

Sesión 4

Contrastes de hipótesis sobre cadenas de Markov

En la práctica anterior, vimos cómo realizar cálculos con una cadena de Markov a partir de su matriz de transición. En esta práctica, veremos como estudiar la bondad de ajuste de unos datos a este tipo de modelos, con varios conjuntos de datos reales.

4.1. Estudio del ajuste a una cadena de Markov

En las sesiones 3 y 4, estudiamos uno de los procesos estocásticos a tiempo discreto más importantes: las cadenas de Markov. Como vimos en la práctica 3, 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 es el estado del sistema en el instante i .

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

Una cadena de Markov se dice *homogénea* cuando la probabilidad de pasar del estado i al j no depende del instante de tiempo en que nos encontramos:

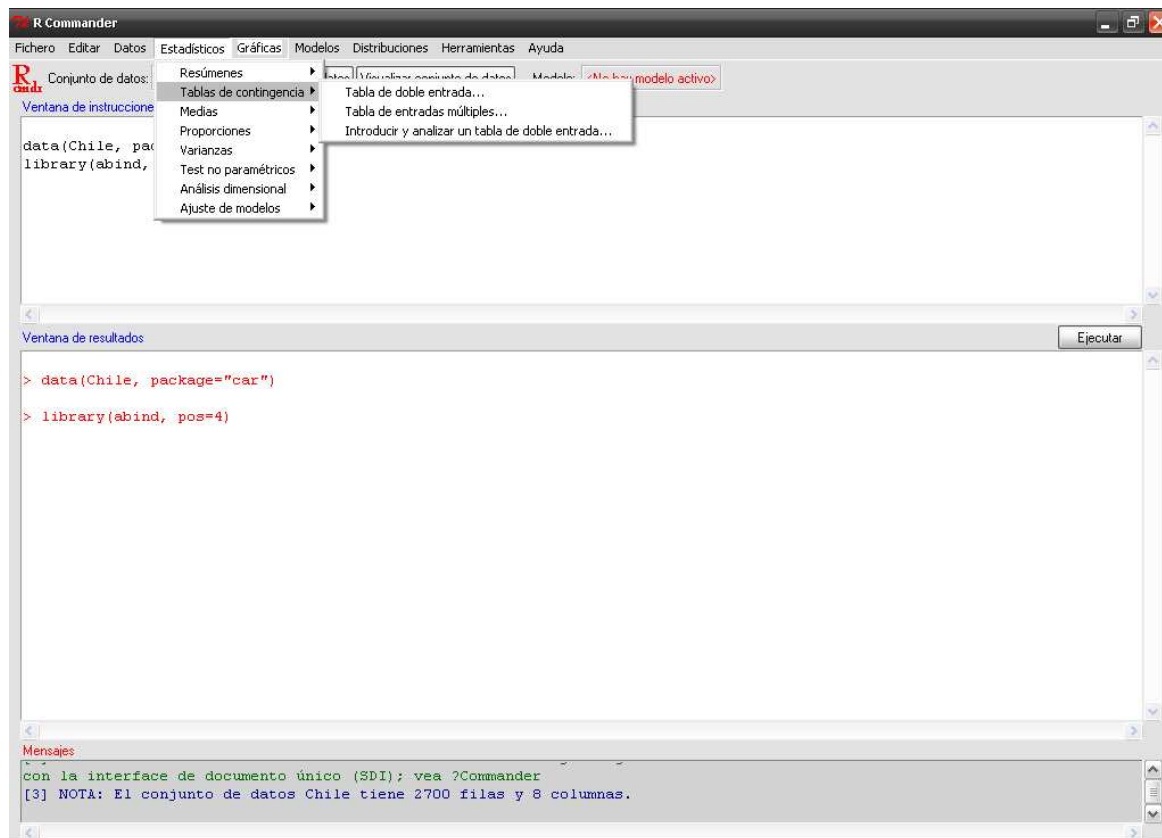
$$P(X_{n+1} = j | X_n = i) = P(X_2 = j | X_1 = i).$$

En muchas ocasiones, antes de aplicar los resultados teóricos existentes acerca de las cadenas de Markov es necesario comprobar estadísticamente estas dos condiciones (la condición de Markov y la homogeneidad). Podemos hacerlo de manera sencilla a partir de los contrastes de hipótesis de homogeneidad que habéis visto en la asignatura de Inferencia Estadística.

Ejemplo 4.1. *Uno de los campos en los que las cadenas de Markov han tenido aplicación es de la meteorología. En esta práctica vamos a considerar las secuencias de días lluviosos y soleados en Oviedo durante el otoño de 2012, donde un día se considera lluvioso ($X=1$) cuando la precipitación es superior a 0.1mm por m^2 . Consideramos la secuencia de ceros y unos entre el 1 de noviembre de 2012 y el 24 de enero de 2013.*

Para comprobar la hipótesis de Markov, vamos a estudiar si la distribución de probabilidad de $X_{t+1} | X_t$ coincide con la de $X_{t+1} | X_t, X_{t-1}$. Para ello, recopilamos los datos para cada una de las combinaciones de valores de X_t, X_{t-1} , y aplicamos un test χ^2 de homogeneidad. Podemos

hacer esto yendo a Estadísticos → Tablas de contingencia → Introducir una tabla de doble entrada. El número de filas representa el número de variables (distribuciones) que estamos comparando, mientras que el número de columnas representa el número de valores posibles de cada una de estas distribuciones (en este caso 2).



El test χ^2 de contingencia compara las frecuencias observadas en cada clase con las que esperaríamos caso de que la hipótesis de igualdad de distribuciones fuese cierta. Para aplicarlo, debemos realizar los siguientes pasos:

- Con la opción Datos → Cargar conjunto de datos cargamos el fichero *Lluvia.Rdata*.
- Creamos un vector de ceros y unos que nos indica si el día ha sido seco (0) o lluvioso (1):

```
x=rep(0,85)
```

```
for (i in 1:85)
```

```
if (Datos[i,2]>0.1) x[i]=1
```

- Calculamos las frecuencias absolutas de las secuencias (0,1), (0,0,1) y (1,0,1). Por ejemplo, para ver las frecuencias de (0,0) y de (0,1), haríamos lo siguiente:

```
freq00=0
```

```

freq01=0
for (i in 1:84){
  if (x[i]==0 & x[i+1]==1)
    freq01=freq01+1
  else {if (x[i]==0 & x[i+1]==0) freq00=freq00+1} }

```

obteniendo

```

> freq01
[1] 17

```

```

> freq00
[1] 25

```

De manera análoga se obtendría la distribución de frecuencias condicionada a (0,0) y a (1,0):

```

freq000=0
freq001=0
freq101=0
freq100=0

for (i in 1:83){

  if (x[i]==0 & x[i+1]==0 & x[i+2]==0) {freq000=freq000+1} else

  if (x[i]==0 & x[i+1]==0 & x[i+2]==1) {freq001=freq001+1} else

  if (x[i]==1 & x[i+1]==0 & x[i+2]==1) {freq101=freq101+1} else

  if (x[i]==1 & x[i+1]==0 & x[i+2]==0) {freq100=freq100+1} }

```

Se obtiene

```

> freq001
[1] 9

```

```

> freq000
[1] 16

```

```

> freq101
[1] 8

```

```

> freq100
[1] 8

```

- *Con estas frecuencias, introducimos y analizamos la tabla de doble entrada. En este caso, obtenemos la siguiente salida:*

data: .Table X-squared = 0.7878, df = 1, p-value = 0.3748

de donde deducimos que, al ser el p -valor mayor que 0.05, no rechazamos la hipótesis de homogeneidad en las distribuciones. Para estar seguros de que se cumple la condición de Markov, deberíamos estudiar de manera análoga si la distribución de $X_{t+1}|X_t = 1$ es independiente del valor de X_{t-1} .

4.2. Ejercicios

1. (a) Utiliza los datos del ejemplo 4.1 y contrasta si la distribución de $X_{t+1}|X_t = 1$ es independiente del valor de X_{t-1} .
(b) Consideremos de nuevo el fichero *Lluvia*, en el que aparecen los datos de precipitación en Oviedo entre el 1 de noviembre de 2012 y el 24 de enero de 2013. Supongamos que ahora clasificamos los días en tres grupos: secos (precipitación $< 0,1\text{mm}$), lluvioso (precipitación entre 0.1 y 5mm) y muy lluvioso (precipitación $> 5\text{mm}$). Obtén la tabla de frecuencias correspondiente y analiza si se cumple la hipótesis de Markov.
2. El fichero de datos *Acciones.Rdata* contiene los datos acerca de la evolución de las acciones de una compañía, en términos cualitativos (se consideran únicamente los valores ‘Baja’, y ‘Sube’).
 - (a) Crea un vector x que tome el valor 0 cuando la acción baja y 1 cuando sube.
 - (b) Utiliza este vector para estudiar si la evolución de la variable satisface la condición de Markov.