

Simulación computacional de cadenas de Markov

Presentación basada en el capítulo 3 de *Finite Markov Chains and Algorithmic Applications* (Haggström, 2002)

Sebastián Castro

Seminario de Probabilidad y Estadística
Centro de Matemática, Facultad de Ciencias, Udelar

Viernes 10 de mayo de 2013

Índice general

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Sección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Repaso de notación y definiciones

Para comenzar, repasamos algunas de las notaciones y definiciones correspondientes al capítulo 2 (*Markov chains*), del libro de Häggström.

Definición: Sea P una matriz $k \times k$ con elementos $\{P_{i,j} : i, j = 1, \dots, k\}$. Un proceso estocástico (X_0, X_1, \dots) con espacio de estados finito $S = \{s_1, \dots, s_k\}$ se dice que es una **cadena de Markov (homogénea)** con matriz de transición P , si para todo n , para todos $i, j \in \{1, \dots, k\}$ y todos $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$ se tiene

$$P(X_{n+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) =$$

$$P(X_{n+1} = s_j | X_n = s_i) = P_{i,j}$$

Repaso de notación y definiciones (cont.)

Las **matrices de transición** satisfacen $P_{i,j} \geq 0$ para todo $i, j \in \{1, \dots, k\}$, y $\sum_{j=1}^k P_{i,j} = 1$ para todo $i \in \{1, \dots, k\}$.

La **distribución inicial** es representada como un vector fila

$$\mu^{(0)} = \left(\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_k^{(0)} \right) =$$

$$\left(P(X_0 = s_1), P(X_0 = s_2), \dots, P(X_0 = s_k) \right).$$

Teorema

Para una cadena de Markov (X_0, X_1, \dots) con espacio de estados $\{s_1, \dots, s_k\}$, distribución inicial $\mu^{(0)}$ y matriz de transición P , tenemos que la **distribución a tiempo n** satisface $\mu^{(n)} = \mu^{(0)} P^n$.

Repaso de notación y definiciones (cont.)

Definición: Sean $P^{(1)}, P^{(2)}, \dots$ una secuencia de matrices de transición $k \times k$. Un proceso estocástico (X_0, X_1, \dots) con espacio de estados finito $S = \{s_1, \dots, s_k\}$ se dice que es una **cadena de Markov no homogénea** con matrices de transición $P^{(1)}, P^{(2)}, \dots$, si para todo n , todos $i, j \in \{1, \dots, k\}$ y todos $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$ se tiene

$$P(X_{n+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) =$$

$$P(X_{n+1} = s_j | X_n = s_i) = P_{i,j}^{(n+1)}$$

Repaso de notación y definiciones (cont.)

El siguiente resultado es una generalización del teorema anterior, y permite calcular las distribuciones $\mu^{(1)}, \mu^{(2)}, \dots$ a tiempos $1, 2, \dots$ de una cadena de Markov no homogénea.

Teorema

Sea (X_0, X_1, \dots) una cadena de Markov no homogénea con espacio de estados $\{s_1, \dots, s_k\}$, distribución inicial $\mu^{(0)}$ y matrices de transición $P^{(1)}, P^{(2)}, \dots$. Para cada n la **distribución a tiempo n** satisface

$$\mu^{(n)} = \mu^{(0)} P^{(1)} P^{(2)} \dots P^{(n)}$$

Sección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Introducción

- ▶ Un tema clave en muchas (sino la mayoría) de las aplicaciones prácticas de la teoría de Markov, es la habilidad de simular cadenas de Markov en una computadora.

- ▶ Comenzamos diciendo una “mentira”:

en la mayoría de los lenguajes de programación de alto nivel se tiene acceso a algún generador de números aleatorios que produce una secuencia U_0, U_1, \dots de variables aleatorias iid, uniformemente distribuidas en el intervalo $[0, 1]$.

- ▶ Esta afirmación no es cierta por al menos dos razones:

Introducción (cont.)

- (A) Los números U_0, U_1, \dots obtenidos de un generador de números aleatorios no están distribuidos uniformemente en $[0,1]$. En realidad tienen una expresión binaria (o decimal) finita y son, por lo tanto, racionales. Por el contrario, se puede mostrar que una variable aleatoria que esté (verdaderamente) distribuida uniformemente en $[0, 1]$ (o de hecho cualquier variable aleatoria continua) es irracional con probabilidad 1.
- (B) ¡Los valores U_0, U_1, \dots no son ni siquiera aleatorios! En realidad, son generados por algún procedimiento determinístico. Por esta razón, los generadores de números aleatorios son algunas veces (y con mayor precisión) llamados generadores de números *pseudo-aleatorios*.

Introducción (cont.)

- ▶ La más importante de estas objeciones es (B), porque (A) tiende a no ser un problema muy grande cuando el número de dígitos binarios o decimales es razonablemente grande.
- ▶ En las últimas décadas, se ha puesto un gran esfuerzo en construir generadores de números (pseudo-) aleatorios cuyo resultado sea tan indistinguible como fuera posible de una verdadera secuencia *iid* de variables aleatorias uniformes en $[0, 1]$.

Introducción (cont.)

- ▶ Hoy en día existen generadores que parecen hacer ésto muy bien (pasando todos los tests estadísticos estándar para tales generadores). .
- ▶ Por esta razón simplemente haremos el supuesto (incorrecto) de que tenemos acceso a una secuencia *iid* de variables aleatorias uniformes $[0, 1]$.
- ▶ Siempre deberíamos tener en mente que ésta es una fuente potencial de errores en la simulación por computadora.

Sección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Simulación de cadenas de Markov finitas

- ▶ Entonces, ¿cómo simulamos una cadena de Markov (X_0, X_1, \dots) con espacio de estados $S = \{s_1, \dots, s_k\}$, distribución inicial $\mu^{(0)}$ y matriz de transición P ?
- ▶ Los números aleatorios U_0, U_1, \dots forman un ingrediente principal.
- ▶ Los otros ingredientes principales son dos funciones, que llamamos **función de iniciación** y **función de actualización**.

Subsección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Función de iniciación

La *función de iniciación* $\psi : [0, 1] \rightarrow S$ es una función del intervalo unitario al espacio de estados S , la cual se utiliza para generar el valor inicial X_0 .

Asumimos que:

- (i) ψ es constante por tramos,
- (ii) para cada $s \in S$, el largo total de los intervalos en los cuales $\psi(x) = s$ es igual a $\mu^{(0)}(s)$.

Otra forma de establecer la propiedad (ii) es

$$\int_0^1 I_{\{\psi(x)=s\}} dx = \mu^{(0)}(s),$$

para cada $s \in S$, donde $I_{\{\psi(x)=s\}}$ es la función indicatriz del conjunto $\{\psi(x) = s\}$.

Función de iniciación (cont.)

Si tenemos esta función ψ podemos generar X_0 del primer número aleatorio U_0 haciendo $X_0 = \psi(U_0)$.

Esto da la distribución correcta de X_0 , dado que para cada $s \in S$ tenemos que

$$P(X_0 = s) = P(\psi(U_0) = s) = \int_0^1 I_{\{\psi(x)=s\}} dx = \mu^{(0)}(s).$$

Llamamos a ψ una *función de iniciación válida* para la cadena de Markov (X_0, X_1, \dots) si $\int_0^1 I_{\{\psi(x)=s\}} dx = \mu^{(0)}(s)$ para todo $s \in S$.

Función de iniciación (cont.)

Funciones de iniciación válidas son fáciles de construir: con espacio de estados $S = \{s_1, \dots, s_k\}$ y distribución inicial $\mu^{(0)}$, podemos establecer:

$$\psi(x) = \begin{cases} s_1 & \text{para } x \in [0, \mu^{(0)}(s_1)) \\ s_2 & \text{para } x \in [\mu^{(0)}(s_1), \mu^{(0)}(s_1) + \mu^{(0)}(s_2)) \\ \vdots & \vdots \\ s_i & \text{para } x \in \left[\sum_{j=1}^{i-1} \mu^{(0)}(s_j), \sum_{j=1}^i \mu^{(0)}(s_j) \right) \\ \vdots & \vdots \\ s_k & \text{para } x \in \left[\sum_{j=1}^{k-1} \mu^{(0)}(s_j), 1 \right] \end{cases}$$

Función de iniciación (cont.)

Necesitamos verificar que esta elección de ψ satisface las propiedades (i) y (ii) mencionadas anteriormente. La propiedad (i) es directa y para (ii) verificamos:

$$\int_0^1 I_{\{\psi(x)=s_i\}} dx = \sum_{j=1}^i \mu^{(0)}(s_j) - \sum_{j=1}^{i-1} \mu^{(0)}(s_j) = \mu^{(0)}(s_i),$$

para $i = 1, \dots, k$. Esto significa que ψ es una función de iniciación válida para la cadena de Markov (X_0, X_1, \dots) y que puede utilizarse para generar el valor inicial X_0 .

Si también resolvemos cómo generar X_{n+1} a partir de X_n para cualquier n , podemos usar este procedimiento iterativamente para obtener toda la cadena (X_0, X_1, \dots) .

Subsección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Función de actualización

Para pasar de X_n a X_{n+1} utilizamos el número aleatorio U_{n+1} y una *función de actualización* $\phi : S \times [0, 1] \rightarrow S$, que toma como entrada un estado $s \in S$ y un número entre 0 y 1, y devuelve otro estado $s' \in S$ como salida.

En forma similar que para la función de iniciación ψ , necesitamos que ϕ cumpla ciertas propiedades:

- (i) para un s_i fijo, la función $\phi(s_i, x)$ es constante por tramos (como función de x),
- (ii) para cada $s_i, s_j \in S$ fijos, el largo total de los intervalos en los cuales $\phi(s_i, x) = s_j$ es $P_{i,j}$.

Función de actualización (cont.)

Nuevamente, la propiedad (ii) puede escribirse como

$$\int_0^1 I_{\{\phi(s_i, x) = s_j\}} dx = P_{i,j}$$

para todo $s_i, s_j \in S$. Si la función de actualización satisface esta propiedad, entonces

$$\begin{aligned} P(X_{n+1} = s_j | X_n = s_i) &= P(\phi(s_i, U_{n+1}) = s_j | X_n = s_i) = \\ &P(\phi(s_i, U_{n+1}) = s_j) = \int_0^1 I_{\{\phi(s_i, x) = s_j\}} dx = P_{i,j} \end{aligned}$$

donde se ha utilizado la independencia de U_{n+1} con (U_0, \dots, U_n) , y por lo tanto, con X_n .

Función de actualización (cont.)

Una función ϕ que satisfaga (i) y (ii), se dice que es una *función de actualización válida* para la cadena de Markov (X_0, X_1, \dots) .

La construcción de una función de actualización válida no es más difícil que la construcción de una función de iniciación válida. Por ejemplo, podemos establecer para cada $s_j \in S$

$$\phi(s_j, x) = \begin{cases} s_1 & \text{para } x \in [0, P_{i,1}) \\ s_2 & \text{para } x \in [P_{i,1}, P_{i,1} + P_{i,2}) \\ \vdots & \vdots \\ s_j & \text{para } x \in \left[\sum_{l=1}^{j-1} P_{i,l}, \sum_{l=1}^j P_{i,l} \right) \\ \vdots & \vdots \\ s_k & \text{para } x \in \left[\sum_{l=1}^{k-1} P_{i,l}, 1 \right) \end{cases}$$

Función de actualización (cont.)

Para ver que ésta es una función de actualización válida, notar que para cada $s_i, s_j \in S$ tenemos

$$\int_0^1 I_{\{\phi(s_i, x) = s_j\}} dx = \sum_{l=1}^j P_{i,l} - \sum_{l=1}^{j-1} P_{i,l} = P_{i,j}$$

Por lo tanto, tenemos un procedimiento completo para simular una cadena de Markov:

- (1) construir las funciones de iniciación y actualización válidas ψ y ϕ ,
- (2) hacer $X_0 = \psi(U_0)$, $X_1 = \phi(X_0, U_1)$, $X_2 = \phi(X_1, U_2)$, \dots

Observación: las funciones de iniciación y de actualización no son únicas.

Sección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Simulación de cadenas de Markov no homogéneas finitas

Al finalizar el capítulo, se señala cómo el método descrito anteriormente puede ser generalizado para simular una cadena de Markov no homogénea.

Sea (X_0, X_1, \dots) una cadena de Markov no homogénea con espacio de estados $S = \{s_1, \dots, s_k\}$, distribución inicial $\mu^{(0)}$ y matrices de transición $P^{(0)}, P^{(1)}, \dots$

Podemos obtener la función de iniciación ψ y el valor inicial X_0 como en el caso homogéneo. La función de actualización es construida en forma similar al caso homogéneo, excepto que debido a que la cadena es no homogénea, necesitamos varias funciones de actualización diferentes $\phi^{(1)}, \phi^{(2)}, \dots$, y para éstas debemos tener

$$\int_0^1 I_{\{\phi^{(n)}(s_i, x) = s_j\}} dx = P_{ij}^{(n)},$$

para cada n y cada $s_i, s_j \in S$.

Simulación de cadenas de Markov no homogéneas finitas (cont.)

Tales funciones pueden obtenerse por la obvia generalización:

$$\phi^{(n)}(s_i, x) = \begin{cases} s_1 & \text{para } x \in [0, P_{i,1}^{(n)}) \\ s_2 & \text{para } x \in [P_{i,1}^{(n)}, P_{i,1}^{(n)} + P_{i,2}^{(n)}) \\ \vdots & \vdots \\ s_j & \text{para } x \in \left[\sum_{l=1}^{j-1} P_{i,l}^{(n)}, \sum_{l=1}^j P_{i,l}^{(n)} \right) \\ \vdots & \vdots \\ s_k & \text{para } x \in \left[\sum_{l=1}^{k-1} P_{i,l}^{(n)}, 1 \right] \end{cases}$$

La cadena de Markov no homogénea es luego simulada haciendo:

$$X_0 = \psi(U_0), X_1 = \phi^{(1)}(X_0, U_1), X_2 = \phi^{(2)}(X_1, U_2), \dots$$

Subsección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Ejemplo 1: Modelo de predicción del clima en Los Angeles

Como primer ejemplo se considera la modelización del clima en la ciudad de Los Angeles.

Para simplificar se supone que hay solamente dos estados:

$E = \{s_1, s_2\}$ ($s_1 =$ lluvioso, $s_2 =$ soleado), y que la sucesión de estados climáticos se puede modelar con una cadena de Markov (X_0, X_1, \dots) con matriz de transición:

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{pmatrix}$$

Es decir, por ejemplo, que $P(X_{n+1} = \text{soleado} | X_n = \text{lluvioso}) = 0.5$ y $P(X_{n+1} = \text{lluvioso} | X_n = \text{soleado}) = 0.1$, para todo n .

A partir de una distribución inicial *uniforme* en el espacio de estados, nos proponemos simular una sucesión de $n = 1000$ estados bajo el modelo asumido.

Ejemplo 1: Modelo de predicción del clima en Los Angeles (cont.)

```
> estados <- c('s1', 's2') # espacio de estados
> mu0 <- c(.5, .5) # distribucion inicial
> names(mu0) <- estados
> P <- rbind(c(.5, .5), c(.1, .9)) # matriz de transición
> dimnames(P) <- list(estados, estados)
> P
      s1 s2
s1 0.5 0.5
s2 0.1 0.9

> (func.inic <- cumsum(mu0)) # función de iniciación
      s1 s2
0.5 1.0

> (func.act <- t(apply(P, 1, cumsum))) # función de actualización
      s1 s2
s1 0.5 1
s2 0.1 1
```

Ejemplo 1: Modelo de predicción del clima en Los Angeles (cont.)

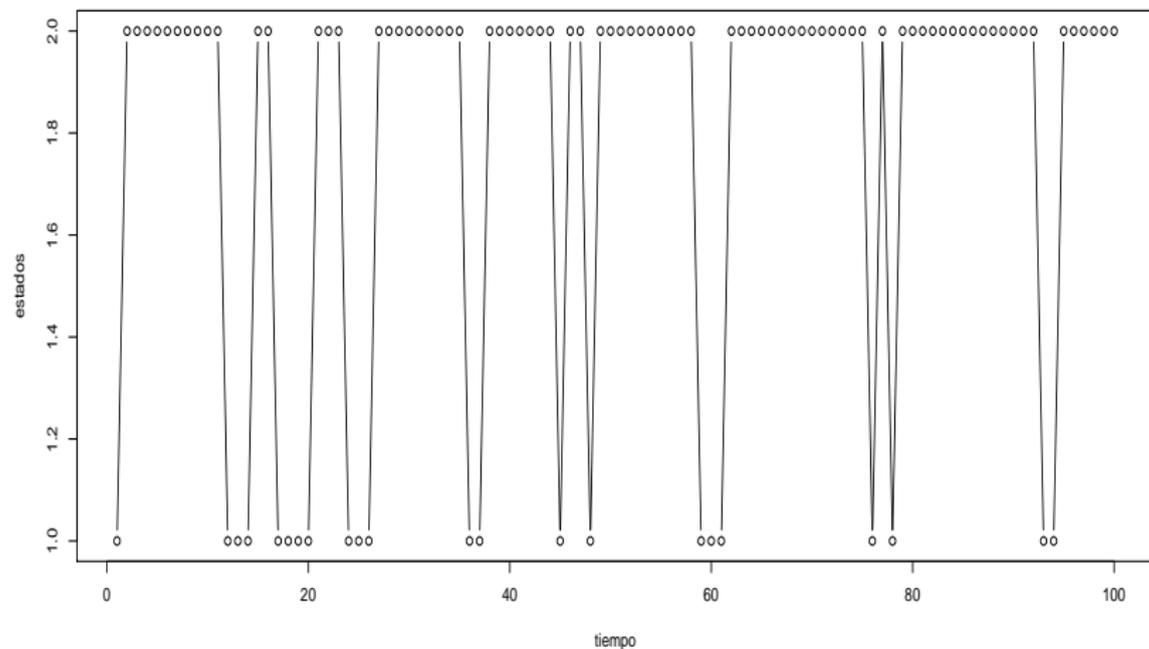
```
n <- 1000 # número de simulaciones deseadas
U <- runif(n) # uniformes U1,...,Un en [0,1]
X <- numeric(n) * NA # vector de valores simulados de la cadena, a completar

# primer valor de la cadena
j = 1; while(U[1] > func.inic[j]) j = j + 1;
X[1] <- estados[j]

# restantes valores de la cadena
for (i in 2:n) {
  j = 1; while(U[i] > func.act[X[i - 1], j]) j = j + 1;
  X[i] <- estados[j]
}

# primeros 25 valores simulados de la cadena
> X[1:25]
[1] "s1" "s2" "s1" "s1" "s1" "s2"
[16] "s2" "s1" "s1" "s1" "s1" "s2" "s2" "s2" "s1" "s1"
```

Ejemplo de trayectoria simulada de la cadena de Markov



Ejemplo 1: Modelo de predicción del clima en Los Angeles

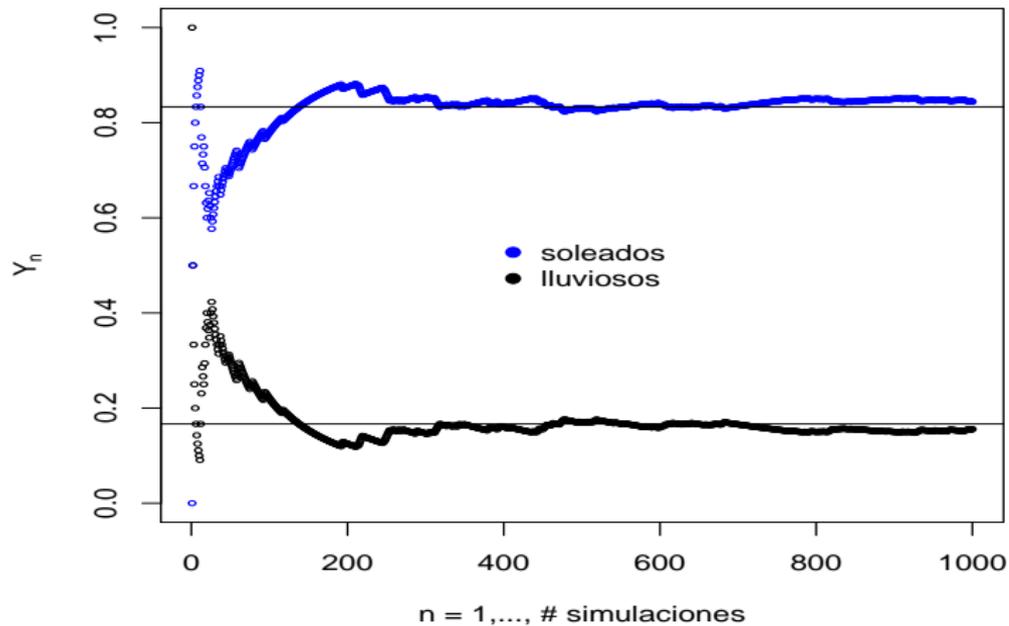
A partir de los n valores simulados, se define el estadístico Y_n que cuenta la proporción de días lluviosos hasta el día n . Es decir:

$$Y_n = \frac{1}{n+1} \sum_{i=0}^n I_{\{X_i = \text{lluvioso}\}}$$

¿Qué sucede con Y_n a medida que n aumenta?

```
> Yn <- cumsum(X == 's1') / 1:length(Yn)
> Yn[1:25]
 [1] 1.0000000 0.5000000 0.3333333 0.2500000 0.2000000 0.1666667
 [7] 0.1428571 0.1250000 0.1111111 0.1000000 0.0909090 0.1666667
[13] 0.2307692 0.2857142 0.2666667 0.2500000 0.2941176 0.3333333
[19] 0.3684210 0.4000000 0.3809523 0.3636363 0.3478260 0.3750000
[25] 0.4000000
```

Proporción acumulada de visitas a cada estado



Observaciones y “conjetura”

La observación del sistema simulado a través del modelo permite generar algunas hipótesis sobre el mismo. Por ejemplo, si

$$\mu^{(n)} = (P(X_n = \text{lluvioso}), P(X_n = \text{soleado})),$$

y la distribución inicial es uniforme entonces

$$\mu^{(n)} \rightarrow (1/6, 5/6), \text{ cuando } n \rightarrow \infty?$$

¿y si la distribución inicial no es uniforme?

En este caso existen formas precisas de responder a estas preguntas (próximos capítulos).

Subsección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

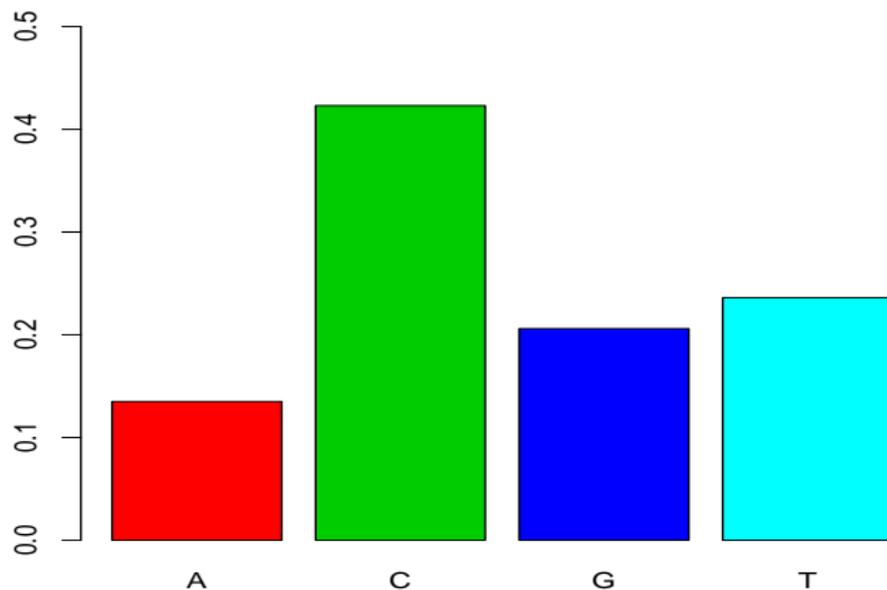
Se considera una cadena de Markov como modelo para la sucesión de nucleótidos en el ADN. El espacio de estados es entonces $E = \{A, C, G, T\}$ ($A = \text{adenina}$, $C = \text{citosina}$, $G = \text{guanina}$, $T = \text{timina}$), y a modo de ejemplo se utiliza la matriz de transición entre estados

$$P = \begin{pmatrix} 0.2 & 0.25 & 0.4 & 0.15 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0.35 & 0.35 \\ 0.1 & 0.45 & 0.2 & 0.25 \end{pmatrix}$$

Utilizando una distribución inicial uniforme, se simulan $n = 1000$ observaciones de esta cadena, como en el ejemplo anterior, a través de la función `rMarkov` implementada en R:

```
X <- rMarkov(P, 1000)
```

Frecuencia relativa de visitas a cada estado



Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

Para cada par de estados $i, j \in E$, se cuentan la cantidad de transiciones del estado i al estado j , divididas por el número de transiciones de i a cualquier otro estado.

$$\hat{P}_{i,j} = \frac{\sum_{l=0}^{n-1} I_{\{X_l=i \ \& \ X_{l+1}=j\}}}{\sum_{l=0}^{n-1} I_{\{X_l=i\}}}$$

$$\hat{P} = \begin{pmatrix} 0.22 & 0.22 & 0.39 & 0.17 \\ 0.11 & 0.61 & 0.09 & 0.19 \\ 0.16 & 0.11 & 0.36 & 0.38 \\ 0.11 & 0.48 & 0.17 & 0.23 \end{pmatrix}$$

$$P = \begin{pmatrix} 0.2 & 0.25 & 0.4 & 0.15 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0.35 & 0.35 \\ 0.1 & 0.45 & 0.2 & 0.25 \end{pmatrix}$$

Subsección actual

0. Repaso de notación y definiciones

Cadenas de Markov homogéneas y no homogéneas, matrices de transición, distribución inicial, distribución a tiempo n

1. Introducción a la simulación de variables aleatorias

2. Simulación de cadenas de Markov homogéneas finitas

2.1 Función de iniciación

2.2 Función de actualización

3. Simulación de cadenas de Markov no homogéneas finitas

4. Implementación en lenguaje R

4.1 Ejemplo 1: Modelo de predicción del clima en Los Angeles

4.2 Ejemplo 2: Modelo para la secuencia de nucleótidos en la cadena de ADN

4.3 Función en R para simular una cadena de Markov

Función en R para simular una cadena de Markov

```
rMarkov <- function(P, n = 100, mu0 = rep(1, nrow(P))/nrow(P)) {  
  
  # P: matriz de transición (sin valor por defecto)  
  # n: número de simulaciones (100 por defecto)  
  # mu0: distribución inicial (uniforme en el espacio de estados por defecto)  
  
  # si el espacio de estados está en los nombres de las columnas de P, usarlo  
  if(length(colnames(P)) == 0) estados <- 1:ncol(P) else estados <- colnames(P)  
  
  # función de iniciación:  
  func.inic <- cumsum(mu0)  
  
  # funciones de actualización:  
  func.act <- t(apply(P, 1, cumsum))  
  
  U <- runif(n) # uniformes  $U_1, \dots, U_n$  en  $[0,1]$   
  X <- numeric(n)*NA # vector de valores simulados de la cadena, a completar  
  
  # primer valor de la cadena  
  j = 1; while(U[1] > func.inic[j]) j = j + 1;  
  X[1] <- estados[j]  
  
  # restantes valores de la cadena  
  for (i in 2:n) {  
    j = 1; while(U[i] > func.act[X[i - 1], j]) j = j + 1;  
    X[i] <- estados[j]  
  }  
  X  
}
```