CADENAS DE MARKOV

Calixtro Ames Cynthia Lizeth 1, Nesiosup Vilca Marcia Paulina 2
Huamaní Inga Alex Raúl 3, Romero Gómez Yasmyn Stephany 4

E.P.Ciencia de la computación 1*, Facultad de Ciencias* 1*, Universidad Nacional de Ingeniería* 1*,*
e-mail:ccalixtroa@uni.pe
E.P.Matemática 2*, Facultad de Ciencias* 2*, Universidad Nacional de Ingeniería* 2*,*
e-mail:marcia.nesiosup@gmail.com
E.P.Matemática 3*, Facultad de Ciencias* 3*, Universidad Nacional de Ingeniería* 3*, e- mail:ahuamanii@uni.pe*
E.P.Matemática 4*, Facultad de Ciencias* 4*,Universidad Nacional de Ingeniería*4*, e- mail:*

Resumen

Una cadena de Markov es una serie de eventos, en la cual la probabilidad de que ocurra un evento depende del evento inmediato anterior. En efecto, las cadenas de este tipo tienen memoria, "*recuerdan*" el último evento y esto condiciona las posibilidades de los eventos futuros. 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. En el presente estudio implementamos la simulación de 3 diferentes tipos de cadenas de Markov (irreductibles,periódicas y estacionarias).

Palabras Claves:Cadenas, Markov, Implementación, Estocástico.

Keywords: Chains, Markov, Implementation, Estochastic.

1. INTRODUCCIÓN

Para una mejor comprensión de las cadenas de Markov y la posterior simulación de algunos ejemplos empezaremos por establecer algunos conceptos previos requeridos.

1.1. CONCEPTOS

Procesos Estocásticos

Un proceso estocástico se puede definir equivalentemente de las siguientes dos maneras.
1. Como un conjunto de realizaciones temporales y un índice aleatorio que selecciona una de ellas.
2. Como un conjunto de variables aleatorias X_t indexadas por un índice t , dado que $t \in T$ con $T \subseteq \mathbb{R}$.

Un proceso estocástico se dice de ***tiempo continuo*** si T es un intervalo (usualmente este intervalo se toma como $[0, \infty\rangle$) o de ***tiempo discreto*** si T es un conjunto numerable (solamente puede asumir determinados valores, usualmente se toma $T \subseteq \mathbb{N}$),de donde el conjunto T sera llamado ***espacio parametral*** y las variables aleatorias toman valores en un conjunto S llamado **espacio de estados**.

Trayectoria de un proceso estocástico

Una *trayectoria* de proceso estocástico $x_t : t \in T$ es una función

$$t \longmapsto X_t(w)]$$

para cada $w \in \omega$.
Existen distintos tipos de procesos estocásticos,estos se obtienen al considerar:

- Distintos espacios parametrales.
- Distintos espacios de estado.
- Distintas características de la trayectorias.
- Distintas relaciones de dependencia estocástica entre las variables aleatorias que conforman el proceso.

Veamos algunos ejemplos de procesos estocásticos.

- **Procesos de ensayos independientes**
  Supongamos que el espacio parametral $T = 1, 2, \ldots$ es el conjunto de números naturales y $S \subseteq \mathbb{R}$, cuando las variables aleatorias $(X_1, X_2, \ldots)$ son independientes se dice que este es un proceso estocástico de ensayos independientes (*Recordar que una colección infinita de variables es independiente si toda subcolección finita lo es.*)

- **Procesos de Markov**
  El proceso $X_t : t = 0, 1, \ldots$ (donde $X_t \in S$ es discreto) es un **proceso de Markov** si para cada $x_0, x_1, \ldots, x_{n+1} en S$ se cumple la siguiente propiedad:

  $$\begin{aligned} P(X_{n+1} =& x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) \\ &= P(X_{n+1} = x_{n+1} | X_n = x_n) \end{aligned}$$

  Es decir la historia del proceso hasta el tiempo n-1 es irrelevante cuando se conoce el estado del proceso al tiempo n a esta identidad se le conoce con el nombre de **propiedad de Markov** y es un buen ejemplo del tipo de dependencia estocástica que puede existir entre las variables de un proceso, en este caso las variables aleatorias son discretas y al proceso lo llamaremos **Cadenas de Markov**.

- **Procesos con incrementos independientes**
  El proceso $X_t : t \geq 0$ tiene **incrementos independientes** si las variables aleatorias

  $$X_{t1}, X_{t2} - X_{t1}, \ldots, X_{tn} - Xtn - 1$$

  son independientes para cualesquiera tiempos $0 \leq t_1 < t_2 < \ldots < t_n$.

- **Procesos estacionarios**
  El proceso $X_t : t \geq 0$ es **estacionario** en el sentido estricto si para cualesquiera $t_1, \ldots, t_n, h \geq 0$

  $$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h})$$

  de donde el vector aleatorio del lado derecho tiene la misma distribucion que el vector aleatorio del lado izquierdo. Esto significa que el vector aleatorio del lado izquierdo se puede trasladar en el tiempo como indica el vector del lado derecho sin que cambie su distribucion de probabilidad.

- **Procesos con incrementos estacionarios**
  El proceso $X_t : t \geq 0$ tiene **incrementos estacionarios** si para cualesquiera $0 \leq S < t; h \geq 0$

  $$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h}$$

- **Martingalas**
  El proceso $X_t : t = 0, 1, \ldots$ es una **Martingala** si

  $$E(X_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = x_n$$

pero en realidad hay otras condiciones tecnicas que se piden a un proceso para que sea un martingala.

- **Procesos de Lévy**
  El procese $X_t : t \geq 0$ es un **proceso de Lévy** si sus incrementos son:
  - Independientes
  - Estacionarios
- **Procesos Gausianos**
  El proceso $X_t : t \geq 0$ es un **proceso gausiano** si para cualesquiera tiempos $t_1, \ldots, t_n$

$$(X_{t_1}, \ldots, X_{t_n}) \sim \text{Normal multivariada}$$

No debe pensarse que estas características de los procesos son excluyentes unas de otras,existen varios procesos estocásticos que cumplen varias de estas propiedades al mismo tiempo, por ejemplo el movimiento Browniano que es un proceso estocástico muy importante es a la vez un proceso **gausiano** con **incrementos independientes** y **estacionarios** por lo tanto es un proceso de **Lévy** y puede además demostrarse que es una **Martingala** y que cumple la **propiedad de Markov**.

Con un panorama más amplio de conocimiento acerca de los procesos estocásticos entraremos a profundidad en el concepto y diversos ejemplos de cadenas de Markov.

1.2. CADENAS DE MARKOV IRREDUCTIBLES Y APERIÓDICAS

Definición 1. Espacio de estados
Definimos el *espacio de estados* como el conjunto formado por todos los posibles valores que puede tomar X_n donde el índice n representa la evolución del proceso en el *tiempo*. Para una cadena de Markov tanto el *espacio de estados* como el *tiempo* pueden ser discreto o continuo.

Definición 2. Cadena de Markov
Una cadena de Markov es una secuencia de variables aleatorias $X_0, X_1, X_2, ...$ que toman valores en en *espacio de estados* $\{1, 2, \ldots, M\}$ si para todo $n \geq 0$ se cumple la propiedad de Markov. Donde $P(X_{n+1} = x_{n+1} | X_n = x_n)$ se denomina probabilidad de transición desde el estado i hacia el estado j. Como esta probabilidad no depende de el valor de n se dice que el tiempo es *homogeneo.*
Definición 3. Matriz de transición
Sean $X_0, X1, X_2, \ldots$ una cadena de Markov con espacio de estados $\{1, 2, \ldots, M\}$ y sea

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

la probabilidad de transición del estado i al estado j. Entonces, a la matriz MxM, $P = \{p_{ij}\}$ se le denomina *matriz de transición* de la cadena. Además, la suma de los valores por fila es igual a 1.

Definición 4. Probabilidad de transición en n pasos La probabilidad de transición en n pasos de i hacia j es la probabilidad de ir de i hacia j en n pasos. Denotamos esto como $p_{ij}^{(n)}$ y se define de la siguiente manera:

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$

y se cumple que $p_{ij}^{(n)}$ es el elemento de subíndices (i, j) de la matriz P^n

Definición 5. Distribución Marginal de X_n
Sea el vector fila $q = (q_1, \ldots, q_M)$ con $q_i = P(X_0 = i)$ entonces la *distribución marginal* de X_n está dada por el vector qP^n. Es decir, el j-ésimo componente del vector qP^n es $P(X_n = j)$

Tipos de cadenas de Markov:

1. ***Cadena de Markov Irreductible*** Se dice que una cadena de Markov con matriz de transición P es *irreductible* si para dos estados i y j cualesquiera existe un entero positivo n tal que el elemento de índices (i, j) en la matriz P^n es positivo. Es decir, es posible ir de i a j en un número finito de pasos. En caso contrario se dirá que la cadena de Markov es *reducible.*

2. ***Cadena de Markov Periódica***

   Definición 6. Periodo de un estado
   Se define el periodo d_i de el estado i como

   $$d_i = gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$$

   Esto es, el *gcd* (*máximo común divisor*) del posible número de pasos que puede tomar para regresar al mismo estado, en este caso, al estado i. Cuando un estado tiene periodo igual a 1 entonces se dirá que es *aperiódico*, en caso contrario, se le llamará *periódico.*

   Una cadena de Markov es llamada *aperiódica* cuando todos sus estados son aperiódicos. En caso contrario, se le denomina *periódica.*

1.3. Distribuciones estacionarias

Un vector fila $\pi = (\pi_1, \ldots, \pi_M)$ con $\pi_i \geq 0$ and $\sum_i \pi_i = 1$ es una *distribución estacionaria* para una cadena de Markov con *matriz de transición P* si

$$\pi P = \pi$$

Si π es la distribución de X_0, esto es si $\pi_i = P(X_0 = i)$, entonces πP vendría a ser la distribución marginal de X_1. Esto significa que, si X_0 tiene distribución π, entonces X_1 también tendrá una distribución π. Si X_2 tiene una distribución π entonces X_3 también la tendrá y así sucesivamente. Por lo tanto, una distribución estacionaria de una cadena de Markov es una distribución de probabilidad que se mantiene sin cambios en la cadena a medida que avanza el tiempo.

1.4. Convergencia

Ya hemos declarado informalmente que la distribución estacionaria describe el comportamiento a largo plazo de la cadena, en el sentido de que si ejecutamos la cadena por un largo tiempo, la distribución marginal de X converge a la distribución estacionaria S. El siguiente el teorema establece que esto es cierto siempre que la cadena sea tanto irreductible como aperiódica. Entonces, independientemente de las condiciones iniciales de la cadena, el PMF de Xn convergerá a la distribución estacionaria cuando $n \longrightarrow \infty$. theorema 11.3.6(Convergencia a la distribucion estacionaria). Sea $X_0, X_1, \ldots$ sea una cadena de Markov con distribucion estacionaria S y matrix de transicion Q, tal que la potencia Q^n tiene todas las entradas positivas. Entonces $P(X_n = i)$ converge a S_i mientras $n \longrightarrow \infty$.. En terminos de la matriz de transicion, Q^n converge a una matriz en la que cada fila es S.

$$\begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}^n \rightarrow \begin{pmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{pmatrix} \textit{cuando } n \rightarrow \infty$$

1.5. Reversibilidad

Hemos visto que la distribución estacionaria de una cadena de Markov es extremadamente util para entender su comportamiento a largo plazo. Desafortunamente en general puede ser computacionalmente dificil de encontrar la distribución estacionaria cuando el espacio de estados es grande.

"Definition"Sea $Q = (q_{ij})$ sea la matrix de transicion de una cadena de Markov, Suponiendo que es $S = (S_1, \ldots, S_M)$ con $S_i \geq 0, \sum_i S_i = 1$ tal que

$$S_i q_{ij} = S_j q_{ji}$$

Para todos los estados i and j. Esta ecuacion es llamada *reversibilidad* o *Detalle de la condición de equilibrio*, y nosotros diremos que la cadena *reversible* respecto a S si la contiene.

Dada una matriz de transición, si podemos encontrar un vector no negativo s cuyas componentes suman 1 y que satisfaga la condición de reversibilidad, entonces s es automáticamente una distribución estacionaria.

2. IMPLEMENTACIÓN

A. Simulación de una cadena de Markov

```r
rMarkov <- function(P, n = 100, mu0 = rep(1, nrow(P))/nrow(P)) {
# P: matriz de transici´on (sin valor por defecto)
# n: n´umero de simulaciones (100 por defecto)
# mu0: distribuci´on inicial (uniforme en el espacio de estados por defecto)
# si el espacio de estados est´a en los nombres de las columnas de P, usarlo
if(length(colnames(P)) == 0) estados <- 1:ncol(P) else estados <- colnames(P)
# funci´on de iniciaci´on:
func.inic <- cumsum(mu0)
# funciones de actualizaci´on:
func.act <- t(apply(P, 1, cumsum))
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]
}
X
}
```

Figura 1: Función que simula una cada de Markov.

Si consideramos un espacio de estados $E = \{A, C, G, T\}$ (A=adenina,C=citosina,G=guanina,T=timina)

```r
#sea la matriz de transición entre estados P
vector=c(.2,.25,.4,0.15,.1,.6,.1,.2,.2,.1,.35,.35,.1,.45,.2,.25)
P=matrix(vector,4,4,T)
P
```

0.2	0.25	0.40	0.15
0.1	0.60	0.10	0.20
0.2	0.10	0.35	0.35
0.1	0.45	0.20	0.25

Figura 2: Matriz de transición de la cadena de Markov.

```r
X=rMarkov(P,1000)
X[1:100]
```

```
4 3 4 3 4 2 1 4 1 2 2 4 4 2 3 1 3 4 2 4 4 3 1 4 4 2 2 2 2 1 4 2 4 4 3 3 2 2 4 2 2 2 4 2 2
2 2 1 2 2 4 3 3 1 2 2 2 1 2 2 2 2 2 2 2 3 4 2 2 2 1 2 4 3 2 4 2 1 3 1 2 2 2 1 3 3 1 3 3 3
3 2 2 3 3 4 4 2 2 1
```

Figura 3: Probando la función *rMarkov*.

```r
plot(1:100, X[1:100],type= 'b',col = 'blue',lwd = 1,las =.2, ylab= 'Estados',xlab='Tiempo',lty=10)
```

Figura 4: Graficando la cadena de Markov.

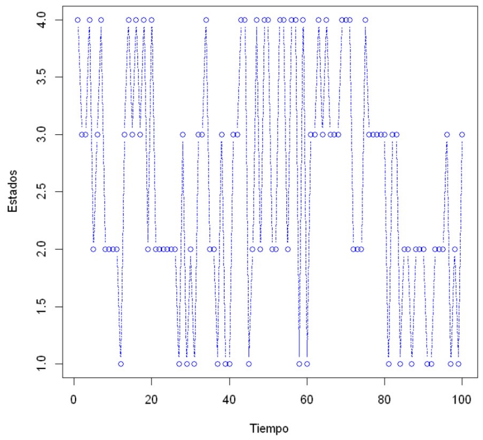

Figura 5: Grafica Tiempo VS Estados.

Referencias

[1] Autor, *Ttulo*, Revista/Editor, (ao)