

El método Markov chain Monte Carlo

IIC3810

Marcelo Arenas y Luis Alberto Croquevielle

Una pregunta pendiente del capítulo anterior

¿Cómo podemos construir un generador (casi) uniforme para una relación?

Recuerde el problema KS definido en el capítulo anterior y la relación:

$$R_{KS} = \{((\vec{a}, b), \vec{x}) \mid \vec{a} \in \mathbb{N}^n \text{ y } \vec{x} \in \{0, 1\}^n \text{ para } n \geq 1, b \in \mathbb{Z} \text{ y } \vec{a} \cdot \vec{x} \leq b\}$$

Vamos a responder primero una pregunta más específica: ¿Cómo podemos construir un generador (casi) uniforme para R_{KS} ?

- La respuesta a esta pregunta va a tener los ingredientes necesarios para responder la pregunta más general

Una secuencia de variables aleatorias para generar R_{KS}

Fije $n \geq 1$, $\vec{a} \in \mathbb{N}^n$ y $b \in \mathbb{N}$

- Y suponga que $\Omega = \{\vec{x} \in \{0, 1\}^n \mid \vec{a} \cdot \vec{x} \leq b\}$

Nótese que $\Omega \neq \emptyset$

Considere una secuencia $\{X_t\}_{t \in \mathbb{N}}$ de variables aleatorias con recorrido Ω

- El dominio de cada variable X_t es D_t , el cual no necesitamos definir

Decimos que Ω es el conjunto de estados de la secuencia $\{X_t\}_{t \in \mathbb{N}}$

Una secuencia de variables aleatorias para generar R_{KS}

Dado $t \in \mathbb{N}$ y $\vec{x} \in \Omega$, nos interesa calcular $\Pr(X_t = \vec{x})$

Para calcular esta probabilidad necesitamos definir la dinámica de la secuencia

- Vale decir, necesitamos definir cómo se cambio de estado al pasar de tiempo t a tiempo $t + 1$

De manera formal, dado $t \in \mathbb{N}$ y $\vec{x}, \vec{y} \in \Omega$, necesitamos entregar:

$$\Pr(X_{t+1} = \vec{y} \mid X_t = \vec{x})$$

Una secuencia de variables aleatorias para generar R_{KS}

Suponga que en tiempo t estamos en el estado $\vec{x} = (x_1, \dots, x_n)$

El estado $\vec{y}$ en el tiempo $t + 1$ se obtiene utilizando el siguiente procedimiento:

GenerarSiguiente($\vec{x}$)

Escoja $c \in \{0, 1\}$ con distribución uniforme

if $c = 0$ **then return** $\vec{x}$

else

Escoja $i \in \{1, \dots, n\}$ con distribución uniforme

$\vec{u} := (x_1, \dots, x_{i-1}, 1 - x_i, x_{i+1}, \dots, x_n)$

if $\vec{a} \cdot \vec{u} \leq b$ **then return** $\vec{u}$

else return $\vec{x}$

Tenemos entonces que $\vec{y} = \text{GenerarSiguiente}(\vec{x})$

Algunas propiedades de la secuencia

Ejercicios

Sea $t \in \mathbb{N}$

1. Dados $\vec{x}, \vec{y} \in \Omega$, calcule $\Pr(X_{t+1} = \vec{y} \mid X_t = \vec{x})$
 - Considere de manera separada los casos $\vec{x} \neq \vec{y}$ y $\vec{x} = \vec{y}$

2. Demuestre que para todo $\vec{x}, \vec{y} \in \Omega$, existe $t' > t$ tal que:

$$\Pr(X_{t'} = \vec{y} \mid X_t = \vec{x}) > 0$$

3. Dados $\vec{x}_0, \vec{x}_1, \dots, \vec{x}_t, \vec{x}_{t+1} \in \Omega$, demuestre que:

$$\Pr(X_{t+1} = \vec{x}_{t+1} \mid X_0 = \vec{x}_0 \wedge X_1 = \vec{x}_1 \wedge \dots \wedge X_t = \vec{x}_t) = \Pr(X_{t+1} = \vec{x}_{t+1} \mid X_t = \vec{x}_t)$$

¿A qué converge la secuencia?

Sea $\vec{x}$ un vector arbitrario en Ω

Suponga que en tiempo 0 estamos en el estado $\vec{x}$, y que en cada instante cambiamos de estado utilizando el procedimiento **GenerarSiguiente**

- ▶ ¿Cuáles son los estados a los que podríamos llegar en un tiempo $t \gg 0$?
¿Cuál es la probabilidad de estar en un estado específico en este tiempo t ?
- ▶ ¿Es posible llegar a una distribución estacionaria, vale decir, un tiempo t' tal que para todo $\vec{y} \in \Omega$ se tiene que $\mathbf{Pr}(X_{t'+1} = \vec{y}) = \mathbf{Pr}(X_{t'} = \vec{y})$?
- ▶ ¿Existe una única distribución estacionaria?

¿Cuáles son las respuestas a las preguntas anteriores si cambiamos $\vec{x}$ por otro vector inicial?

La secuencia converge a la distribución uniforme

Ejercicios

1. Dado $t \in \mathbb{N}$ y $\vec{x} \in \Omega$, demuestre que:

$$\sum_{\vec{y} \in \Omega} \Pr(X_{t+1} = \vec{y} \mid X_t = \vec{x}) = 1$$

$$\sum_{\vec{y} \in \Omega} \Pr(X_{t+1} = \vec{x} \mid X_t = \vec{y}) = 1$$

2. Demuestre que la distribución uniforme es una distribución estacionaria.

- Vale decir, demuestre que si para $t \in \mathbb{N}$ se tiene que

$\Pr(X_t = \vec{x}) = \frac{1}{|\Omega|}$ para cada $\vec{x} \in \Omega$, entonces:

$$\Pr(X_{t+1} = \vec{y}) = \Pr(X_t = \vec{y}) \text{ para cada } \vec{y} \in \Omega$$

Generando R_{KS} con distribución (casi) uniforme

Para generar los elementos de Ω con distribución (casi) uniforme utilizamos el siguiente procedimiento:

```
Sea  $\vec{x}$  un elemento arbitrario de  $\Omega$   
 $t = f(|x|)$   
for  $i := 1$  to  $t$  do  
     $\vec{x} := \mathbf{GenerarSiguiente}(\vec{x})$   
return  $\vec{x}$ 
```

¿Qué condiciones deben cumplirse para que este procedimiento genere Ω con distribución (casi) uniforme y en tiempo polinomial?

Las condiciones para la generación (casi) uniforme

- ▶ La secuencia debe converger a la distribución uniforme desde cualquier punto de partida $\vec{x} \in \Omega$
- ▶ La distribución uniforme debe ser la única distribución a la que converge la secuencia
- ▶ $f(n)$ debe estar acotada superiormente por un polinomio

Las condiciones para la generación (casi) uniforme

- ▶ Debe ser posible calcular **GenerarSiguiente**($\vec{x}$) en tiempo polinomial
- ▶ Después de ejecutar t pasos se debe tener una garantía de que estamos cerca de la distribución uniforme
 - ▶ Obtenemos entonces un generador casi uniforme para los elementos de Ω

Todas estas condiciones han sido estudiadas para las cadenas de Markov.

- ▶ Vamos a introducir y estudiar esta herramienta esencial para el muestreo de variables aleatorias

Las cadenas de Markov

Considere una sucesión $\{X_t\}_{t \in \mathbb{N}}$ de variables aleatorias.

Definición

$\{X_t\}_{t \in \mathbb{N}}$ es una cadena de Markov con conjunto de estados Ω si:

1. Ω es un conjunto finito o infinito enumerable, y $X_t : D_t \rightarrow \Omega$ para cada $t \in \mathbb{N}$
2. Existe $p : \Omega \times \Omega \rightarrow [0, 1]$ tal que para cada $t \in \mathbb{N}$ y cada secuencia $a_0, \dots, a_t, a_{t+1}$ de elementos de Ω :

$$\begin{aligned} \Pr(X_{t+1} = a_{t+1} \mid X_t = a_t \wedge \dots \wedge X_0 = a_0) &= \\ \Pr(X_{t+1} = a_{t+1} \mid X_t = a_t) &= p(a_{t+1}, a_t) \end{aligned}$$

Las cadenas de Markov

En una cadena de Markov la distribución de probabilidades en el tiempo $t + 1$ sólo depende de la distribución de probabilidades en el tiempo t

- Consideramos cadenas discretas: el tiempo es un conjunto infinito enumerable, y el conjunto de estados Ω es finito o infinito enumerable

Además, consideramos cadenas de Markov donde las probabilidades de transición no dependen del tiempo.

- Para cada $t_1, t_2 \in \mathbb{N}$ y $a, b \in \Omega$:

$$\Pr(X_{t_1+1} = a \mid X_{t_1} = b) = \Pr(X_{t_2+1} = a \mid X_{t_2} = b)$$

Estos son llamadas cadenas de Markov homogéneas.

En general, consideramos cadenas de Markov con conjunto de estados Ω finito.

Representando una cadena de Markov

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov con conjunto de estados Ω

- Suponemos además que $p : \Omega \times \Omega \rightarrow [0, 1]$ es la función que define las probabilidades de transición.

Podemos representar cada variable aleatoria X_t como un vector $\vec{x}_t$ tal que:

$$\text{para cada } a \in \Omega \text{ se tiene que } \vec{x}_t[a] = \mathbf{Pr}(X_t = a)$$

Además, podemos representar la cadena de Markov como una matriz P de $|\Omega| \times |\Omega|$:

$$\text{para cada } a, b \in \Omega, \text{ se tiene que } P[a, b] = p(a, b)$$

P es llamada la matriz de transición de la cadena de Markov.

Representando una cadena de Markov

Dado $a \in \Omega$ y $t \in \mathbb{N}$, tenemos que:

$$\begin{aligned}\mathbf{Pr}(X_{t+1} = a) &= \sum_{b \in \Omega} \mathbf{Pr}(X_{t+1} = a \mid X_t = b) \cdot \mathbf{Pr}(X_t = b) = \\ &= \sum_{b \in \Omega} p(a, b) \cdot \mathbf{Pr}(X_t = b) = \sum_{b \in \Omega} P[a, b] \cdot \mathbf{Pr}(X_t = b)\end{aligned}$$

Dado que $\vec{x}_{t+1}[a] = \mathbf{Pr}(X_{t+1} = a)$ y $\vec{x}_t[b] = \mathbf{Pr}(X_t = b)$ para cada $a, b \in \Omega$ y $t \in \mathbb{N}$, concluimos que:

$$P\vec{x}_t = \vec{x}_{t+1}$$

Algunas propiedades de la matriz de transición

Ejercicios

Sea P la matriz de transición de una cadena de Markov.

1. Demuestre que para cada columna de P , se tiene que la suma de sus valores es 1
2. Para cada fila de P , ¿se debe tener que la suma de sus valores es 1?
 - ¿Era cierta esta propiedad para la matriz P de la cadena de Markov para R_{KS} ?

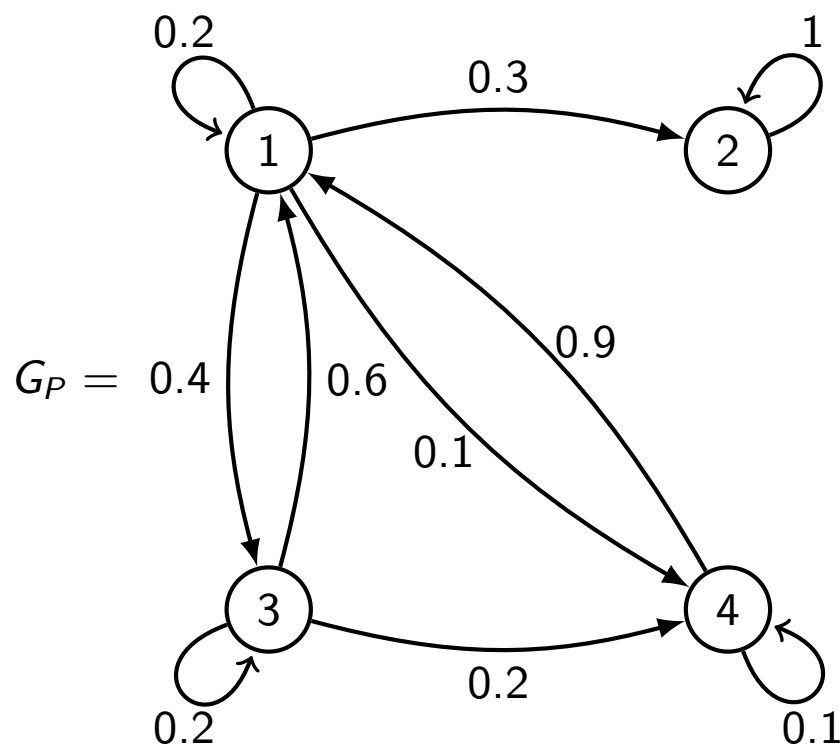
Una cadena de Markov como un grafo

La matriz de transición P de una cadena de Markov con conjunto de estados Ω puede ser vista como un grafo G_P con pesos:

- ▶ Ω es el conjunto de nodos de G_P
- ▶ Dados $a, b \in \Omega$, el peso del arco (a, b) en G es $P[b, a]$
 - ▶ Vale decir, el peso de (a, b) representa la probabilidad de pasar al estado b dado que estábamos en el estado a

Una cadena de Markov como un grafo: un ejemplo

$$\Omega = \{1, 2, 3, 4\} \quad P = \begin{pmatrix} 0.2 & 0 & 0.6 & 0.9 \\ 0.3 & 1 & 0 & 0 \\ 0.4 & 0 & 0.2 & 0 \\ 0.1 & 0 & 0.2 & 0.1 \end{pmatrix}$$



Algunas propiedades de las cadenas de Markov

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov con conjunto de estados Ω y matriz de transición P

- ▶ Además, defina P^0 como la matriz identidad y $P^{t+1} = PP^t$ para todo $t \in \mathbb{N}$

Ejercicios

1. Demuestre que para cada $t \in \mathbb{N}$ y $a, b \in \Omega$ se tiene que:

$$\Pr(X_t = a \mid X_0 = b) = P^t[a, b]$$

2. Demuestre para cada $t_1, t_2 \in \mathbb{N}$ y $a, b \in \Omega$ se tiene que:

$$\Pr(X_{t_1+t_2} = a \mid X_{t_1} = b) = \Pr(X_{t_2} = a \mid X_0 = b)$$

La distribución estacionaria de una cadena de Markov

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov con conjunto de estados Ω y matriz de transición P

► Y sea $\vec{\pi} \in [0, 1]^{|\Omega|}$ tal que $\sum_{a \in \Omega} \vec{\pi}[a] = 1$

Definición

$\vec{\pi}$ es una distribución estacionaria para $\{X_t\}_{t \in \mathbb{N}}$ si $P\vec{\pi} = \vec{\pi}$

Vale decir, $\vec{\pi}$ es una distribución estacionaria para $\{X_t\}_{t \in \mathbb{N}}$ si esta distribución no cambia al realizar una transición de la cadena de Markov

Algunos ejemplos de distribuciones estacionarias

Ejercicios

1. Considere una cadena de Markov con la siguiente matriz de transición:

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

Muestre que esta cadena de Markov tiene una única distribución estacionaria, y construya esta distribución.

2. Construya una cadena de Markov que tenga al menos dos distribuciones estacionarias.
 - ▶ El dominio de esta cadena debe ser finito
3. Construya una cadena de Markov que no tenga distribución estacionaria.
 - ▶ El dominio de esta cadena debe ser infinito enumerable

Una primera propiedad fundamental: irreducibilidad

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov con conjunto de estados Ω y matriz de transición P

Definición

$\{X_t\}_{t \in \mathbb{N}}$ es irreducible si G_P es un grafo fuertemente conexo.

Tenemos que $\{X_t\}_{t \in \mathbb{N}}$ es irreducible si y sólo si para cada $a, b \in \Omega$, existe $t > 0$ tal que:

$$\Pr(X_t = a \mid X_0 = b) > 0$$

Una segunda propiedad fundamental: aperiodicidad

Sea $a \in \Omega$. Si $\{n > 0 \mid \mathbf{Pr}(X_n = a \mid X_0 = a) > 0\}$ no es vacío, entonces el periodo de a es definido como:

$$\text{MCD } \{n > 0 \mid \mathbf{Pr}(X_n = a \mid X_0 = a) > 0\}$$

En caso contrario, el periodo de a no está definido.

Definición

Un estado $a \in \Omega$ es aperiódico si su periodo está definido y es igual a 1. Además, $\{X_t\}_{t \in \mathbb{N}}$ es aperiódica si cada estado $b \in \Omega$ es aperiódico.

Irreducibilidad y aperiodicidad: algunos ejemplos

Ejercicios

1. Muestre que la cadena de Markov definida para la relación R_{KS} es irreducible y aperiódica.
2. Demuestre que en una cadena de Markov irreducible todos los estados tienen el mismo periodo.
 - Concluimos entonces que si una cadena de Markov irreducible tiene un estado aperiódico, entonces la cadena es aperiódica.

Una solución para el segundo ejercicio

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov irreducible con conjunto de estados Ω y matriz de transición P , y sean $b, c \in \Omega$

Suponemos que los periodos de b y c son ℓ_b y ℓ_c , respectivamente.

- ¿Por qué sabemos que estos periodos existen?

Dado que $\{X_t\}_{t \in \mathbb{N}}$ es una cadena de Markov irreducible:

- Existe un camino en G_P desde b a c de largo $k_{b,c}$, y existe un camino en G_P desde c a b de largo $k_{c,b}$

Una solución para el segundo ejercicio

Entonces existe un camino de b a b de largo $k_{b,c} + k_{c,b}$, de lo cual concluimos que $\Pr(X_{k_{b,c}+k_{c,b}} = b \mid X_0 = b) > 0$

- ▶ Dado que ℓ_b es el periodo de b , concluimos que $\ell_b \mid (k_{b,c} + k_{c,b})$

Sea $n > 0$ tal que $\Pr(X_n = c \mid X_0 = c) > 0$

- ▶ Existe un camino en G_P de c a c de largo n

Tenemos entonces que existe un camino en G_P de b a b de largo $k_{b,c} + n + k_{c,b}$, por lo que $\Pr(X_{k_{b,c}+n+k_{c,b}} = b \mid X_0 = b) > 0$

- ▶ Dado que ℓ_b es el periodo de b , concluimos que $\ell_b \mid (k_{b,c} + n + k_{c,b})$
- ▶ Así, dado que $\ell_b \mid (k_{b,c} + k_{c,b})$, deducimos que $\ell_b \mid n$

Una solución para el segundo ejercicio

Por lo tanto, $\ell_b \mid n$ para cada $n > 0$ tal que $\Pr(X_n = c \mid X_0 = c) > 0$

- De lo cual deducimos que $\ell_b \leq \ell_c$, puesto que
 $\ell_c = \text{MCD} \{n > 0 \mid \Pr(X_n = c \mid X_0 = c) > 0\}$

De la misma forma se puede demostrar que $\ell_c \leq \ell_b$, de lo cual concluimos que $\ell_b = \ell_c$



Existencia de un límite: irreducibilidad y aperiodicidad

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov irreducible y aperiódica con conjunto de estados Ω

Teorema

Si $a \in \Omega$, se tiene que:

1. $\lim_{n \rightarrow \infty} \mathbf{Pr}(X_n = a \mid X_0 = b)$ existe para cada $b \in \Omega$
2. $\lim_{n \rightarrow \infty} \mathbf{Pr}(X_n = a \mid X_0 = b) = \lim_{n \rightarrow \infty} \mathbf{Pr}(X_n = a \mid X_0 = c)$ para cada $b, c \in \Omega$

Existencia de un límite: irreducibilidad y aperiodicidad

Dado $a \in \Omega$, definimos:

$$\lambda_a = \lim_{n \rightarrow \infty} \Pr(X_n = a \mid X_0 = b),$$

donde b es un elemento arbitrario en Ω

Lema

Si existe $a \in \Omega$ tal que $\lambda_a = 0$, entonces $\lambda_b = 0$ para todo $b \in \Omega$

Ejercicio

Demuestre el lema.

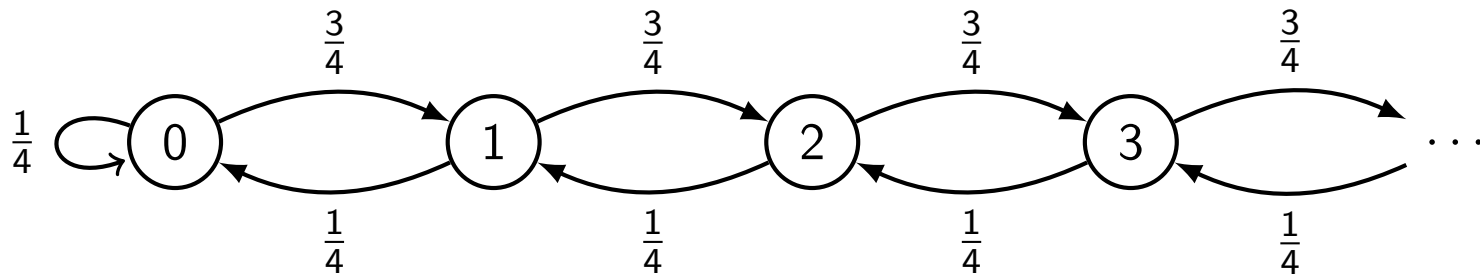
Irreducibilidad y aperiodicidad: un par de ejemplos

Ejercicios

1. Construya una cadena de Markov irreducible, aperiódica y tal que $\lambda_a = 0$ para todo estado a
2. Construya una cadena de Markov irreducible, aperiódica y tal que $\lambda_b > 0$ para todo estado b

Una solución al primer problema

Considere la siguiente cadena de Markov con conjunto de estado $\Omega = \mathbb{N}$:



Esta cadena de Markov es irreducible y aperiódica.

- Es fácil ver que el estado 0 es aperiódico, de lo cual se concluye que la cadena de Markov es aperiódica dado que es irreducible

Una solución al primer problema

Tenemos que:

$$\begin{aligned}\Pr(X_{n+1} = 0 \mid X_0 = 0) &= \Pr(X_{n+1} = 0 \mid X_n = 0) \cdot \Pr(X_n = 0 \mid X_0 = 0) + \\ &\quad \Pr(X_{n+1} = 0 \mid X_n = 1) \cdot \Pr(X_n = 1 \mid X_0 = 0) \\ &= \frac{1}{4} \cdot \Pr(X_n = 0 \mid X_0 = 0) + \frac{1}{4} \cdot \Pr(X_n = 1 \mid X_0 = 0)\end{aligned}$$

Tomando el límite cuando n tiende a infinito obtenemos:

$$\lambda_0 = \frac{1}{4} \cdot \lambda_0 + \frac{1}{4} \cdot \lambda_1$$

Por lo tanto: $3 \cdot \lambda_0 = \lambda_1$

Una solución al primer problema

Considere $i > 0$ y suponga que $3 \cdot \lambda_{i-1} = \lambda_i$

Tenemos que:

$$\begin{aligned}\Pr(X_{n+1} = i \mid X_0 = i) &= \Pr(X_{n+1} = i \mid X_n = i-1) \cdot \Pr(X_n = i-1 \mid X_0 = i) + \\ &\quad \Pr(X_{n+1} = i \mid X_n = i+1) \cdot \Pr(X_n = i+1 \mid X_0 = i) \\ &= \frac{3}{4} \cdot \Pr(X_n = i-1 \mid X_0 = i) + \frac{1}{4} \cdot \Pr(X_n = i+1 \mid X_0 = i)\end{aligned}$$

Una solución al primer problema

Tomando el límite cuando n tiende a infinito obtenemos:

$$\begin{aligned}\lambda_i &= \frac{3}{4} \cdot \lambda_{i-1} + \frac{1}{4} \cdot \lambda_{i+1} \\ &= \frac{3}{4} \cdot \frac{1}{3} \cdot \lambda_i + \frac{1}{4} \cdot \lambda_{i+1} \\ &= \frac{1}{4} \cdot \lambda_i + \frac{1}{4} \cdot \lambda_{i+1}\end{aligned}$$

Por lo tanto: $3 \cdot \lambda_i = \lambda_{i+1}$

Una solución al primer problema

Concluimos que $\lambda_i = 3^i \cdot \lambda_0$ para cada $i \in \mathbb{N}$

Así, dado que $\lambda_i \leq 1$ para cada $i \in \mathbb{N}$, se debe tener que $\lambda_0 = 0$

► Del lema concluimos entonces que $\lambda_i = 0$ para cada $i \in \mathbb{N}$ □

Irreducibilidad, aperiodicidad y el caso finito

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov irreducible, aperiódica y con conjunto de estados Ω finito

Lema

$$\sum_{a \in \Omega} \lambda_a = 1$$

Demostración: fije un elemento $b \in \Omega$

Para cada $n \in \mathbb{N}$ tenemos que:

$$\sum_{a \in \Omega} \mathbf{Pr}(X_n = a \mid X_0 = b) = 1$$

Irreducibilidad, aperiodicidad y el caso finito

Por lo tanto tenemos que:

$$\lim_{n \rightarrow \infty} \left(\sum_{a \in \Omega} \mathbf{Pr}(X_n = a \mid X_0 = b) \right) = 1$$

Además, sabemos que:

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\sum_{a \in \Omega} \mathbf{Pr}(X_n = a \mid X_0 = b) \right) &= \sum_{a \in \Omega} \lim_{n \rightarrow \infty} \mathbf{Pr}(X_n = a \mid X_0 = b) \\ &= \sum_{a \in \Omega} \lambda_a \end{aligned}$$

Concluimos que $\sum_{a \in \Omega} \lambda_a = 1$

□

Irreducibilidad, aperiodicidad y el caso finito

¿Por qué la demostración anterior no es válida si Ω es un conjunto infinito enumerable?

- ¿Qué paso de la demostración no funciona?

Como consecuencia de los resultados anteriores obtenemos:

Corolario

Si $\{X_t\}_{t \in \mathbb{N}}$ es una cadena de Markov irreducible, aperiódica y con un conjunto de estados Ω finito, entonces $\lambda_a > 0$ para cada $a \in \Omega$

Existencia y unicidad de la distribución estacionaria: el caso finito

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov irreducible, aperiódica, con conjunto de estados Ω finito y matriz de transición P

- Y sea $\vec{\pi} \in [0, 1]^{|\Omega|}$ un vector tal que $\vec{\pi}[a] = \lambda_a$ para cada $a \in \Omega$

Teorema

$\vec{\pi}$ es la única distribución estacionaria para $\{X_t\}_{t \in \mathbb{N}}$

Demostración: Primero tenemos que demostrar que $P\vec{\pi} = \vec{\pi}$

- Vale decir, tenemos que demostrar que $(P\vec{\pi})[a] = \vec{\pi}[a]$ para cada $a \in \Omega$

Existencia y unicidad en el caso finito: demostración

Dado $a \in \Omega$ y un estado arbitrario $c \in \Omega$, tenemos que:

$$\begin{aligned}(P\vec{\pi})[a] &= \sum_{b \in \Omega} P[a, b] \cdot \vec{\pi}[b] \\&= \sum_{b \in \Omega} \left(\Pr(X_1 = a \mid X_0 = b) \cdot \lim_{n \rightarrow \infty} \Pr(X_n = b \mid X_0 = c) \right) \\&= \lim_{n \rightarrow \infty} \sum_{b \in \Omega} \left(\Pr(X_1 = a \mid X_0 = b) \cdot \Pr(X_n = b \mid X_0 = c) \right) \\&= \lim_{n \rightarrow \infty} \sum_{b \in \Omega} \left(\Pr(X_{n+1} = a \mid X_n = b) \cdot \Pr(X_n = b \mid X_0 = c) \right) \\&= \lim_{n \rightarrow \infty} \Pr(X_{n+1} = a \mid X_0 = c) \\&= \lim_{m \rightarrow \infty} \Pr(X_m = a \mid X_0 = c) \\&= \lambda_a = \vec{\pi}[a]\end{aligned}$$

Existencia y unicidad en el caso finito: demostración

En segundo lugar, tenemos que demostrar que para toda distribución estacionaria $\vec{\alpha}$ para $\{X_t\}_{t \in \mathbb{N}}$, se tiene que $\vec{\alpha} = \vec{\pi}$

Suponga que $P\vec{\alpha} = \vec{\alpha}$. Dado $a \in \Omega$ tenemos que:

$$\begin{aligned}\vec{\alpha}[a] &= (P\vec{\alpha})[a] \\ &= \sum_{b \in \Omega} P[a, b] \cdot \vec{\alpha}[b] \\ &= \sum_{b \in \Omega} P[a, b] \cdot (P\vec{\alpha})[b] \\ &= \sum_{b \in \Omega} P[a, b] \cdot \left(\sum_{c \in \Omega} P[b, c] \cdot \vec{\alpha}[c] \right) \\ &= \sum_{b \in \Omega} \sum_{c \in \Omega} P[a, b] \cdot P[b, c] \cdot \vec{\alpha}[c] \\ &= \sum_{c \in \Omega} \left(\sum_{b \in \Omega} P[a, b] \cdot P[b, c] \right) \cdot \vec{\alpha}[c]\end{aligned}$$

Existencia y unicidad en el caso finito: demostración

$$\begin{aligned} &= \sum_{c \in \Omega} P^2[a, c] \cdot \vec{\alpha}[c] \\ &= \sum_{c \in \Omega} P^2[a, c] \cdot (P\vec{\alpha})[c] \\ &= \sum_{c \in \Omega} P^2[a, c] \cdot \left(\sum_{d \in \Omega} P[c, d] \cdot \vec{\alpha}[d] \right) \\ &= \sum_{c \in \Omega} \sum_{d \in \Omega} P^2[a, c] \cdot P[c, d] \cdot \vec{\alpha}[d] \\ &= \sum_{d \in \Omega} \left(\sum_{c \in \Omega} P^2[a, c] \cdot P[c, d] \right) \cdot \vec{\alpha}[d] \\ &= \sum_{d \in \Omega} P^3[a, d] \cdot \vec{\alpha}[d] \end{aligned}$$

Existencia y unicidad en el caso finito: demostración

Siguiendo con este proceso, concluimos que para todo $n \geq 1$:

$$\vec{\alpha}[a] = \sum_{b \in \Omega} P^n[a, b] \cdot \vec{\alpha}[b]$$

Por lo tanto:

$$\vec{\alpha}[a] = \sum_{b \in \Omega} \mathbf{Pr}(X_n = a \mid X_0 = b) \cdot \vec{\alpha}[b]$$

Existencia y unicidad en el caso finito: demostración

Tomando el límite cuando n tiende a infinito obtenemos:

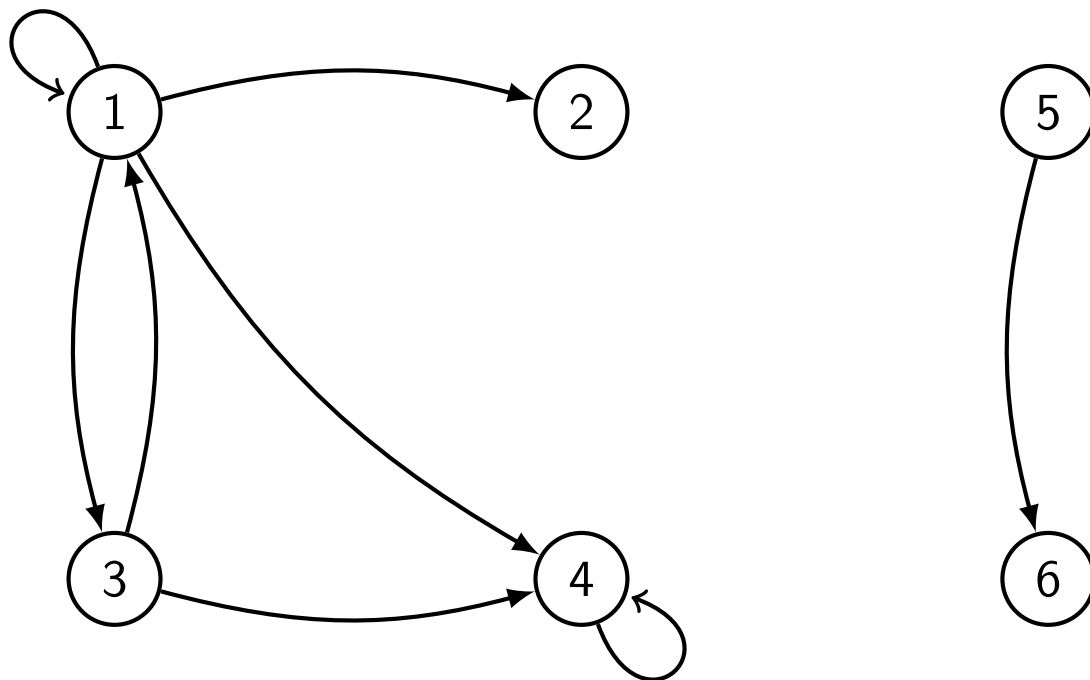
$$\begin{aligned}\vec{\alpha}[a] &= \lim_{n \rightarrow \infty} \left(\sum_{b \in \Omega} \Pr(X_n = a \mid X_0 = b) \cdot \vec{\alpha}[b] \right) \\ &= \sum_{b \in \Omega} \left(\lim_{n \rightarrow \infty} \Pr(X_n = a \mid X_0 = b) \right) \cdot \vec{\alpha}[b] \\ &= \sum_{b \in \Omega} \lambda_a \cdot \vec{\alpha}[b] \\ &= \lambda_a \cdot \sum_{b \in \Omega} \vec{\alpha}[b] \\ &= \lambda_a \cdot 1 = \lambda_a = \vec{\pi}[a]\end{aligned}$$

Concluimos que $\vec{\alpha} = \vec{\pi}$, puesto que $\vec{\alpha}[a] = \vec{\pi}[a]$ para todo $a \in \Omega$



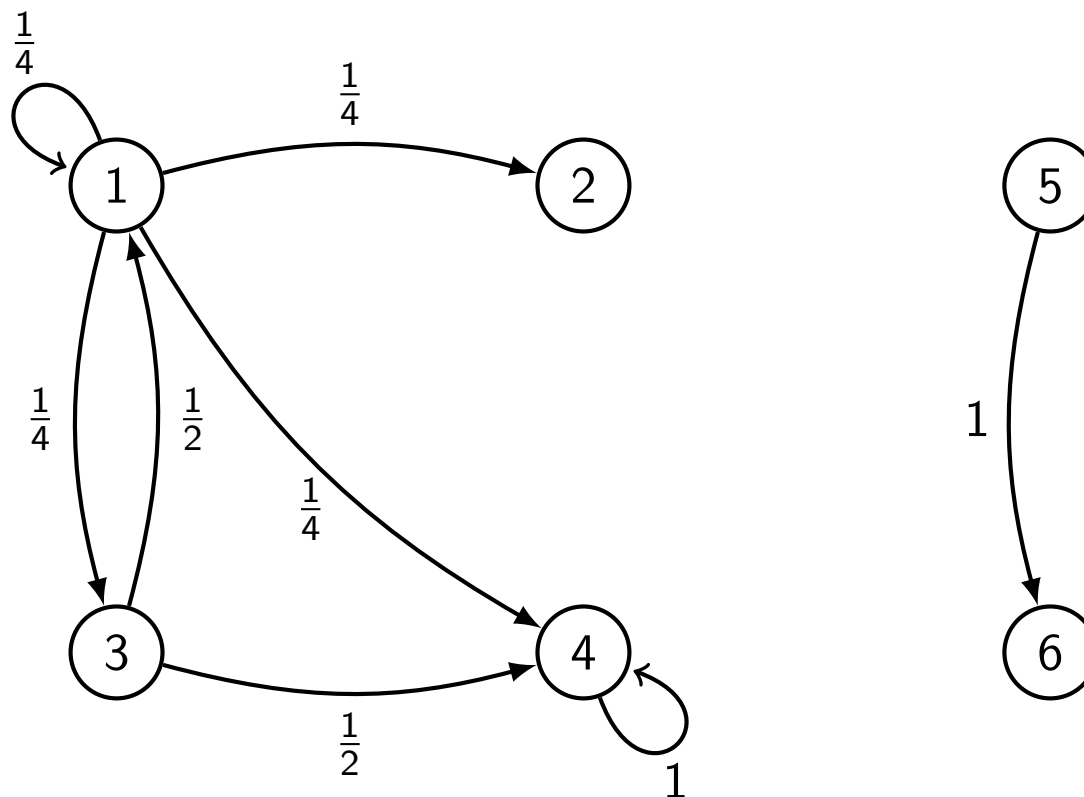
Un ejemplo importante: PageRank

La estructura de links en una pequeña Web:



Un ejemplo importante: PageRank

La estructura de links vista como una caminata aleatoria:



Un ejemplo importante: PageRank

Podemos representar la caminata aleatoria como una matriz P

- $P[i, j]$ es la probabilidad de ir de la página j a la página i

La matriz P de nuestro ejemplo:

$$P = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Un ejemplo importante: PageRank

El ranking de una página i se define como la probabilidad x_i de que el caminante aleatorio llegue a i

- ¿Qué características de la Web influyen en el valor x_i ?

En el ejemplo tenemos que:

$$x_i = \sum_{j=1}^6 P[i, j] \cdot x_j$$

Así, si $\vec{x}$ es un vector tal que $\vec{x}[i] = x_i$ para cada $i \in \{1, \dots, 6\}$, entonces:

$$P\vec{x} = \vec{x}$$

PageRank como una cadena de Markov

El proceso anterior se puede generalizar para la Web real.

- Suponemos que la Web tiene N páginas, por lo que P es una matriz de $N \times N$

Además, definimos $\vec{\pi}$ como un vector tal que $\vec{\pi}[i]$ es la probabilidad de que el caminante aleatorio llegue a la página i , para cada $i \in \{1, \dots, N\}$

PageRank como una cadena de Markov

P puede ser considerada como la matriz de transición de una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ con conjunto de estados $\Omega = \{1, \dots, N\}$

- ▶ El vector $\vec{\pi}$ es entonces una distribución estacionaria de $\{X_t\}_{t \in \mathbb{N}}$

¿Cómo podemos asegurar que $\vec{\pi}$ existe y es único? ¿Cómo podemos calcular $\vec{\pi}$?

¡Irreducibilidad y aperiodicidad son las propiedades que necesitamos!

- ▶ Dado que el conjunto de estados de la cadena de Markov es finito

Irreducibilidad, aperiodicidad y PageRank

Dos preguntas por responder:

- ▶ ¿Podemos asegurar que la cadena de Markov dada por P es irreducible? No
- ▶ ¿Podemos asegurar que la cadena de Markov dada por P es aperiódica? No

¿Cómo solucionamos estos problemas?

Irreducibilidad, aperiodicidad y PageRank

Para solucionar los problemas suponemos que el caminante aleatorio también puede decidir moverse a cualquier página sin considerar la estructura de la Web.

- Esto es controlado por un parámetro α

Sea U una matriz de $N \times N$ tal que $U[i, j] = \frac{1}{N}$, y sea $\alpha \in (0, 1)$

Definimos una matriz W_α de $N \times N$ como:

$$W_\alpha = \alpha \cdot P + (1 - \alpha) \cdot U$$

Irreducibilidad, aperiodicidad y PageRank

Con probabilidad α el caminante utiliza los links para moverse, y con probabilidad $(1 - \alpha)$ se mueve a cualquier página (sin considerar la estructura de la Web).

- ▶ Cuanto más grande es α , más consideramos la estructura de la Web

Tenemos que W_α define una cadena de Markov irreducible y aperiódica.

- ▶ ¿Por qué?

La definición de PageRank

Consideramos un valor fijo de α

- ▶ En el artículo original sobre PageRank se consideraba $\alpha = 0.85$

Definimos entonces $\vec{\pi}$ como la distribución estacionaria de la cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ cuya matriz de transición es W_α

- ▶ Esta distribución existe y es única porque la cadena de Markov es irreducible, aperiódica y tiene un conjunto finito de estados

El cálculo de PageRank

Para cada página i , calculamos $\vec{\pi}[i]$ considerando la igualdad:

$$\vec{\pi}[i] = \lim_{n \rightarrow \infty} \mathbf{Pr}(X_n = i \mid X_0 = 1)$$

Vale decir, la caminata comienza en la página 1 (una página arbitraria), y el valor $\vec{\pi}[i]$ se calcula utilizando un valor de n suficientemente grande.

- Obtenemos una aproximación del valor de $\vec{\pi}[i]$

El cálculo de PageRank

Las ideas anteriores se traducen en la práctica en el siguiente algoritmo **AproxPageRank** para aproximar $\vec{\pi}$

En **AproxPageRank** utilizamos un parámetro $\varepsilon \in (0, 1)$ que indica cuándo debe detenerse el algoritmo.

- ▶ Si la distancia entre los últimos vectores generados es menor que ε el algoritmo se detiene

En **AproxPageRank** la similitud entre dos vectores $\vec{a}$ y $\vec{b}$ es medida a través de la distancia Manhattan:

$$||\vec{a} - \vec{b}||_1 = \sum_{i=1}^N |\vec{a}[i] - \vec{b}[i]|$$

El algoritmo AproxPageRank

AproxPageRank(ε)

Sea $\vec{\pi}_0$ tal que $\vec{\pi}_0[1] = 1$ y $\vec{\pi}_0[i] = 0$ para cada $i \in \{2, \dots, N\}$

Sea $\vec{\pi}_1 = W_\alpha \vec{\pi}_0$

if $\|\vec{\pi}_0 - \vec{\pi}_1\|_1 < \varepsilon$ **then return** $\vec{\pi}_1$

else

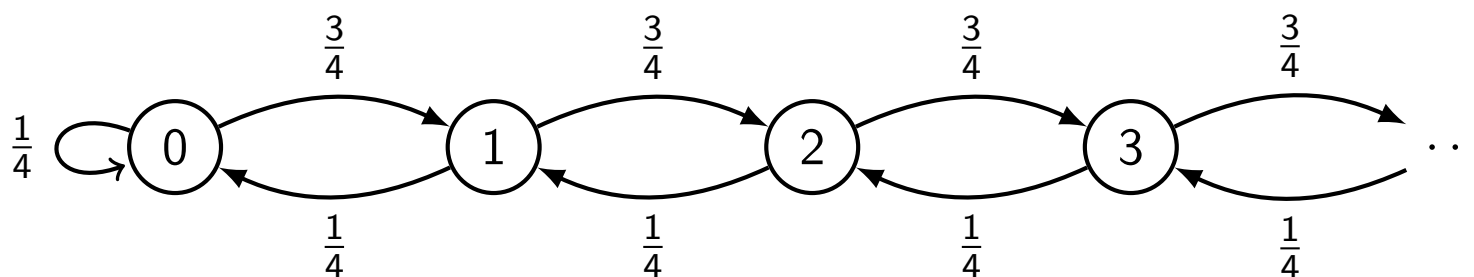
for $n := 2$ **to** ∞ **do**

$\vec{\pi}_n = W_\alpha \vec{\pi}_{n-1}$

if $\|\vec{\pi}_{n-1} - \vec{\pi}_n\|_1 < \varepsilon$ **then return** $\vec{\pi}_n$

Un último comentario: el caso infinito

Considere nuevamente la siguiente cadena de Markov con conjunto de estados $\Omega = \mathbb{N}$:



Demostramos que para esta cadena de Markov se tiene que $\lambda_i = 0$ para todo $i \in \mathbb{N}$

- Queremos definir una condición que nos asegure que los valores λ_i forman una distribución estacionaria para la cadena de Markov

Una tercera propiedad fundamental: recurrencia

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov con conjunto de estados Ω

- Y sea $b \in \Omega$

Definimos T_b como una variable aleatoria que registra el primer instante de tiempo mayor a 0 en que llegamos a b , suponiendo que en el tiempo 0 estábamos en b

- Vale decir, si el valor de T_b es n , con $n \geq 1$, entonces $X_0 = b$, $X_1 \neq b$, ..., $X_{n-1} \neq b$ y $X_n = b$

Este tiempo de retorno a b es llamado *hitting time*

Una tercera propiedad fundamental: recurrencia

Definition

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov con conjunto de estados Ω , y sea $b \in \Omega$

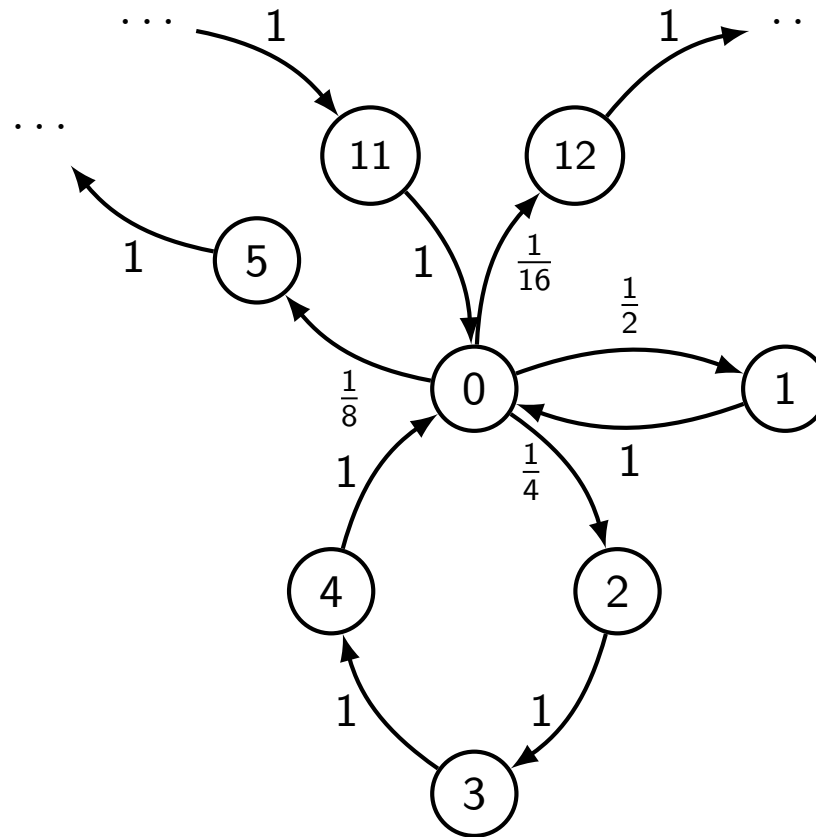
1. b es recurrente si $\sum_{n \geq 1} \Pr(T_b = n) = 1$
2. b es recurrente positivo si b es recurrente y $\mathbf{E}[T_b] \in \mathbb{R}^+$

Ejercicio

De una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ con conjunto de estados Ω y un estado $b \in \Omega$ tal que b es recurrente pero no recurrente positivo.

Una solución para el ejercicio

Considere la siguiente cadena de Markov con conjunto de estados $\Omega = \mathbb{N}$:



Tenemos que 0 es un estado recurrente pero no recurrente positivo.

Existencia y unicidad de la distribución estacionaria: el caso general

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov irreducible, aperiódica y **donde cada estado es recurrente positivo**

- ▶ Y sea $\vec{\pi} \in [0, 1]^{|\Omega|}$ un vector tal que $\vec{\pi}[a] = \lambda_a$ para cada $a \in \Omega$, donde Ω es el conjunto de estados de $\{X_t\}_{t \in \mathbb{N}}$
 - ▶ Nótese que Ω puede ser un conjunto infinito enumerable

Teorema

$\vec{\pi}$ es la única distribución estacionaria para $\{X_t\}_{t \in \mathbb{N}}$