Markov Chain Monte Carlo

Ya vimos cómo los problemas de conteo se pueden reducir a problemas de generación uniforme.

- ¿Pero cómo podemos resolver el problema de generación uniforme? Las cadenas de Markov resultan ser una herramienta muy útil para esto

Markov Chain Monte Carlo

Ya vimos cómo los problemas de conteo se pueden reducir a problemas de generación uniforme.

- ¿Pero cómo podemos resolver el problema de generación uniforme? Las cadenas de Markov resultan ser una herramienta muy útil para esto

Supongamos, más en general, que nos interesa generar elementos de un conjunto Ω según una distribución de probabilidad $\vec{\pi}$

Markov Chain Monte Carlo

Ya vimos cómo los problemas de conteo se pueden reducir a problemas de generación uniforme.

- ¿Pero cómo podemos resolver el problema de generación uniforme? Las cadenas de Markov resultan ser una herramienta muy útil para esto

Supongamos, más en general, que nos interesa generar elementos de un conjunto Ω según una distribución de probabilidad $\vec{\pi}$

- Por ejemplo, Ω podría ser el conjunto de valuaciones que satisfacen una fórmula proposicional, y $\vec{\pi}$ la distribución uniforme sobre estas valuaciones

Markov Chain Monte Carlo

Ya vimos cómo los problemas de conteo se pueden reducir a problemas de generación uniforme.

- ¿Pero cómo podemos resolver el problema de generación uniforme? Las cadenas de Markov resultan ser una herramienta muy útil para esto

Supongamos, más en general, que nos interesa generar elementos de un conjunto Ω según una distribución de probabilidad $\vec{\pi}$

- Por ejemplo, Ω podría ser el conjunto de valuaciones que satisfacen una fórmula proposicional, y $\vec{\pi}$ la distribución uniforme sobre estas valuaciones

¿Cómo podemos hacer esto?

- La clave está en construir una cadena de Markov cuya distribución estacionaria sea $\vec{\pi}$

Markov Chain Monte Carlo

Así, dado un conjunto Ω finito y una distribución de probabilidades $\vec{\pi}$ sobre Ω, queremos definir una cadena de Markov cuyos estados sean los elementos de Ω y cuya distribución estacionaria sea $\vec{\pi}$

Markov Chain Monte Carlo

Así, dado un conjunto Ω finito y una distribución de probabilidades $\vec{\pi}$ sobre Ω, queremos definir una cadena de Markov cuyos estados sean los elementos de Ω y cuya distribución estacionaria sea $\vec{\pi}$

Mostraremos una estrategia general para definir una cadena de Markov con estas características: el algoritmo de Metropolis-Hastings

Metropolis-Hastings: el primer ingrediente

Proposición

Sea $\{X_t\}_{t \in \mathbb{N}}$ *una cadena de Markov con matriz de transición* P. *Si para todo* $a, b \in \Omega$ *se cumple que:*

$$P[b, a] \cdot \vec{\pi}[a] \quad = \quad P[a, b] \cdot \vec{\pi}[b],$$

entonces $\vec{\pi}$ *es una distribución estacionaria para* $\{X_t\}_{t \in \mathbb{N}}$

Una demostración de la proposición

Demostración: Sea $a \in \Omega$. Tenemos que:

$$\begin{aligned}
(P\vec{\pi})[a] &= \sum_{b \in \Omega} P[a,b] \cdot \vec{\pi}[b] \\
&= \sum_{b \in \Omega} P[b,a] \cdot \vec{\pi}[a] \\
&= \vec{\pi}[a] \cdot \sum_{b \in \Omega} P[b,a] \\
&= \vec{\pi}[a] \cdot \sum_{b \in \Omega} \mathbf{Pr}(X_1 = b \mid X_0 = a) \\
&= \vec{\pi}[a] \cdot 1 \;=\; \vec{\pi}[a]
\end{aligned}$$

Vale decir, $P\vec{\pi} = \vec{\pi}$, de lo cual concluimos que $\vec{\pi}$ es una distribución estacionaria de $\{X_t\}_{t \in \mathbb{N}}$ $\square$

Metropolis-Hastings: el primer ingrediente

A la propiedad de que para todo $a, b \in \Omega$ se tenga que:

$$P[b,a] \cdot \vec{\pi}[a] \quad = \quad P[a,b] \cdot \vec{\pi}[b]$$

se le llama **condición de reversibilidad**

La proposición anterior nos indica entonces que para que la cadena de Markov tenga a $\vec{\pi}$ como distribución estacionaria es suficiente que cumpla la condición de reversibilidad.

Metropolis-Hastings: el primer ingrediente

A la propiedad de que para todo $a, b \in \Omega$ se tenga que:

$$P[b, a] \cdot \vec{\pi}[a] \quad = \quad P[a, b] \cdot \vec{\pi}[b]$$

se le llama **condición de reversibilidad**

La proposición anterior nos indica entonces que para que la cadena de Markov tenga a $\vec{\pi}$ como distribución estacionaria es suficiente que cumpla la condición de reversibilidad.

- ¿Cómo definimos una cadena de Markov que cumpla esta condición?

Metropolis-Hastings: la idea del algoritmo

A priori, no existe ninguna garantía de que para todo a, b se cumpla que:

$$P[b, a] \cdot \vec{\pi}[a] \quad = \quad P[a, b] \cdot \vec{\pi}[b]$$

Si no se cumple esta propiedad, entonces existe un par a, b tal que:

$$P[b, a] \cdot \vec{\pi}[a] \quad > \quad P[a, b] \cdot \vec{\pi}[b]$$

¿Qué hacemos en este caso?

Metropolis-Hastings: la idea del algoritmo

Lo que hacemos es definir una nueva matriz de transiciones Q, a partir de P, de la siguiente forma:

$$Q[c, d] \quad = \quad P[c, d] \cdot \alpha[c, d] \qquad \text{para } c, d \in \Omega$$

En esta definición, α es construido de tal forma que Q cumpla la condición de reversibilidad.

Metropolis-Hastings: la idea del algoritmo

Considerando un par a, b tal que:

$$P[b, a] \cdot \vec{\pi}[a] \quad > \quad P[a, b] \cdot \vec{\pi}[b]$$

Una idea natural para lograr que esto sea una igualdad es hacer más pequeño el lado izquierdo.

- El lado derecho no es necesario hacerlo más pequeño

Metropolis-Hastings: la idea del algoritmo

Considerando un par a, b tal que:

$$P[b, a] \cdot \vec{\pi}[a] \quad > \quad P[a, b] \cdot \vec{\pi}[b]$$

Una idea natural para lograr que esto sea una igualdad es hacer más pequeño el lado izquierdo.

- El lado derecho no es necesario hacerlo más pequeño

Por lo tanto, definimos $\alpha[a, b] = 1$ y

$$\alpha[b, a] \quad = \quad \frac{P[a, b] \cdot \vec{\pi}[b]}{P[b, a] \cdot \vec{\pi}[a]}$$

Metropolis-Hastings: la idea del algoritmo

Análogamente, para un par a, b tal que $P[b, a] \cdot \vec{\pi}[a] < P[a, b] \cdot \vec{\pi}[b]$, definimos $\alpha[b, a] = 1$ y

$$\alpha[a, b] \quad = \quad \frac{P[b, a] \cdot \vec{\pi}[a]}{P[a, b] \cdot \vec{\pi}[b]},$$

que es lo mismo que la ecuación en el otro caso sólo que con los roles de a y b intercambiados.

Metropolis-Hastings: la idea del algoritmo

Análogamente, para un par a, b tal que $P[b,a] \cdot \vec{\pi}[a] < P[a,b] \cdot \vec{\pi}[b]$, definimos $\alpha[b,a] = 1$ y

$$\alpha[a,b] \quad = \quad \frac{P[b,a] \cdot \vec{\pi}[a]}{P[a,b] \cdot \vec{\pi}[b]},$$

que es lo mismo que la ecuación en el otro caso sólo que con los roles de a y b intercambiados.

En resumen, α se define como:

$$\alpha[a,b] \quad = \quad \begin{cases} \text{mín}\left\{\dfrac{P[b,a] \cdot \vec{\pi}[a]}{P[a,b] \cdot \vec{\pi}[b]}, 1\right\} & \text{si } P[a,b] \cdot \vec{\pi}[b] > 0 \\ 1 & \text{en otro caso} \end{cases}$$

Metropolis-Hastings: la idea del algoritmo

Por tanto, nos gustaría definir $Q[a, b] = P[a, b] \cdot \alpha[a, b]$ y así tener una matriz de transición Q que cumple la condición de reversibilidad.

- Sin embargo, no podemos asegurar que Q sea la matriz de transición de una cadena de Markov

Metropolis-Hastings: la idea del algoritmo

Por tanto, nos gustaría definir $Q[a,b] = P[a,b] \cdot \alpha[a,b]$ y así tener una matriz de transición Q que cumple la condición de reversibilidad.

- Sin embargo, no podemos asegurar que Q sea la matriz de transición de una cadena de Markov

Para asegurar esto, definimos para cada $b \in \Omega$:

$$r(b) \quad = \quad 1 - \sum_{x \in \Omega} P[x,b] \cdot \alpha[x,b],$$

y agregamos $r(b)$ a la probabilidad de quedarse en el estado b

Metropolis-Hastings: la idea del algoritmo

Por tanto, nos gustaría definir $Q[a, b] = P[a, b] \cdot \alpha[a, b]$ y así tener una matriz de transición Q que cumple la condición de reversibilidad.

- Sin embargo, no podemos asegurar que Q sea la matriz de transición de una cadena de Markov

Para asegurar esto, definimos para cada $b \in \Omega$:

$$r(b) \quad = \quad 1 - \sum_{x \in \Omega} P[x, b] \cdot \alpha[x, b],$$

y agregamos $r(b)$ a la probabilidad de quedarse en el estado b

La definición de Q es entonces:

$$Q[a, b] \quad = \quad \begin{cases} P[a, b] \cdot \alpha[a, b] + r(b) & \text{si } a = b \\ P[a, b] \cdot \alpha[a, b] & \text{si } a \neq b \end{cases}$$

Metropolis-Hastings: la formalización del algoritmo

Sea $\vec{\pi}$ una distribución de probabilidades sobre un dominio finito Ω, y sea P la matriz de transición de una cadena de Markov con conjunto de estados Ω

- Y sea $\alpha : \Omega \times \Omega \to [0, 1]$ definida como en las transparencias anteriores

Metropolis-Hastings: la formalización del algoritmo

Sea $\vec{\pi}$ una distribución de probabilidades sobre un dominio finito Ω, y sea P la matriz de transición de una cadena de Markov con conjunto de estados Ω

- Y sea $\alpha : \Omega \times \Omega \to [0, 1]$ definida como en las transparencias anteriores

Además, sea a_0 un elemento arbitrario de Ω

Metropolis-Hastings: la formalización del algoritmo

Sea $\vec{\pi}$ una distribución de probabilidades sobre un dominio finito Ω, y sea P la matriz de transición de una cadena de Markov con conjunto de estados Ω

- Y sea $\alpha : \Omega \times \Omega \to [0, 1]$ definida como en las transparencias anteriores

Además, sea a_0 un elemento arbitrario de Ω

El siguiente algoritmo simula N transiciones de acuerdo a la cadena de Markov con matriz de transición Q descrita anteriormente.

- Así, la cadena de Markov simulada tiene como distribución estacionaria a la distribución $\vec{\pi}$

Metropolis-Hastings: la formalización del algoritmo

MH($\vec{\pi}$, P, a_0, N)
    for $j := 1$ **to** N **do**
        Generar $b \in \Omega$ con distribución $P[\cdot, a_{j-1}]$
        Generar $u \in [0, 1]$ con distribución uniforme
        if $u \leq \alpha[b, a_{j-1}]$ **then**
            $a_j := b$
        else
            $a_j := a_{j-1}$
    return $\{a_1, a_2, \ldots, a_N\}$

Metropolis-Hastings: la formalización del algoritmo

¿Cómo interpretamos la salida del algoritmo?

- Vamos a demostrar que para cada $j \in \{1, \ldots, N\}$, se tiene que a_j es generado según la distribución de probabilidades $Q^j \vec{x}_0$, donde $\vec{x}_0$ es un vector tal que $\vec{x}_0[a_0] = 1$ y $\vec{x}_0[b] = 0$ para todo $b \neq a_0$

Metropolis-Hastings: la formalización del algoritmo

¿Cómo interpretamos la salida del algoritmo?

- Vamos a demostrar que para cada $j \in \{1, \ldots, N\}$, se tiene que a_j es generado según la distribución de probabilidades $Q^j \vec{x}_0$, donde $\vec{x}_0$ es un vector tal que $\vec{x}_0[a_0] = 1$ y $\vec{x}_0[b] = 0$ para todo $b \neq a_0$

En este contexto, Q es la matriz que describimos antes, con distribución estacionaria $\vec{\pi}$

- Así, el algoritmo toma un vector inicial $\vec{x}_0$ y realiza N transiciones según la matriz de transición Q

Markov Chain Monte Carlo: algunos comentarios

El algoritmo de Metropolis-Hastings nos entrega una estrategia general para muestrear de acuerdo a una distribución de interés $\vec{\pi}$

- Si N es suficientemente grande, entonces a_N será generado de acuerdo a una distribución similar a la estacionaria

Markov Chain Monte Carlo: algunos comentarios

El algoritmo de Metropolis-Hastings nos entrega una estrategia general para muestrear de acuerdo a una distribución de interés $\vec{\pi}$

- Si N es suficientemente grande, entonces a_N será generado de acuerdo a una distribución similar a la estacionaria

Sin embargo, el algoritmo no entrega a priori ningún criterio respecto a qué se entiende por suficientemente grande.

- ¿Cuál debe ser el valor de N para que la cadena de Markov se acerque lo suficiente a la distribución estacionaria?

Markov Chain Monte Carlo: algunos comentarios

Por ejemplo, para aplicar el algoritmo a problemas de conteo necesitamos un generador casi uniforme.

- Eso significa que $\vec{\pi}$ es la distribución uniforme, y que queremos estar muy cerca de la distribución estacionaria
- A la vez, no queremos que N sea demasiado grande ya que queremos obtener un algoritmo de tiempo polinomial

Markov Chain Monte Carlo: algunos comentarios

Otro tema importante es el de la convergencia de la cadena de Markov a la distribución $\vec{\pi}$

- El algoritmo de Metropolis-Hastings nos asegura que $\vec{\pi}$ es distribución estacionaria, pero no nos asegura que sea la única
- Aún más, el algoritmo no nos asegura que la cadena de Markov converja a $\vec{\pi}$ independientemente del vector $\vec{x}_0$ inicial

Markov Chain Monte Carlo: algunos comentarios

Otro tema importante es el de la convergencia de la cadena de Markov a la distribución $\vec{\pi}$

- El algoritmo de Metropolis-Hastings nos asegura que $\vec{\pi}$ es distribución estacionaria, pero no nos asegura que sea la única
- Aún más, el algoritmo no nos asegura que la cadena de Markov converja a $\vec{\pi}$ independientemente del vector $\vec{x}_0$ inicial

Para conseguir estas garantías se debe exigir más condiciones a la cadena de Markov caracterizada por Q

- En particular, es suficiente pedir que la cadena de Markov sea irreducible y aperiódica

Markov Chain Monte Carlo: algunos comentarios

Cómo mencionamos antes, en los problemas de conteo consideramos $\vec{\pi}$ como la distribución uniforme.

- Entonces, dado que $\vec{\pi}[a] = \vec{\pi}[b]$ para todo a, b, la condición de reversibilidad se reduce a que $P[a, b] = P[b, a]$

Markov Chain Monte Carlo: algunos comentarios

Cómo mencionamos antes, en los problemas de conteo consideramos $\vec{\pi}$ como la distribución uniforme.

- Entonces, dado que $\vec{\pi}[a] = \vec{\pi}[b]$ para todo a, b, la condición de reversibilidad se reduce a que $P[a, b] = P[b, a]$

Por lo tanto, no siempre será necesario recurrir al algoritmo de Metropolis-Hastings en este contexto.

- Nos basta con definir una cadena de Markov cuya matriz de transición sea simétrica, y que además sea irreducible y aperiódica
- Esto puede lograrse en muchos casos. En otros, sin embargo, es necesario recurrir al algoritmo de Metropolis-Hastings

Metropolis-Hastings: la correctitud del algoritmo

Proposición

Sea Ω un conjunto finito, π una distribución sobre Ω y P la matriz de transición de una cadena de Markov con conjunto de estados Ω. Además, sea Q una matriz tal que:

$$Q[a, b] \quad = \quad \begin{cases} P[a, b] \cdot \alpha[a, b] + r(b) & \text{si } a = b \\ P[a, b] \cdot \alpha[a, b] & \text{si } a \neq b \end{cases}$$

para todo $a, b \in \Omega$, donde $r(b) = 1 - \sum_{x \in \Omega} P[x, b] \cdot \alpha[x, b]$ y

$$\alpha[a, b] \quad = \quad \begin{cases} \min\left\{\dfrac{P[b, a] \cdot \vec{\pi}[a]}{P[a, b] \cdot \vec{\pi}[b]}, 1\right\} & \text{si } P[a, b] \cdot \vec{\pi}[b] > 0 \\ 1 & \text{en otro caso} \end{cases}$$

Entonces Q es la matriz de transición de una cadena de Markov y $Q[c, d] \cdot \vec{\pi}[c] = Q[c, d] \cdot \vec{\pi}[d]$ para todo $c, d \in \Omega$

Metropolis-Hastings: la correctitud del algoritmo

Ejercicio

Demuestre la proposición anterior.

Metropolis-Hastings: la correctitud del algoritmo

Ahora, solo falta demostrar que el algoritmo de Metropolis-Hastings efectúa las transiciones de acuerdo a las probabilidades especificadas por Q

Metropolis-Hastings: la correctitud del algoritmo

Ahora, solo falta demostrar que el algoritmo de Metropolis-Hastings efectúa las transiciones de acuerdo a las probabilidades especificadas por Q

En el siguiente teorema, consideramos toda la notación dada en la proposición anterior y en la descripción del algoritmo de Metropolis-Hastings.

Teorema

Para todo $c \in \Omega$ y $j \in \{1, \ldots, N\}$ se tiene que:

$$\mathbf{Pr}(a_j = c) \quad = \quad (Q^j \vec{x}_0)[c]$$

Metropolis-Hastings: la correctitud del algoritmo

Ahora, solo falta demostrar que el algoritmo de Metropolis-Hastings efectúa las transiciones de acuerdo a las probabilidades especificadas por Q

En el siguiente teorema, consideramos toda la notación dada en la proposición anterior y en la descripción del algoritmo de Metropolis-Hastings.

Teorema

Para todo $c \in \Omega$ y $j \in \{1, \ldots, N\}$ se tiene que:

$$\mathbf{Pr}(a_j = c) \quad = \quad (Q^j \vec{x}_0)[c]$$

Demostraremos el teorema sin requerir que $\vec{x}_0$ sea un vector canónico.

- $\vec{x}_0$ representa entonces la distribución de probabilidades de acuerdo a la cual se eligió a_0

La demostración del teorema

Demostración: haremos la demostración por inducción, partiendo por el caso base $j = 1$

- La demostración del caso inductivo, como se verá después, es completamente análoga a la del caso base

Sea $c \in \Omega$. Tenemos entonces que:

$$\mathbf{Pr}(a_1 = c) \ = \ \mathbf{Pr}(a_1 = c \wedge u \leq \alpha[b, a_0]) + \mathbf{Pr}(a_1 = c \wedge u > \alpha[b, a_0])$$

Denotamos el primer término como p_1 y el segundo como p_2

- Estudiamos a continuación cada uno de estos términos

La demostración del teorema

$$
\begin{aligned}
p_1 &= \mathbf{Pr}(a_1 = c \wedge u \leq \alpha[b, a_0]) \\
&= \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(a_1 = c \wedge u \leq \alpha[b, a_0] \wedge b = y \wedge a_0 = x) \\
&= \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(a_1 = c \mid u \leq \alpha[b, a_0] \wedge b = y \wedge a_0 = x) \cdot \\
& \qquad\qquad\qquad\qquad \mathbf{Pr}(u \leq \alpha[b, a_0] \wedge b = y \wedge a_0 = x) \\
&= \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(y = c) \cdot \mathbf{Pr}(u \leq \alpha[b, a_0] \mid b = y \wedge a_0 = x) \cdot \\
& \qquad\qquad\qquad\qquad \mathbf{Pr}(b = y \wedge a_0 = x) \\
&= \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(y = c) \cdot \alpha[y, x] \cdot \mathbf{Pr}(b = y \mid a_0 = x) \cdot \mathbf{Pr}(a_0 = x)
\end{aligned}
$$

La demostración del teorema

Por lo tanto, tenemos que:

$$
\begin{aligned}
p_1 &= \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(y = c) \cdot \alpha[y, x] \cdot P[y, x] \cdot \mathbf{Pr}(a_0 = x) \\
&= \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(y = c) \cdot \alpha[y, x] \cdot P[y, x] \cdot \vec{x}_0[x] \\
&= \sum_{x \in \Omega} \alpha[c, x] \cdot P[c, x] \cdot \vec{x}_0[x]
\end{aligned}
$$

La demostración del teorema

Por otra parte, tenemos que:

$$
\begin{aligned}
p_2 = {} & \mathbf{Pr}(a_1 = c \wedge u > \alpha[b, a_0]) \\
= {} & \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(a_1 = c \wedge u > \alpha[b, a_0] \wedge b = y \wedge a_0 = x) \\
= {} & \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(a_1 = c \mid u > \alpha[b, a_0] \wedge b = y \wedge a_0 = x) \cdot \\
& \qquad \mathbf{Pr}(u > \alpha[b, a_0] \wedge b = y \wedge a_0 = x) \\
= {} & \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(x = c) \cdot \mathbf{Pr}(u > \alpha[b, a_0] \mid b = y \wedge a_0 = x) \cdot \\
& \qquad \mathbf{Pr}(b = y \wedge a_0 = x) \\
= {} & \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(x = c) \cdot (1 - \alpha[y, x]) \cdot \mathbf{Pr}(b = y \mid a_0 = x) \cdot \mathbf{Pr}(a_0 = x)
\end{aligned}
$$

La demostración del teorema

Por lo tanto, tenemos que:

$$
\begin{aligned}
p_2 = \ & \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(x = c) \cdot (1 - \alpha[y, x]) \cdot P[y, x] \cdot \mathbf{Pr}(a_0 = x) \\
= \ & \sum_{x \in \Omega} \sum_{y \in \Omega} \mathbf{Pr}(x = c) \cdot (1 - \alpha[y, x]) \cdot P[y, x] \cdot \vec{x}_0[x] \\
= \ & \sum_{y \in \Omega} (1 - \alpha[y, c]) \cdot P[y, c] \cdot \vec{x}_0[c] \\
= \ & \vec{x}_0[c] \cdot \sum_{y \in \Omega} (1 - \alpha[y, c]) \cdot P[y, c] \\
= \ & \vec{x}_0[c] \cdot \left(\sum_{y \in \Omega} P[y, c] - \sum_{y \in \Omega} \alpha[y, c] \cdot P[y, c] \right) \\
= \ & \vec{x}_0[c] \cdot \left(1 - \sum_{y \in \Omega} \alpha[y, c] \cdot P[y, c] \right) \\
= \ & \vec{x}_0[c] \cdot r(c)
\end{aligned}
$$

La demostración del teorema

Tenemos entonces que:

$$\begin{aligned}
\mathbf{Pr}(a_1 = c) &= p_1 + p_2 \\
&= \left(\sum_{x \in \Omega} \alpha[c,x] \cdot P[c,x] \cdot \vec{x}_0[x] \right) + \vec{x}_0[c] \cdot r(c) \\
&= (\alpha[c,c] \cdot P[c,c] \cdot \vec{x}_0[c] + r(c) \cdot \vec{x}_0[c]) + \\
&\qquad \sum_{x \in \Omega \,:\, x \neq c} \alpha[c,x] \cdot P[c,x] \cdot \vec{x}_0[x] \\
&= Q[c,c] \cdot \vec{x}_0[c] + \sum_{x \in \Omega \,:\, x \neq c} Q[c,x] \cdot \vec{x}_0[x] \\
&= \sum_{x \in \Omega} Q[c,x] \cdot \vec{x}_0[x] \\
&= (Q\vec{x}_0)[c]
\end{aligned}$$

La demostración del teorema

Con esto queda demostrado el caso base.

La demostración del teorema

Con esto queda demostrado el caso base.

La demostración del caso inductivo es totalmente análoga.

- Si suponemos que la propiedad es cierta para a_k y queremos probar para a_{k+1}, la demostración es igual a la que acabamos de hacer, solo que en lugar del vector $\vec{x}_0$ consideramos el vector $(Q^k \vec{x}_0)$
  - Por hipótesis de inducción sabemos que a_k distribuye de acuerdo al vector $(Q^k \vec{x}_0)$

La demostración del teorema

Con esto queda demostrado el caso base.

La demostración del caso inductivo es totalmente análoga.

- Si suponemos que la propiedad es cierta para a_k y queremos probar para a_{k+1}, la demostración es igual a la que acabamos de hacer, solo que en lugar del vector $\vec{x}_0$ consideramos el vector $(Q^k \vec{x}_0)$
  - Por hipótesis de inducción sabemos que a_k distribuye de acuerdo al vector $(Q^k \vec{x}_0)$

Así, queda demostrado el teorema. $\square$