)

trans=matrix(c(0.3,0.7,0.4,0.6),2,2,byrow=T)

inic %*% trans
  [,1] [,2]
[1,] 0.35 0.65
```

Así, las probabilidades de los estados al cabo de 5 iteraciones serían:

```
for (i in 1:5){
  inic=inic %*% trans}

> inic
  [,1] [,2]
[1,] 0.363635 0.636365
```

3.2. La deriva genética

Consideramos a continuación un ejemplo procedente de la Biología, llamado la *deriva genética* (genetic drift). Con este nombre se designan los cambios en la distribución de frecuencias de los distintos alelos de un gen debido al muestreo aleatorio. Los alelos en los descendientes son una muestra de la población de los padres, y la teoría de la probabilidad puede utilizarse para determinar si un gen determinado va a sobrevivir o no. Debido a que la distribución en la

generación $n + 1$ depende únicamente de la de la generación n -ésima, pero no de las anteriores, la evolución de esta distribución a lo largo del paso de las generaciones cumple la condición de Markov.

Podemos ilustrar el proceso de deriva genética a través de urnas y bolas. Supongamos que tenemos una urna A con 20 bolas, 15 de las cuales son rojas y 5 de las cuales son azules. Cada paso de una generación a otra puede ser modelizado por un muestreo con reemplazamiento: seleccionamos una bola al azar de la urna A y colocamos en la urna B una bola del mismo color. A continuación, devolvemos la bola extraída a la urna A . Si repetimos el proceso 20 veces, obtenemos una urna B con 20 bolas entre rojas y azules, pero en donde la composición puede haber cambiado.



Al tratarse de un muestreo con reemplazamiento, el número de bolas azules en la segunda generación seguiría una distribución de probabilidad binomial $B(20, p)$, siendo $p = 5/20$; en generaciones sucesivas debemos actualizar la p . Tendríamos así

$$P(X_{j+1} = k | X_j = x) = P(B(20, \frac{x}{20}) = k) \quad \forall k = 0, \dots, 20.$$

Veamos cómo simular el proceso anterior con RCommander:

```
alelos=function(n,x,m){#creamos una función dependiente del total n de individuos,
                        #el número inicial x de individuos de la clase de interés
                        #y el número de generaciones a estudiar.

for(i in 1:m)

x=rbinom(1,n,x/n) #actualizamos la proporción de bolas azules
                  #mediante la fórmula anterior.

return(x)}
```

Uno de los aspectos más interesantes en este tipo de problemas consiste en calcular la probabilidad de supervivencia de un determinado tipo de alelo después de m generaciones. Nótese que si el alelo desaparece en alguna de las m repeticiones ya desaparece para todas las generaciones siguientes.

Al igual que en los ejemplos anteriores, podemos estimar dicha probabilidad mediante simulación:

```
n=20
x=5
```

```

m=5 #valores iniciales y número de repeticiones

superv=rep(0,21) #creamos el vector en que almacenaremos las frecuencias

for(j in 1:1000){ z=alelos(n,x,m)

superv[z+1]=superv[z+1]+1}

```

Obtendríamos por ejemplo

```

> superv[1]/1000
[1] 0.188

```

lo que significa que en 188 de las 1000 simulaciones al final no había ninguna bola azul en la urna, es decir, el alelo correspondiente había desaparecido.

Consideremos a continuación una situación ligeramente más complicada, en la que el rojo hace el papel análogo (aunque no exactamente igual) al de un gen dominante y el azul el de gen recesivo. Concretamente, para generar un nuevo individuo de la siguiente urna vamos a seleccionar dos bolas; si ambas son azules, entonces la nueva bola es de color azul; para cualquier otra combinación, la nueva bola es roja. En ese caso, la probabilidad de que una nueva bola fuese azul si tenemos x bolas azules de n sería

$$p = P\left(B\left(2, \frac{x}{n}\right) = 2\right) = \left(\frac{x}{n}\right)^2,$$

y el número de bolas azules en la siguiente urna seguiría por tanto una distribución binomial $B(20, p)$.

De acuerdo con esto, una simulación del número de bolas azules al cabo de 5 generaciones sería la siguiente:

```

alelosmod=function(n,x,m){
for(i in 1:m)
x=rbinom(1,n,(x/n)^2)
return(x)}

```

Como era de esperar, de esta forma es mucho más probable que las bolas azules acaben desapareciendo de la urna. De hecho, incluso aunque empezemos con 15 bolas azules de las 20, al cabo de 5 generaciones es casi seguro que van a desaparecer:

```

n=20

x=15

m=5

superv=rep(0,21)

for(j in 1:1000){

```

```
z=alelosmod(n,x,m)
```

```
superv[z+1]=superv[z+1]+1}
```

```
> superv[1]/1000
```

```
[1] 0.927
```

3.3. Ejercicios

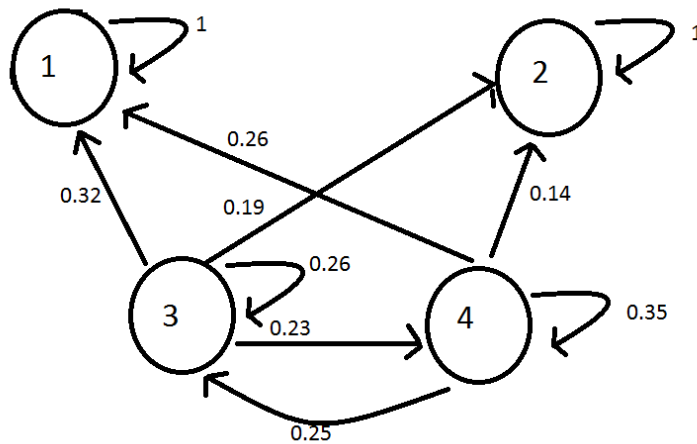
1. En su informe anual, la agencia de rating Standard & Poor's estima mediante técnicas estadísticas la probabilidad de transición de las calificaciones de los bonos al cabo de un año. La siguiente tabla da la matriz de transición correspondiente al año 2001:

Initial ratings	Ratings at year-end							
	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	0.9366	0.0583	0.0040	0.0009	0.0002	0	0	0
AA	0.0066	0.9172	0.0694	0.0049	0.0006	0.0009	0.0002	0.0002
A	0.0007	0.0225	0.9176	0.0518	0.0049	0.0020	0.0001	0.0004
BBB	0.0003	0.0026	0.0483	0.8924	0.0444	0.0081	0.0016	0.0023
BB	0.0003	0.0006	0.0044	0.0666	0.8323	0.0746	0.0105	0.0107
B	0	0.0010	0.0032	0.0046	0.0572	0.8362	0.0384	0.0594
CCC	0.0015	0	0.0029	0.0088	0.0191	0.1028	0.6123	0.2526
Default	0	0	0	0	0	0	0	1.0000

- (a) Determina la matriz de transición para la evolución de un bono en un período de 5 años.
 - (b) Calcula la probabilidad de que la evolución de una calificación sea AAA-AA-A-AA-AA.
 - (c) Supongamos que compramos un bono con calificación B. Calcula la probabilidad de que se haya convertido en bono basura (Default) antes de 10 años.
2. Una empresa agroindustrial ¹ aplica cadenas de Markov para estimar la probabilidad de pago entre distintos tipos de clientes morosos. En una de las categorías, se consideran cuatro estados de deuda:
 - 1- Deudas pagadas.
 - 2- Deudas incobrables.
 - 3- Deudas con antigüedad de un año.
 - 4- Deudas con antigüedad de dos años,

y se estima que las probabilidades de transición de unos estados a otros vienen dadas por el siguiente grafo:

¹Fuente: http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/Cabanillas_CE/Cap2.PDF



- (a) Determina la matriz de transición.
- (b) Calcula la probabilidad de que una deuda con antigüedad de dos años esté pagada al cabo de 8 años.