Cadenas de Markov

IIC3810

Marcelo Arenas, Luis Alberto Croquevielle y Thomas Reisenegger

Motivación: una pregunta pendiente del capítulo anterior

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

Motivación: 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:

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

Motivación: 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:

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

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 $\Omega \neq \varnothing$

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 $\Omega \neq \varnothing$

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

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 $\Omega \neq \varnothing$

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

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 $\Omega \neq \varnothing$

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 $\mathbf{Pr}(X_t = \vec{x})$

Una secuencia de variables aleatorias para generar R_{KS}

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

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 $t + 1$

Una secuencia de variables aleatorias para generar R_{KS}

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

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 definir:

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

Una secuencia de variables aleatorias para generar R_{KS}

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

Una secuencia de variables aleatorias para generar R_{KS}

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

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 $i \in \{1, \ldots, n\}$ con distribución uniforme
        $\vec{u} := (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$
        if $\vec{a} \cdot \vec{u} \leq b$ **then return** $\vec{u}$
        else return $\vec{x}$

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

Algunas propiedades de la secuencia

Ejercicios

Sea $t \in \mathbb{N}$

1. Dados $\vec{x}, \vec{y} \in \Omega$, calcule $\mathbf{Pr}(X_{t+1} = \vec{y} \mid X_t = \vec{x})$
    - 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:
$$\mathbf{Pr}(X_{t'} = \vec{y} \mid X_t = \vec{x}) \quad > \quad 0$$

3. Dados $\vec{x}_0, \vec{x}_1, \ldots, \vec{x}_t, \vec{x}_{t+1} \in \Omega$, demuestre que:
$$\mathbf{Pr}(X_{t+1} = \vec{x}_{t+1} \mid X_0 = \vec{x}_0 \wedge X_1 = \vec{x}_1 \wedge \cdots \wedge X_t = \vec{x}_t) \ = \ \mathbf{Pr}(X_{t+1} = \vec{x}_{t+1} \mid X_t = \vec{x}_t)$$

¿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**

¿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?

¿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 $\mathbf{Pr}(X_{t'+1} = \vec{y}) = \mathbf{Pr}(X_{t'} = \vec{y})$?

¿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?

¿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:

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

2. Demuestre que la distribución uniforme es una distribución estacionaria.
   - Vale decir, demuestre que si para $t \in \mathbb{N}$ se tiene que $\mathbf{Pr}(X_t = \vec{x}) = \frac{1}{|\Omega|}$ para cada $\vec{x} \in \Omega$, entonces:

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

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 Ω
> $t = f(|\vec{a}| + |b|)$
> **for** $i := 1$ **to** t **do**
> $\vec{x} :=$ **GenerarSiguiente**$(\vec{x})$
> **return** $\vec{x}$

Donde $|\vec{a}| + |b|$ es el tamaño de la entrada $(\vec{a}, b)$

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 Ω
> $t = f(|\vec{a}| + |b|)$
> **for** $i := 1$ **to** t **do**
> $\vec{x} :=$ **GenerarSiguiente**$(\vec{x})$
> **return** $\vec{x}$

Donde $|\vec{a}| + |b|$ es el tamaño de la entrada $(\vec{a}, b)$

¿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 $\vec{x} \in \Omega$

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

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

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

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 este herramienta esencial para el muestreo de variables aleatorias

Las cadenas de Markov

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

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 $X_t : D_t \to \Omega$ para cada $t \in \mathbb{N}$*
2. *Existe $p : \Omega \times \Omega \to [0,1]$ tal que para cada $t \in \mathbb{N}$ y cada secuencia $a_0, \ldots, a_t, a_{t+1}$ de elementos de Ω:*

$$\mathbf{Pr}(X_{t+1} = a_{t+1} \mid X_t = a_t \wedge \cdots \wedge X_0 = a_0) = \mathbf{Pr}(X_{t+1} = a_{t+1} \mid X_t = a_t) = p(a_{t+1}, a_t)$$

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

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 $a, b \in \Omega$:

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

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 $a, b \in \Omega$:

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

Estos son llamadas cadenas de Markov homogéneas.

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 $a, b \in \Omega$:

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

Estos son llamadas cadenas de Markov homogéneas.

En general, consideramos cadenas de Markov con conjuntos finitos de estados.

Representando una cadena de Markov

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

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

Representando una cadena de Markov

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

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

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

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

Representando una cadena de Markov

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

- Suponemos además que $p : \Omega \times \Omega \to [0, 1]$ 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:

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

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:

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

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

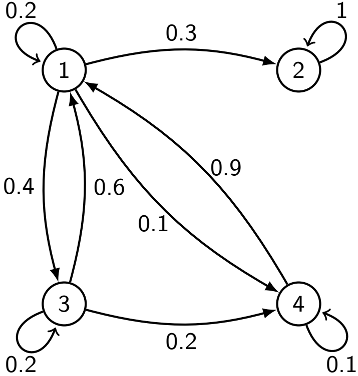
La matriz de transición P de una cadena de Markov con conjunto de estados Ω puede ser vista como un grafo G_P con pesos:

- 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

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

$G_P =$



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 $t \in \mathbb{N}$

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:

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

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

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

La demostración de la existencia del límite

Para cada $k \geq 1$, tenemos que:

$$\begin{array}{rcl}
M_a^{(k \cdot n_0 + n)} - m_a^{(k \cdot n_0 + n)} & = & M_a^{(n_0 + (k-1) \cdot n_0 + n)} - m_a^{(n_0 + (k-1) \cdot n_0 + n)} \\
& \leq & \big(1 - \ell \cdot t\big) \cdot \big(M_a^{((k-1) \cdot n_0 + n)} - m_a^{((k-1) \cdot n_0 + n)}\big) \\
& = & \big(1 - \ell \cdot t\big) \cdot \big(M_a^{(n_0 + (k-2) \cdot n_0 + n)} - m_a^{(n_0 + (k-2) \cdot n_0 + n)}\big) \\
& \leq & \big(1 - \ell \cdot t\big)^2 \cdot \big(M_a^{((k-2) \cdot n_0 + n)} - m_a^{((k-2) \cdot n_0 + n)}\big) \\
& \leq & \cdots \\
& \leq & \big(1 - \ell \cdot t\big)^k \cdot \big(M_a^{(n)} - m_a^{(n)}\big)
\end{array}$$

Notése que la desigualdad también es válida para $k = 0$

La demostración de la existencia del límite

Considerando $n = 0$, obtenemos la sub-sucesión $\left(M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}\right)_{k \in \mathbb{N}}$ de la sucesión $\left(M_a^{(n)} - m_a^{(n)}\right)_{n \in \mathbb{N}}$

Dado que $0 \leq (1 - \ell \cdot t) < 1$ (ya que $t, \ell > 0$), para esta sub-sucesión se tiene que:

$$0 \leq \lim_{k \to \infty} (M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}) \ \leq \lim_{k \to \infty} \left(1 - \ell \cdot t\right)^k \cdot \left(M_c^{(n)} - m_c^{(n)}\right) \ = \ 0$$

Vale decir,

$$\lim_{k \to \infty} (M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}) \quad = \quad 0$$

La demostración de la existencia del límite

Para todo $n \in \mathbb{N}$ se tiene que:

- $m_a^{(n)} \leq m_a^{(n+1)}$
- $M_a^{(n)} \geq M_a^{(n+1)}$
- $m_a^{(n)} \leq M_a^{(n)}$

Por lo tanto, para todo $n \in \mathbb{N}$ se tiene que:

- $M_a^{(n)} - m_a^{(n)} \geq 0$
- $M_a^{(n+1)} - m_a^{(n+1)} \leq M_a^{(n)} - m_a^{(n)}$

La demostración de la existencia del límite

Dado que $\left(M_a^{(n)} - m_a^{(n)}\right)_{n\in\mathbb{N}}$ es una sucesión de números reales monótona decreciente y acotada inferiormente, concluimos que existe $u \in \mathbb{R}$ tal que:

$$\lim_{n\to\infty} \left(M_a^{(n)} - m_a^{(n)}\right) = u$$

La demostración de la existencia del límite

Dado que $\left(M_a^{(n)} - m_a^{(n)}\right)_{n \in \mathbb{N}}$ es una sucesión de números reales monótona decreciente y acotada inferiormente, concluimos que existe $u \in \mathbb{R}$ tal que:

$$\lim_{n \to \infty} \left(M_a^{(n)} - m_a^{(n)}\right) \quad = \quad u$$

Además, para la sub-sucesión $\left(M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}\right)_{k \in \mathbb{N}}$ de $\left(M_a^{(n)} - m_a^{(n)}\right)_{n \in \mathbb{N}}$ se cumple que:

$$\lim_{k \to \infty} (M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}) \quad = \quad 0$$

La demostración de la existencia del límite

Dado que $\left(M_a^{(n)} - m_a^{(n)}\right)_{n \in \mathbb{N}}$ es una sucesión de números reales monótona decreciente y acotada inferiormente, concluimos que existe $u \in \mathbb{R}$ tal que:

$$\lim_{n \to \infty} \left(M_a^{(n)} - m_a^{(n)}\right) \quad = \quad u$$

Además, para la sub-sucesión $\left(M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}\right)_{k \in \mathbb{N}}$ de $\left(M_a^{(n)} - m_a^{(n)}\right)_{n \in \mathbb{N}}$ se cumple que:

$$\lim_{k \to \infty} (M_a^{(k \cdot n_0)} - m_a^{(k \cdot n_0)}) \quad = \quad 0$$

¡Concluimos que $u = 0$!

La demostración de la existencia del límite

Dado que $\left(M_a^{(n)} - m_a^{(n)}\right)_{n\in\mathbb{N}}$ es una sucesión de números reales monótona decreciente y acotada inferiormente, concluimos que existe $u \in \mathbb{R}$ tal que:

$$\lim_{n\to\infty} \left(M_a^{(n)} - m_a^{(n)}\right) \quad = \quad u$$

Además, para la sub-sucesión $\left(M_a^{(k\cdot n_0)} - m_a^{(k\cdot n_0)}\right)_{k\in\mathbb{N}}$ de $\left(M_a^{(n)} - m_a^{(n)}\right)_{n\in\mathbb{N}}$ se cumple que:

$$\lim_{k\to\infty} \left(M_a^{(k\cdot n_0)} - m_a^{(k\cdot n_0)}\right) \quad = \quad 0$$

¡Concluimos que $u = 0$!

- ¿Cómo se demuestra que las dos sucesiones tienen el mismo límite? ¿Por qué es importante aquí que $n_0 \geq 1$? □

Irreducibilidad, aperiodicidad y el caso infinito

Irreducibilidad y aperiodicidad **no** son condiciones suficientes para asegurar que una cadena de Markov con un conjunto infinito (enumerable) de estados tiene una distribución estacionaria única.

- De hecho, bajo estas condiciones ni siquiera se puede asegurar que la cadena tiene una distribución estacionaria

Irreducibilidad, aperiodicidad y el caso infinito

Irreducibilidad y aperiodicidad **no** son condiciones suficientes para asegurar que una cadena de Markov con un conjunto infinito (enumerable) de estados tiene una distribución estacionaria única.

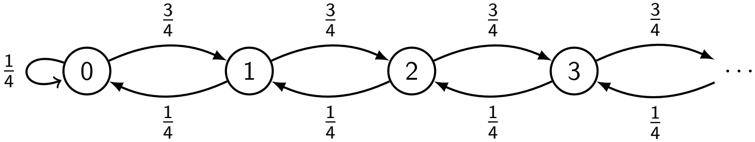
- De hecho, bajo estas condiciones ni siquiera se puede asegurar que la cadena tiene una distribución estacionaria

Ejercicio

Construya una cadena de Markov irreducible, aperiódica, con un conjunto infinito de estados y que no tenga distribución estacionaria.

Una solución al ejercicio

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 ejercicio

Suponga que P es la matriz de transición de la cadena de Markov, y $\vec{\lambda}$ es una distribución estacionaria para ella, vale decir,

$$P\vec{\lambda} \quad = \quad \vec{\lambda}$$

Además, suponga que $\vec{\lambda}[i] = \lambda_i$ para cada $i \in \mathbb{N}$

- Tenemos que $\sum_{i \in \mathbb{N}} \lambda_i = 1$

Una solución al ejercicio

Considerando las definiciones de P y $\vec{\lambda}$ obtenemos:

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

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

Una solución al ejercicio

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

Tenemos que:

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

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

Una solución al ejercicio

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

Una solución al ejercicio

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 $\lambda_0 = 0$

Una solución al ejercicio

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 $\lambda_0 = 0$

Concluimos entonces que $\lambda_i = 0$ para cada $i \in \mathbb{N}$, lo cual contradice la condición $\sum_{i \in \mathbb{N}} \lambda_i = 1$

- No podemos entonces tener una distribución estacionaria para P $\square$

¿Cómo podemos asegurar convergencia en el caso infinito?

Vamos a ver que irreducibilidad y aperiodicidad sí son condiciones útiles en el caso infinito, pero cuando con consideradas junto con una tercera propiedad fundamental.

La noción de estado recurrente

Sea b un estado en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$

La noción de estado recurrente

Sea b un estado en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$

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*

La noción de estado recurrente

Sea b un estado en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$

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 $X_0 = b$, $X_1 \neq b$, $\ldots$, $X_{n-1} \neq b$ y $X_n = b$

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

Definición

b es recurrente si $\sum_{n \geq 1} \mathbf{Pr}(T_b = n) = 1$, y b es transitorio en caso contrario.

Una caracterización de los estado recurrentes

Sea b un estado en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$

Defina:

$$p_b = \sum_{n \geq 1} \mathbf{Pr}(T_b = n)$$

Una caracterización de los estado recurrentes

Sea b un estado en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$

Defina:

$$p_b = \sum_{n \geq 1} \mathbf{Pr}(T_b = n)$$

Ejercicio

Demuestre que $p_b \in [0, 1]$

- Nótese que no estamos asumiendo que b es recurrente
- ¿Es posible tener una cadena de Markov donde $p_b = 0$?

Una caracterización de los estado recurrentes

Suponga que la matriz de transición de $\{X_t\}_{t \in \mathbb{N}}$ es P, y recuerde que G_P es grafo que representa a P

Partiendo desde el estado b, consideramos caminatas sobre G_P que sólo utilizan arcos con probabilidad mayor a cero y tiene largo infinito.

- Podemos realizar estas caminatas incluso si $\{X_t\}_{t \in \mathbb{N}}$ tiene un conjunto finito de estados

Defina una variable aleatoria V_b que cuenta el número total de veces que b es visitado en las caminatas.

- El punto de partida b es considerado como la primera visita a este estado

Una caracterización de los estado recurrentes

Dado un número natural $n \geq 1$, tenemos que:

$$\mathbf{Pr}(V_b = n) \quad = \quad p_b^{n-1} \cdot (1 - p_b)$$

Una caracterización de los estado recurrentes

Dado un número natural $n \geq 1$, tenemos que:

$$\mathbf{Pr}(V_b = n) \quad = \quad p_b^{n-1} \cdot (1 - p_b)$$

Proposición

b es recurrente si y sólo si $\mathbf{E}[V_b] = \infty$

Una caracterización de los estado recurrentes

Dado un número natural $n \geq 1$, tenemos que:

$$\mathbf{Pr}(V_b = n) \quad = \quad p_b^{n-1} \cdot (1 - p_b)$$

Proposición

b es recurrente si y sólo si $\mathbf{E}[V_b] = \infty$

Ejercicio

Demuestre la proposición.

Una caracterización de los estado recurrentes

Dado un número natural $n \geq 1$, tenemos que:

$$\mathbf{Pr}(V_b = n) \quad = \quad p_b^{n-1} \cdot (1 - p_b)$$

Proposición

b es recurrente si y sólo si $\mathbf{E}[V_b] = \infty$

Ejercicio

Demuestre la proposición.

¡Tenemos entonces que b es recurrente si y sólo si b es visitado una cantidad infinita de veces!

Una segunda caracterización de los estado recurrentes

Partiendo desde el estado b, nuevamente consideramos caminatas sobre G_P que sólo utilizan arcos con probabilidad mayor a cero y tiene largo infinito.

Dado $n \in \mathbb{N}$, sea $V_b^{(n)}$ una variable aleatoria tal que $V_b^{(n)} = 1$ si el estado en la caminata es b después de recorrer n arcos en G_P, y $V_b^{(n)} = 0$ en caso contrario.

Tenemos que:

$$V_b = \sum_{n \in \mathbb{N}} V_b^{(n)}$$

Una segunda caracterización de los estado recurrentes

Concluimos que:

$$\mathbf{E}[V_b] \quad = \quad \sum_{n \in \mathbb{N}} \mathbf{E}[V_b^{(n)}]$$

Concluimos que:

$$\mathbf{E}[V_b] \quad = \quad \sum_{n \in \mathbb{N}} \mathbf{E}[V_b^{(n)}]$$

Así, dado que $V_b^{(n)} \sim \mathbf{Ber}(\mathbf{Pr}(X_n = b \mid X_0 = b))$, concluimos que:

$$\mathbf{E}[V_b] \quad = \quad \sum_{n \in \mathbb{N}} \mathbf{Pr}(X_n = b \mid X_0 = b)$$

Una segunda caracterización de los estado recurrentes

Concluimos que:

$$\mathbf{E}[V_b] = \sum_{n \in \mathbb{N}} \mathbf{E}[V_b^{(n)}]$$

Así, dado que $V_b^{(n)} \sim \mathbf{Ber}(\mathbf{Pr}(X_n = b \mid X_0 = b))$, concluimos que:

$$\mathbf{E}[V_b] = \sum_{n \in \mathbb{N}} \mathbf{Pr}(X_n = b \mid X_0 = b)$$

Corolario

b es recurrente si y sólo si

$$\sum_{n \in \mathbb{N}} \mathbf{Pr}(X_n = b \mid X_0 = b) = \infty$$

Una tercera propiedad fundamental: estados recurrentes positivos

Definition

Un estado b en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ es recurrente positivo si b es recurrente y $\mathbf{E}[T_b] \in \mathbb{R}$

Una tercera propiedad fundamental: estados recurrentes positivos

Definition

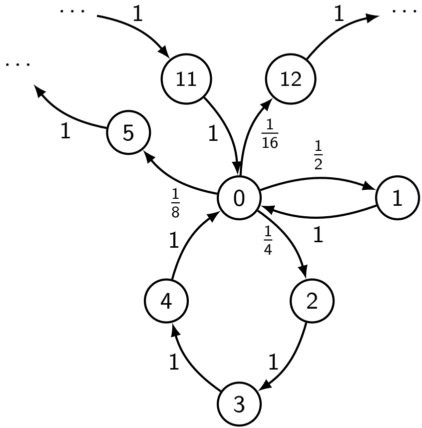
Un estado b en una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ es recurrente positivo si b es recurrente y $\mathbf{E}[T_b] \in \mathbb{R}$

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.

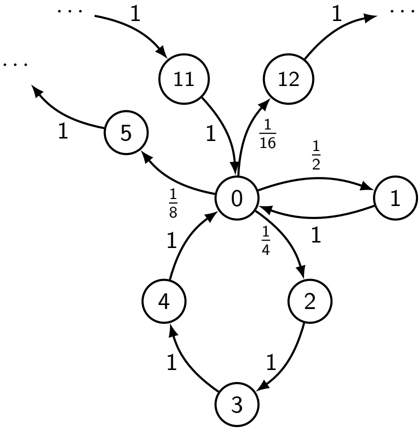
Una solución para el ejercicio

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



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
  - $\lambda_a = \lim_{n \to \infty} \mathbf{Pr}(X_n = a \mid X_0 = b)$, vale decir, λ_a es definido como en el caso finito considerando un estado arbitrario $b \in \Omega$ (todos los puntos de partida dan el mismo resultado)

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
  - $\lambda_a = \lim_{n \to \infty} \mathbf{Pr}(X_n = a \mid X_0 = b)$, vale decir, λ_a es definido como en el caso finito considerando un estado arbitrario $b \in \Omega$ (todos los puntos de partida dan el mismo resultado)

Teorema

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

Un recordatorio de Álgebra Lineal

Estudiar las cadenas de Markov a través de sus matrices de transición nos puede ayudar a enteneder mejor su comportamiento.

Un recordatorio de Álgebra Lineal

Estudiar las cadenas de Markov a través de sus matrices de transición nos puede ayudar a enteneder mejor su comportamiento. Para esto, vamos a recordar unos conceptos de Álgebra Lineal.

Un recordatorio de Álgebra Lineal

Estudiar las cadenas de Markov a través de sus matrices de transición nos puede ayudar a enteneder mejor su comportamiento. Para esto, vamos a recordar unos conceptos de Álgebra Lineal.

Definición

Sea A una matriz cuadrada. Un **vector propio** *de A es un vector (no nulo) $\bar{x}$ tal que $A\bar{x} = \lambda\bar{x}$, donde $\lambda \in \mathbb{R}$. Este λ a su vez es un* **valor propio** *de M.*

Un recordatorio de Álgebra Lineal

Estudiar las cadenas de Markov a través de sus matrices de transición nos puede ayudar a enteneder mejor su comportamiento. Para esto, vamos a recordar unos conceptos de Álgebra Lineal.

Definición

Sea A una matriz cuadrada. Un **vector propio** *de A es un vector (no nulo) $\bar{x}$ tal que $A\bar{x} = \lambda\bar{x}$, donde $\lambda \in \mathbb{R}$. Este λ a su vez es un* **valor propio** *de M.*

Naturalmente, si $\vec{x}$ es vector propio de M, para cualquier valor $a \in \mathbb{R}$ se tiene que $a \cdot \vec{x}$ también lo es.

Un recordatorio de Álgebra Lineal

Definición

El **polinomio característico** *de una matriz cuadrada A es el determinante de $A - \lambda I$, donde I es la matriz identidad.*

Definición

El **polinomio característico** *de una matriz cuadrada A es el determinante de $A - \lambda I$, donde I es la matriz identidad.*

Notemos que $A\vec{x} = \lambda\vec{x}$ equivale a que $(A - \lambda I)\vec{x}$ sea el vector 0 para un vector no nulo $\vec{x}$. Es decir, $A - \lambda I$ es no invertible, y su determinante es 0.

Definición

El **polinomio característico** *de una matriz cuadrada A es el determinante de $A - \lambda I$, donde I es la matriz identidad.*

Notemos que $A\vec{x} = \lambda\vec{x}$ equivale a que $(A - \lambda I)\vec{x}$ sea el vector 0 para un vector no nulo $\vec{x}$. Es decir, $A - \lambda I$ es no invertible, y su determinante es 0.

Por lo tanto, λ es raíz del polinomio característico de A si y sólo si es valor propio.

Un recordatorio de Álgebra Lineal

Corolario

Una matriz A de tamaño $n \times n$ tiene a lo más n valores propios. Además, son los mismos que los de su matriz transpuesta A^T.

Corolario

Una matriz A de tamaño $n \times n$ tiene a lo más n valores propios. Además, son los mismos que los de su matriz transpuesta A^T.

El grado de λ en el determinante de $A - \lambda I$ es a lo más n, así que la primera parte se cumple por el teorema fundamental del álgebra.

Corolario

Una matriz A de tamaño $n \times n$ tiene a lo más n valores propios. Además, son los mismos que los de su matriz transpuesta A^T.

El grado de λ en el determinante de $A - \lambda I$ es a lo más n, así que la primera parte se cumple por el teorema fundamental del álgebra.

Para la segunda parte hay que recordar el determinante es simétrico respecto a filas y columnas, por lo que $det(A) = det(A^T)$. Luego, $det(A - \lambda I) = det(A^T - \lambda I)$, y sus raíces son las mismas.

Valores propios de una cadena de Markov

Teorema

Considere una cadena de Markov **arbitraria** *con una matriz de transición P. Se tiene que 1 es valor propio de P.*

Valores propios de una cadena de Markov

Teorema

Considere una cadena de Markov **arbitraria** *con una matriz de transición P. Se tiene que 1 es valor propio de P.*

Demostración. Tenemos que las columnas de de P suman 1, y por lo tanto las filas de P^T también.

Valores propios de una cadena de Markov

Teorema

Considere una cadena de Markov **arbitraria** *con una matriz de transición P. Se tiene que 1 es valor propio de P.*

Demostración. Tenemos que las columnas de de P suman 1, y por lo tanto las filas de P^T también.

Considere el siguiente vector:

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Se tiene que $\vec{x}$ es un vector propio de P^T con valor propio 1. Luego 1 es valor propio de P.

Valores propios de una cadena de Markov

Si una cadena de Markov tiene una distribución estacionaria $\vec{\pi}$ entonces cumple con $P\vec{\pi} = 1 \cdot \pi$. Es decir, tiene como valor propio 1.

Valores propios de una cadena de Markov

Si una cadena de Markov tiene una distribución estacionaria $\vec{\pi}$ entonces cumple con $P\vec{\pi} = 1 \cdot \pi$. Es decir, tiene como valor propio 1.

Sin embargo, que una matriz tenga valor propio 1 aún no nos dice que su vector propio sea una distribución de probabilidades.

Valores propios de una cadena de Markov

Si una cadena de Markov tiene una distribución estacionaria $\vec{\pi}$ entonces cumple con $P\vec{\pi} = 1 \cdot \pi$. Es decir, tiene como valor propio 1.

Sin embargo, que una matriz tenga valor propio 1 aún no nos dice que su vector propio sea una distribución de probabilidades. Podría tener componentes negativas.

Valores propios de una cadena de Markov

Vamos a usar una herramienta clásica del Álgebra lineal

Valores propios de una cadena de Markov

Vamos a usar una herramienta clásica del Álgebra lineal

Teorema (Perron-Frobenius)

Sea A una matriz finita sin componentes negativas. Existe un vector propio de A cuyas componentes son todas positivas.

Valores propios de una cadena de Markov

Vamos a usar una herramienta clásica del Álgebra lineal

Teorema (Perron-Frobenius)

Sea A una matriz finita sin componentes negativas. Existe un vector propio de A cuyas componentes son todas positivas.

Corolario

Toda cadena de Markov finita tiene una distribución estacionaria.

Demostración. Sea P su matriz de transición, y sea $\vec{x}$ un vector propio positivo con valor propio λ.

Valores propios de una cadena de Markov

Demostración. Sea P su matriz de transición, y sea $\vec{x}$ un vector propio positivo con valor propio λ.

Normalizamos este vector de tal manera que la suma de sus componentes es 1. Sigue siendo vector propio para λ.

Valores propios de una cadena de Markov

Demostración. Sea P su matriz de transición, y sea $\vec{x}$ un vector propio positivo con valor propio λ.

Normalizamos este vector de tal manera que la suma de sus componentes es 1. Sigue siendo vector propio para λ.

Luego tenemos que $P\vec{x} = \lambda\vec{x}$. Como la suma de todas las columnas de P es 1, entonces la suma de los componentes de $\lambda\vec{x}$ también es 1.

Demostración. Sea P su matriz de transición, y sea $\vec{x}$ un vector propio positivo con valor propio λ.

Normalizamos este vector de tal manera que la suma de sus componentes es 1. Sigue siendo vector propio para λ.

Luego tenemos que $P\vec{x} = \lambda\vec{x}$. Como la suma de todas las columnas de P es 1, entonces la suma de los componentes de $\lambda\vec{x}$ también es 1.

Finalmente, como $\sum_i x_i = \sum_i \lambda x_i$, entonces $\lambda = 1$, por lo que $\vec{x}$ es una distribución estacionaria. $\square$