

Los métodos Markov chain Monte Carlo (MCMC)

IIC3810

Marcelo Arenas, Luis Alberto Croquevielle y Thomas
Reisenegger

Markov Chain Monte Carlo

Ya vimos como 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 son una herramienta útil para esto

En general, las cadenas de Markov son una herramienta útil para generar elementos de un conjunto Ω según una distribución de probabilidad $\vec{\pi}$

- ▶ Para esto, la distribución estacionaria de la cadena de Markov debe ser $\vec{\pi}$

Ejemplo

Sea φ una fórmula proposicional satisfacible, y sea $\{x_1, \dots, x_n\}$ el conjunto de variables mencionadas en φ

Definimos Ω como el conjunto de valuaciones $\sigma : \{x_1, \dots, x_n\} \rightarrow \{0, 1\}$ y $\vec{\pi}$ como una distribución de probabilidad tal que:

► Si $\sigma_1(\varphi) = \sigma_2(\varphi) = 1$, entonces $\vec{\pi}[\sigma_1] = \vec{\pi}[\sigma_2]$

►
$$\sum_{\sigma : \sigma(\varphi)=0} \vec{\pi}[\sigma] \leq \frac{1}{2}$$

Una cadena de Markov con distribución estacionaria $\vec{\pi}$ puede ser utilizada para generar valuaciones que satisfacen φ

► ¿Cómo se hace esto? ¿Cuántos pasos toma generar una valuación?

Markov Chain Monte Carlo

Un método Markov Chain Monte Carlo utiliza una cadena de Markov para generar elementos de acuerdo con la distribución estacionaria de la cadena

- ▶ En el capítulo anterior estudiamos uno de estos métodos

En este capítulo vamos a ver respuestas para dos preguntas fundamentales para los métodos Markov Chain Monte Carlo

- ▶ ¿Es posible crear una cadena de Markov con distribución estacionaria π a partir de una cadena de Markov que no tenga a π como distribución estacionaria?
- ▶ ¿Cuántos pasos deben ser ejecutados para generar elementos de acuerdo a una distribución de probabilidad cercana a la distribución estacionaria de una cadena de Markov?
 - ▶ ¿Cómo se define que dos distribuciones sean cercanas?

Una respuesta a la primera pregunta

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 $\vec{\pi}$

- ▶ Queremos definir esta cadena de Markov a partir de una cadena de Markov dada (que no tiene a $\vec{\pi}$ como distribución estacionaria)

Mostraremos una estrategia general para esta tarea: **el algoritmo de Metropolis-Hastings**

Metropolis-Hastings: el primer ingrediente

Proposición

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

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

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

Una demostración de la proposición

Demostración. Dado $a \in \Omega$, tenemos que:

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

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



Metropolis-Hastings: el primer ingrediente

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

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

se le llama **condición de reversibilidad**

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

Reversibilidad no es una condición necesaria

Ejercicio

Construya una cadena de Markov irreducible, aperiódica, con matriz de transición P no simétrica y que tenga a la distribución uniforme $\vec{\pi}$ como distribución estacionaria.

- Nótese que esta cadena no satisface la propiedad de reversibilidad para $\vec{\pi}$

Una solución para el ejercicio: $P = \begin{pmatrix} 0.25 & 0.25 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.75 & 0 \end{pmatrix}$

Metropolis-Hastings: la idea del algoritmo

¿Cómo definimos una cadena de Markov que cumpla la condición de reversibilidad para una distribución $\vec{\pi}$?

No existe ninguna garantía de que para todo a, b se cumpla que:

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

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

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

¿Qué hacemos en este caso?

Metropolis-Hastings: la idea del algoritmo

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

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

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

Metropolis-Hastings: la idea del algoritmo

Considere un par a, b tal que:

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

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

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

Metropolis-Hastings: la idea del algoritmo

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

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

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:

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

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.

- Puesto que no podemos asegurar que $\sum_{a \in \Omega} Q[a, b] = 1$ para todo $b \in \Omega$

De hecho sólo podemos asegurar que $\sum_{a \in \Omega} Q[a, b] \leq 1$ para todo $b \in \Omega$

- ¿Por qué?

Metropolis-Hastings: la idea del algoritmo

Para resolver el problema, definimos para cada $b \in \Omega$:

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

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

De esta forma, la definición de Q es:

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

Ejercicio

Demuestre que Q es la matriz de transición de una cadena de Markov.

Metropolis-Hastings: la formalización del algoritmo

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

- ▶ Y sea $\alpha : \Omega \times \Omega \rightarrow [0, 1]$ 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.

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  $\{a_1, a_2, \dots, a_N\}$ 
```

Metropolis-Hastings: la formalización del algoritmo

¿Cómo interpretamos la salida del algoritmo?

- Vamos a demostrar que para cada $j \in \{1, \dots, N\}$, 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

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

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

- ▶ Más adelante vamos a analizar en detalle cómo conseguir estas garantías manipulando la matriz P

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, $\vec{\pi}$ 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:

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

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

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

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

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

Metropolis-Hastings: la correctitud del algoritmo

Teorema

Para todo $c \in \Omega$ y $j \in \{0, 1, \dots, N\}$ se tiene que:

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

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

Haremos la demostración por inducción, partiendo por el caso base $j = 0$

La demostración del caso base es trivial, dado que $Q^0 = I$ y, por hipótesis, a_0 se elige de acuerdo a la distribución de probabilidades representada por $\vec{x}_0$, o sea, para todo $c \in \Omega$ se tiene:

$$\Pr(a_0 = c) = \vec{x}_0[c] = (Q^0 \vec{x}_0)[c]$$

La demostración del teorema

Ahora, suponemos que el resultado es cierto para $j = k$, y consideramos $j = k + 1$

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

$$\Pr(a_{k+1} = c) = \Pr(a_{k+1} = c \wedge u \leq \alpha[b, a_k]) + \Pr(a_{k+1} = c \wedge u > \alpha[b, a_k])$$

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

$$\begin{aligned} p_1 &= \Pr(a_{k+1} = c \wedge u \leq \alpha[b, a_k]) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(a_{k+1} = c \wedge u \leq \alpha[b, a_k] \wedge b = y \wedge a_k = x) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(a_{k+1} = c \mid u \leq \alpha[b, a_k] \wedge b = y \wedge a_k = x) \cdot \\ &\quad \Pr(u \leq \alpha[b, a_k] \wedge b = y \wedge a_k = x) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(y = c) \cdot \Pr(u \leq \alpha[b, a_k] \mid b = y \wedge a_k = x) \cdot \\ &\quad \Pr(b = y \wedge a_k = x) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(y = c) \cdot \alpha[y, x] \cdot \Pr(b = y \mid a_k = x) \cdot \Pr(a_k = x) \end{aligned}$$

La demostración del teorema

Por lo tanto, tenemos por hipótesis de inducción que:

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

La demostración del teorema

Por otra parte, tenemos que:

$$\begin{aligned} p_2 &= \Pr(a_{k+1} = c \wedge u > \alpha[b, a_k]) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(a_{k+1} = c \wedge u > \alpha[b, a_k] \wedge b = y \wedge a_k = x) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(a_{k+1} = c \mid u > \alpha[b, a_k] \wedge b = y \wedge a_k = x) \cdot \\ &\quad \Pr(u > \alpha[b, a_k] \wedge b = y \wedge a_k = x) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(x = c) \cdot \Pr(u > \alpha[b, a_k] \mid b = y \wedge a_k = x) \cdot \\ &\quad \Pr(b = y \wedge a_k = x) \\ &= \sum_{x \in \Omega} \sum_{y \in \Omega} \Pr(x = c) \cdot (1 - \alpha[y, x]) \cdot \Pr(b = y \mid a_k = x) \cdot \Pr(a_k = x) \end{aligned}$$

La demostración del teorema

Por lo tanto, tenemos por hipótesis de inducción que:

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

La demostración del teorema

Tenemos entonces que:

$$\begin{aligned}\Pr(a_{k+1} = c) &= p_1 + p_2 \\&= \left(\sum_{x \in \Omega} \alpha[c, x] \cdot P[c, x] \cdot (Q^k \vec{x}_0)[x] \right) + (Q^k \vec{x}_0)[c] \cdot r(c) \\&= (\alpha[c, c] \cdot P[c, c] \cdot (Q^k \vec{x}_0)[c] + r(c) \cdot (Q^k \vec{x}_0)[c]) + \\&\quad \sum_{x \in \Omega : x \neq c} \alpha[c, x] \cdot P[c, x] \cdot (Q^k \vec{x}_0)[x] \\&= Q[c, c] \cdot (Q^k \vec{x}_0)[c] + \sum_{x \in \Omega : x \neq c} Q[c, x] \cdot (Q^k \vec{x}_0)[x] \\&= \sum_{x \in \Omega} Q[c, x] \cdot (Q^k \vec{x}_0)[x] \\&= (Q^{k+1} \vec{x}_0)[c]\end{aligned}$$

La demostración del teorema

Con esto queda demostrado el caso inductivo.

Así, queda demostrado el teorema.



$\vec{\pi}$ como la única distribución estacionaria del algoritmo de Metropolis-Hastings

¿Cómo podemos asegurar que $\vec{\pi}$ es la única distribución estacionaria de la cadena de Markov con matriz de transición Q ?

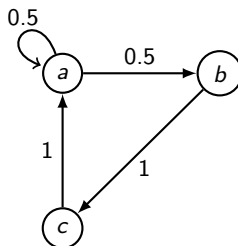
- ▶ ¿Cómo podemos asegurar que la cadena de Markov converge a $\vec{\pi}$ independientemente del vector $\vec{x}_0$ inicial?

Para tener las condiciones anteriores la cadena de Markov debe ser irreducible y aperiódica.

- ▶ ¿Para lograr esto basta con que la cadena de Markov con matriz de transición P sea irreducible y aperiódica?

No basta partir de una cadena irreducible y aperiódica

Considere la siguiente cadena de Markov:



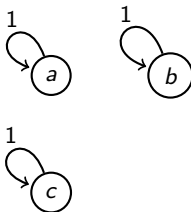
Esta cadena de Markov es irreducible y aperiódica, y su única distribución estacionaria es:

$$\begin{pmatrix} 0.5 \\ 0.25 \\ 0.25 \end{pmatrix}$$

No basta partir de una cadena irreducible y aperiódica

Queremos utilizar el algoritmo de Metropolis-Hastings para obtener a partir de P la distribución uniforme $\bar{\pi}[a] = \bar{\pi}[b] = \bar{\pi}[c] = \frac{1}{3}$ como distribución estacionaria de una cadena de Markov.

En este caso tenemos que $P[a, b] = P[b, c] = P[c, a] = 0$ y $\alpha[b, a] = \alpha[c, b] = \alpha[a, c] = 0$, por lo que el algoritmo genera la siguiente cadena de Markov:



No basta partir de una cadena irreducible y aperiódica

Vale decir, tenemos que: $Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

La cadena de Markov generada por el algoritmo de Metropolis-Hastings no es irreducible ni tiene una única distribución estacionaria.

- ▶ Y además converge a distintas distribuciones dependiendo de la distribución de partida

¿Cómo podemos solucionar estos problemas?

Una solución simple y general

Suponga que tenemos una cadena de Markov con conjunto de estados Ω y matriz de transición P

En el algoritmo de Metropolis-Hastings suponemos que es posible hacer muestreo de acuerdo con P

- ▶ Vale decir, dado $a \in \Omega$, suponemos que es posible generar de manera aleatoria $b \in \Omega$ de acuerdo a la distribución probabilidad $P[\cdot, a]$
- ▶ Más aun, suponemos que b puede ser generado de manera eficiente

Para solucionar los problemas mencionados en las transparencias anteriores vamos a modificar P

- ▶ Lo importante es que debe ser posible hacer muestreo de acuerdo con la matriz modificada de manera eficiente

Una solución simple y general

Sea $\theta \in (0, 1)$ y U una matriz de $|\Omega| \times |\Omega|$ tal que

$$U[a, b] = \frac{1}{|\Omega|}$$

Vale decir, U representa una cadena de Markov en la cual es posible moverse de un estado a cualquier otro con distribución uniforme.

Considerando P , θ y U defina:

$$P_\theta = \theta \cdot P + (1 - \theta) \cdot U$$

Una solución simple y general

Tenemos que:

- ▶ P_θ es irreducible y aperiodica
- ▶ Es posible hacer muestreo de manera eficiente de acuerdo a P_θ si es posible hacer muestreo de manera eficiente de acuerdo a P
- ▶ La matriz Q_θ generada por el algoritmo de Metropolis-Hastings es irreducible y aperiódica
 - ▶ Por lo tanto, utilizando P_θ solucionamos los problemas mencionados en las transparencias anteriores

Ejercicio

Demuestre las propiedades mencionadas arriba.

Una respuesta a la segunda pregunta: velocidad de convergencia

¿Cómo podemos asegurar que una cadena de Markov convergerá a la distribución estacionaria en un tiempo razonable?

Para responder esta pregunta, primero necesitamos definir una métrica que indique la distancia que hay entre la distribución estacionaria y la distribución de probabilidad de la cadena en un tiempo $t \geq 0$

Distancia de variación

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

Definición

Dado $x \in \Omega$, la distancia de variación en tiempo $t \geq 0$ con respecto a un estado inicial x se define como:

$$\Delta_x(t) = \frac{1}{2} \sum_{y \in \Omega} |P^t[y, x] - \vec{\pi}[y]|$$

Mixing time de una cadena de Markov

Definición

Dada una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ y $\varepsilon \in (0, 1)$, el mixing time de $\{X_t\}_{t \in \mathbb{N}}$ a partir de un estado inicial x se define como:

$$\tau_x(\varepsilon) = \min\{t \in \mathbb{N} \mid \forall t' \geq t : \Delta_x(t') \leq \varepsilon\}$$

¿Cómo podemos acotar el mixing time de una cadena de Markov?

La técnica de los caminos canónicos

Desde ahora en adelante, asumiremos una notación estándar para los elementos de una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$:

- ▶ Ω es su conjunto de estados
- ▶ P es su matriz de transición
- ▶ $\vec{\pi}$ es su distribución estacionaria

Suponiendo que $\{X_t\}_{t \in \mathbb{N}}$ es reversible, para cada $x, y \in \Omega$ definimos

$$Q(\{x, y\}) = P[x, y] \cdot \vec{\pi}[y] = P[y, x] \cdot \vec{\pi}[x]$$

Intuitivamente, $\{x, y\}$ representa a un arco en un grafo no dirigido.

La técnica de los caminos canónicos

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov reversible, irreducible y aperiódica

Asociamos a $\{X_t\}_{t \in \mathbb{N}}$ un grafo $H_{P, \pi} = (\Omega, E_{P, \pi})$ no dirigido, donde:

$$E_{P, \pi} = \{\{x, y\} \in 2^\Omega \mid Q(\{x, y\}) > 0\}$$

Además, dados $x, y \in \Omega$, un camino canónico de x a y se define como un camino $\gamma_{x,y} = e_1, \dots, e_r$ desde x a y en $H_{P, \pi}$

- ▶ Suponemos que $r \geq 1$, y denotamos al largo r del camino como $|\gamma_{x,y}|$
- ▶ Utilizamos la notación $e \in \gamma_{x,y}$ para indicar que e es un arco (no dirigido) en el camino $\gamma_{x,y}$

La técnica de los caminos canónicos

Suponga que por cada par $x, y \in \Omega$ escogemos un camino canónico $\gamma_{x,y}$, y definimos Γ como el conjunto de estos caminos.

Definición

La máxima carga de un arco se define como

$$\rho(\Gamma) = \max_{e \in E_P, \vec{\pi}} \left(\frac{1}{Q(e)} \sum_{\gamma_{x,y} \in \Gamma : e \in \gamma_{x,y}} \vec{\pi}[x] \cdot \vec{\pi}[y] \cdot |\gamma_{x,y}| \right)$$

Intuitivamente, queremos que no haya arcos con mucha carga para no formar *cuellos de botella*.

La técnica de los caminos canónicos: un resultado fundamental

Teorema

Sea $\{X_t\}_{t \in \mathbb{N}}$ una cadena de Markov reversible, irreducible y aperiódica, tal que $P[x, x] \geq \frac{1}{2}$ para todo $x \in \Omega$. Además, sea Γ un conjunto de caminos canónicos para cada par de estados en Ω .

Para cada $x \in \Omega$ y $\varepsilon \in (0, 1)$, se tiene que:

$$\tau_x(\varepsilon) \leq \rho(\Gamma) \cdot \left(\ln \frac{1}{\vec{\pi}[x]} + \ln \frac{1}{\varepsilon} \right)$$

Un primer ejemplo: generación aleatoria de strings binarios

Consideremos el problema de generar con distribución uniforme strings binarios de largo $n \geq 1$ a través de una cadena de Markov.

- ▶ Usaremos el conjunto de estados $\Omega = \{0, 1\}^n$

Para movernos entre estados invertiremos a lo más un bit del string.

- ▶ Con probabilidad $\frac{1}{2}$ no nos movemos de estado
- ▶ Si nos movemos, se invierte uno de los n bits con distribución uniforme

Un primer ejemplo: generación aleatoria de strings binarios

Sea $\vec{\pi} : \Omega \rightarrow \mathbb{R}$ la distribución uniforme sobre los estados.

Suponiendo que $d : \Omega \times \Omega \rightarrow \mathbb{N}$ es la distancia de Hamming entre dos strings binarios, tenemos que:

$$P[x, y] = \begin{cases} \frac{1}{2} & \text{si } d(x, y) = 0 \\ \frac{1}{2 \cdot n} & \text{si } d(x, y) = 1 \\ 0 & \text{si } d(x, y) \geq 2 \end{cases}$$

De esto se concluye que $\vec{\pi}$ es la (única) distribución estacionaria de la cadena de Markov.

► ¿Por qué?

Un primer ejemplo: generación aleatoria de strings binarios

La cadena de Markov definida satisface las condiciones del teorema.

- ▶ La cadena es reversible, irreducible, aperiódica y $P[x, x] \geq \frac{1}{2}$ para cada $x \in \Omega$

Podemos usar entonces la técnica de los caminos canónicos para acotar el mixing time.

- ▶ La clave es escoger un conjunto Γ de caminos canónicos tal que $\rho(\Gamma)$ sea polinomial en n , que en este caso es el tamaño de la entrada

Utilizamos como primer ejemplo a la generación aleatoria de strings binarios porque es simple escoger un conjunto de caminos Γ tal que $\rho(\Gamma)$ es polinomial en n

Un primer ejemplo: generación aleatoria de strings binarios

Dado $x = x_1 \cdots x_n$ e $y = y_1 \cdots y_n$, consideramos el camino canónico $\gamma_{x,y} = e_1 \cdots e_n$, donde para $i \in \{1, \dots, n\}$:

$$e_i = \{y_1 \cdots y_{i-1} \textcolor{red}{x}_i x_{i+1} \cdots x_n, y_1 \cdots y_{i-1} \textcolor{red}{y}_i x_{i+1} \cdots x_n\}$$

Vale decir, el camino de x a y corresponde a cambiar uno a uno de izquierda a derecha los bits de x por los de y

- ▶ Si los strings tienen un bit en común, utilizamos un loop
- ▶ Nótese que $|\gamma_{xy}| = n$

Acotando el número de caminos

Necesitamos obtener una cota superior para la cantidad de caminos que pasan por cada arco.

- Consideramos un arco e arbitrario

Para algún $i \in \{1, \dots, n\}$, tenemos que:

$$e = \{y_1 \cdots y_{i-1} x_i x_{i+1} \cdots x_n, y_1 \cdots y_{i-1} y_i x_{i+1} \cdots x_n\}$$

Acotando el número de caminos

Dados $x, y \in \Omega$, si $e \in \gamma_{x,y}$, entonces:

- ▶ Los bits $x_i, \dots, x_n$ de x están fijos, y tenemos 2^{i-1} opciones para los otros bits de x
- ▶ Los bits $y_1, \dots, y_i$ de y están fijos, y tenemos 2^{n-i} opciones para los otros bits de y

Concluimos que la cantidad total de caminos que pueden contener a e es

$$2^{i-1} \cdot 2^{n-i} = 2^{n-1}$$

Acotando superiormente $\rho(\Gamma)$

Recordemos que la carga máxima de un arco es

$$\rho(\Gamma) = \max_{e \in E_{P, \vec{\pi}}} \left(\frac{1}{Q(e)} \sum_{\gamma_{x,y} \in \Gamma : e \in \gamma_{x,y}} \vec{\pi}[x] \cdot \vec{\pi}[y] \cdot |\gamma_{x,y}| \right)$$

Tenemos que:

- ▶ $\forall x \in \Omega : \vec{\pi}[x] = \frac{1}{2^n}$
- ▶ $\forall e \in E_{P, \vec{\pi}}$ tal que $e = \{x, y\} : Q(e) = P[x, y] \cdot \vec{\pi}[y] \geq \frac{1}{2 \cdot n} \cdot \frac{1}{2^n}$
- ▶ $\forall e \in E_{P, \vec{\pi}} : |\{\gamma_{x,y} \in \Gamma \mid e \in \gamma_{x,y}\}| = 2^{n-1}$
- ▶ $\forall x, y \in \Omega : |\gamma_{x,y}| = n$

Acotando superiormente $\rho(\Gamma)$

Desarrollando la expresión obtenemos:

$$\begin{aligned}\rho(\Gamma) &= \max_{e \in E_P, \vec{\pi}} \left(\frac{1}{Q(e)} \sum_{\gamma_{x,y} \in \Gamma : e \in \gamma_{x,y}} \vec{\pi}[x] \cdot \vec{\pi}[y] \cdot |\gamma_{x,y}| \right) \\ &\leq \max_{e \in E_P, \vec{\pi}} \left(2 \cdot n \cdot 2^n \sum_{\gamma_{x,y} \in \Gamma : e \in \gamma_{x,y}} \frac{1}{2^n} \cdot \frac{1}{2^n} \cdot n \right) \\ &= \max_{e \in E_P, \vec{\pi}} \left(2 \cdot n \cdot 2^n \cdot \frac{1}{2^n} \cdot \frac{1}{2^n} \cdot n \cdot 2^{n-1} \right) \\ &= \max_{e \in E_P, \vec{\pi}} n^2 \\ &= n^2\end{aligned}$$

Acotando el mixing time en la generación de strings binarios

Recordemos que para $x \in \Omega$ y $\varepsilon \in (0, 1)$:

$$\tau_x(\varepsilon) \leq \rho(\Gamma) \cdot \left(\ln \frac{1}{\bar{\pi}[x]} + \ln \frac{1}{\varepsilon} \right)$$

Por lo tanto, tenemos que:

$$\begin{aligned} \tau_x(\varepsilon) &\leq n^2 \cdot \left(\ln \frac{1}{\bar{\pi}[x]} + \ln \frac{1}{\varepsilon} \right) = \\ & n^2 \cdot \left(\ln 2^n + \ln \frac{1}{\varepsilon} \right) = n^2 \cdot \left(n \cdot \ln 2 + \ln \frac{1}{\varepsilon} \right) \end{aligned}$$

Concluimos que la cadena converge rápido desde cualquier estado inicial.

Un FPAUG para strings binarios

Ejercicio

Utilizando el resultado anterior construya un FPAUG para el problema de generar string binarios con distribución uniforme.

- Su algoritmo debe realizar a lo más la siguiente cantidad de pasos en la cadena de Markov usada:

$$\left\lceil n^2 \cdot \left(n \cdot \ln 2 + \ln \frac{2}{\epsilon} \right) \right\rceil$$

Un segundo ejemplo: permutaciones de una lista

Dado $n \geq 2$, consideramos el problema de generar permutaciones aleatorias de una lista con n elementos y sin repeticiones.

Para resolver este problema consideramos una cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ con conjunto de estados Ω tal que:

$$\Omega = \{L \mid L \text{ es una lista de largo } n \text{ con elementos } 1, \dots, n\}$$

Un segundo ejemplo: permutaciones de una lista

La dinámica de $\{X_t\}_{t \in \mathbb{N}}$ es definida de la siguiente forma.

- ▶ Dada una lista L :
 - ▶ Con probabilidad $\frac{1}{2}$ no se realiza cambios a L
 - ▶ Con probabilidad $\frac{1}{2}$ se realiza cambios a L . En este caso se escoge con distribución uniforme un par (i, j) tal que $i, j \in \{1, \dots, n\}$ e $i \neq j$, y se intercambia $L[i]$ por $L[j]$

Permutaciones de una lista: distribución estacionaria

La matriz de transición P de la cadena de Markov $\{X_t\}_{t \in \mathbb{N}}$ satisface lo siguiente para cada $L_1, L_2 \in \Omega$:

$$P[L_1, L_2] = \begin{cases} \frac{1}{2} & \text{si } L_1 = L_2 \\ \frac{1}{n \cdot (n-1)} & \text{si } L_1 \text{ y } L_2 \text{ difieren exactamente en dos posiciones} \\ 0 & \text{en otro caso} \end{cases}$$

Ejercicio

Sea $\bar{\pi} : \Omega \rightarrow \mathbb{R}$ la distribución uniforme sobre Ω :

$$\bar{\pi}[L] = \frac{1}{n!} \quad \text{para cada } L \in \Omega$$

Demuestre que $\bar{\pi}$ es la única distribución estacionaria de $\{X_t\}_{t \in \mathbb{N}}$, a la cual converge la cadena independientemente del estado inicial.

Permutaciones de una lista: caminos canónicos

Dados $L_1, L_2 \in \Omega$, definimos un camino canónico γ_{L_1, L_2} desde L_1 a L_2 de la siguiente forma.

- ▶ Suponiendo que $L_1 = [a_1, \dots, a_n]$ y $L_2 = [b_1, \dots, b_n]$, los arcos de γ_{L_1, L_2} van colocando a cada elemento b_i en la posición i que le corresponde en L_2 , desde $i = 1$ hasta $i = n$
- ▶ Si un elemento b_i estaba en la posición i en su turno, entonces no se realiza ninguna acción

Tenemos que $|\gamma_{L_1, L_2}| \leq n$

Definimos Γ como $\{\gamma_{L_1, L_2} \mid L_1, L_2 \in \Omega\}$

Permutaciones de una lista: caminos canónicos

Ejemplo

Si $L_1 = [3, 1, 4, 2]$ y $L_2 = [4, 2, 1, 3]$, el camino canónico de L_1 a L_2 se construye realizando las siguientes acciones:

1. Colocar el elemento 4 en la posición 1, lo que equivale a permutar las posiciones 1 y 3 en la lista L_1 generando $L_3 = [4, 1, 3, 2]$
2. Colocar el elemento 2 en la posición 2, lo que equivale a permutar las posiciones 2 y 4 en la lista L_3 generando $L_4 = [4, 2, 3, 1]$
3. Colocar el elemento 1 en la posición 3, lo que equivale a permutar las posiciones 3 y 4 en la lista L_4 generando $L_2 = [4, 2, 1, 3]$

Tenemos entonces que $\gamma_{L_1, L_2} = \{L_1, L_3\}, \{L_3, L_4\}, \{L_4, L_2\}$

Permutaciones de una lista: caminos por arco

No hay una forma sencilla de contar la cantidad de caminos que pasan por un arco dado.

- Vamos a acotar superiormente este número

Considere un arco $\{L_1, L_2\}$ del grafo $H_{P, \pi}$ tal que $L_1 \neq L_2$, y suponga que:

$$k = \min_{i \in \{1, \dots, n\}} L_1[i] \neq L_2[i]$$

Vamos a definir una función inyectiva $f_{\{L_1, L_2\}} : \{\gamma_{L_3, L_4} \mid \{L_1, L_2\} \in \gamma_{L_3, L_4}\} \rightarrow \Omega$

- Concluimos que $|\{\gamma_{L_3, L_4} \mid \{L_1, L_2\} \in \gamma_{L_3, L_4}\}| \leq |\Omega| = n!$

Permutaciones de una lista: caminos por arco

Suponga que $L_1 = [a_1, \dots, a_n]$, y sean $L_3, L_4 \in \Omega$ tales que $\{L_1, L_2\} \in \gamma_{L_3, L_4}$

- Nótese que $L_3 \neq L_4$

Nos referimos a la posición de un elemento en las permutaciones L_3 , L_4 y L_1 como *inicial*, *final* y *actual*, respectivamente.

Definimos $f_{\{L_1, L_2\}}(\gamma_{L_3, L_4})$ como una lista $L \in \Omega$ tal que:

- Los elementos $a_1, \dots, a_{k-1}$ son colocados en L en sus posiciones iniciales en L_3
- Los $n - (k - 1)$ elementos restantes de L_1 son colocados en L en el orden final en que aparecen en L_4

Permutaciones de una lista: caminos por arco

Podemos comprobar que $f_{\{L_1, L_2\}}$ es inyectiva obteniendo L_3 y L_4 a partir de $f_{\{L_1, L_2\}}(\gamma_{L_3, L_4})$

- Nótese que γ_{L_3, L_4} está determinado por L_3 y L_4

Podemos recuperar L_4 de la siguiente forma:

- Las posiciones finales de $a_1, \dots, a_{k-1}$ son las mismas que sus posiciones actuales
- El orden final de los elementos restantes se puede obtener directamente de $f_{\{L_1, L_2\}}(\gamma_{L_3, L_4})$

Permutaciones de una lista: caminos por arco

Podemos recuperar L_3 notando que:

- ▶ Las posiciones iniciales de $a_1, \dots, a_{k-1}$ se obtienen directamente de $f_{\{L_1, L_2\}}(\gamma_{L_3, L_4})$
- ▶ Las transiciones realizadas antes de $\{L_1, L_2\}$ en el camino γ_{L_3, L_4} corresponden a poner a un elemento de $a_1, \dots, a_{k-1}$ en su posición final
- ▶ Podemos deducir las posiciones iniciales de los $n - (k - 1)$ elementos restantes a partir de las transiciones que fueron realizadas antes de $\{L_1, L_2\}$ en el camino γ_{L_3, L_4}

Permutaciones de una lista: carga máxima

Recordemos que la carga máxima de un arco es

$$\rho(\Gamma) = \max_{e \in E_{P, \vec{\pi}}} \left(\frac{1}{Q(e)} \sum_{\gamma_{L_1, L_2} \in \Gamma : e \in \gamma_{L_1, L_2}} \vec{\pi}[L_1] \cdot \vec{\pi}[L_2] \cdot |\gamma_{L_1, L_2}| \right)$$

Tenemos que:

- ▶ $\forall L \in \Omega : \vec{\pi}[L] = \frac{1}{n!}$
- ▶ $\forall e \in E_{P, \vec{\pi}}$ tal que $e = \{L_1, L_2\} : Q(e) = P[L_1, L_2] \cdot \vec{\pi}[L_2] \geq \frac{1}{n \cdot (n-1)} \cdot \frac{1}{n!}$
- ▶ $\forall e \in E_{P, \vec{\pi}}$ tal que $e = \{L_1, L_2\}$ y $L_1 \neq L_2 : |\{\gamma_{L_3, L_4} \in \Gamma \mid e \in \gamma_{L_3, L_4}\}| \leq n!$
- ▶ $\forall e \in E_{P, \vec{\pi}}$ tal que $e = \{L, L\} : |\{\gamma_{L_3, L_4} \in \Gamma \mid e \in \gamma_{L_3, L_4}\}| = 1 \leq n!$
- ▶ $\forall L_1, L_2 \in \Omega : |\gamma_{L_1, L_2}| \leq n$

Permutaciones de una lista: carga máxima

Desarrollando la expresión obtenemos:

$$\begin{aligned}\rho(\Gamma) &= \max_{e \in E_P, \vec{\pi}} \left(\frac{1}{Q(e)} \sum_{\gamma_{L_1, L_2} \in \Gamma : e \in \gamma_{L_1, L_2}} \vec{\pi}[L_1] \cdot \vec{\pi}[L_2] \cdot |\gamma_{L_1, L_2}| \right) \\ &\leq \max_{e \in E_P, \vec{\pi}} \left(n \cdot (n-1) \cdot n! \sum_{\gamma_{L_1, L_2} \in \Gamma : e \in \gamma_{L_1, L_2}} \frac{1}{n!} \cdot \frac{1}{n!} \cdot n \right) \\ &\leq \max_{e \in E_P, \vec{\pi}} \left(n \cdot (n-1) \cdot n! \cdot \frac{1}{n!} \cdot \frac{1}{n!} \cdot n \cdot n! \right) \\ &= \max_{e \in E_P, \vec{\pi}} n^2 \cdot (n-1) \\ &= n^2 \cdot (n-1)\end{aligned}$$

Permutaciones de una lista: mixing time

Recordemos que para $x \in \Omega$ y $\varepsilon \in (0, 1)$:

$$\tau_x(\varepsilon) \leq \rho(\Gamma) \cdot \left(\ln \frac{1}{\bar{\pi}[x]} + \ln \frac{1}{\varepsilon} \right)$$

Por lo tanto, tenemos que:

$$\begin{aligned} \tau_x(\varepsilon) &\leq n^2 \cdot (n-1) \cdot \left(\ln \frac{1}{\bar{\pi}[x]} + \ln \frac{1}{\varepsilon} \right) = \\ &n^2 \cdot (n-1) \cdot \left(\ln n! + \ln \frac{1}{\varepsilon} \right) \leq n^2 \cdot (n-1) \cdot \left(n \cdot \ln n + \ln \frac{1}{\varepsilon} \right) \end{aligned}$$

Concluimos que la cadena converge rápido desde cualquier estado inicial.

Un FPAUG para las permutaciones de una lista

Ejercicio

Utilizando el resultado anterior construya un FPAUG para el problema de generar permutaciones aleatorias de una lista con n elementos y sin repeticiones.

- Su algoritmo debe realizar a lo más la siguiente cantidad de pasos en la cadena de Markov usada:

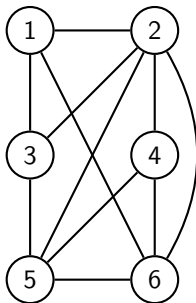
$$\left\lceil n^2 \cdot (n - 1) \cdot \left(n \cdot \ln n + \ln \frac{2}{\varepsilon} \right) \right\rceil$$

Un tercer ejemplo: matchings de un grafo

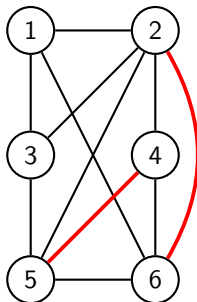
Dado un grafo $G = (V, E)$ no dirigido, un matching en G es un conjunto de arcos $M \subseteq E$ que satisface las siguientes condiciones:

- ▶ Ningún arco en M es un loop $\{a, a\} = \{a\}$
- ▶ Para cada par de arcos $\{a_1, b_1\}, \{a_2, b_2\} \in M$ tales que $\{a_1, b_1\} \neq \{a_2, b_2\}$, se tiene que $\{a_1, b_1\} \cap \{a_2, b_2\} = \emptyset$

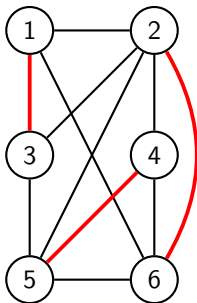
Matchings de un grafo: ejemplos



Matchings de un grafo: ejemplos



Matchings de un grafo: ejemplos



Contando el número de matchings de un grafo

Dado un grafo G no dirigido, sea $\#MATCHING(G)$ una función que retorna el número de matchings de G

Teorema

$\#MATCHING$ es $\#P$ -completo y $L_{\#MATCHING} \in PTIME$

Entonces puede existir un FPRAS para $\#MATCHING$

- ▶ ¿Qué herramientas podemos usar para construir este FPRAS?

Construyendo un FPRAS para #MATCHING

Considere la siguiente relación:

$$R_{\text{MATCHING}} = \{(G, M) \mid G \text{ es un grafo no dirigido y} \\ M \text{ es un matching de } G\}$$

Proposición

R_{MATCHING} es auto-reducible.

- ¿Cómo puede ser codificado M para que se cumpla esta propiedad?

Por el teorema de Jerrum, Valiant & Vazirani, necesitamos entonces un FPAUG para R_{MATCHING} para obtener un FPRAS para #MATCHING

- Usamos una cadena de Markov para construir este FPAUG

Una cadena de Markov para los matchings de un grafo

Sea $G = (V, E)$ un grafo no dirigido

Consideramos una cadena de Markov con conjunto de estados:

$$\Omega = \{M \subseteq E \mid M \text{ es un matching de } G\}$$

Una cadena de Markov para los matchings de un grafo

Dado $M \in \Omega$, el procedimiento para ejecutar una transición es el siguiente:

1. Con probabilidad $\frac{1}{2}$ nos quedamos en M
2. En caso contrario, escogemos un arco $e \in M$, con $e = \{u, v\}$ y $u \neq v$, de manera aleatoria y con distribución uniforme, y hacemos lo siguiente:
 - 2.1 si $e \in M$, nos movemos a $M \setminus \{e\}$
 - 2.2 si $e \cap e' = \emptyset$ para cada $e' \in M$, nos movemos a $M \cup \{e\}$
 - 2.3 si $|e \cap e'| = 1$ para algún $e' \in E$, y $e \cap e'' = \emptyset$ para todo $e'' \in M \setminus \{e'\}$, entonces nos movemos a $(M \setminus \{e'\}) \cup \{e\}$
 - 2.4 en otro caso, nos quedamos en M

Llamamos a los cambios 2.1, 2.2 y 2.3 transiciones tipo 1, 2 y 3, respectivamente

Una cadena de Markov para los matchings de un grafo

Ejercicios

Considere la cadena de Markov definida en las transparencias anteriores.

1. Demuestre que la cadena es irreducible y aperiódica
2. Demuestre que la cadena es reversible y su distribución estacionaria es la distribución uniforme sobre Ω

Matchings de un grafo: caminos canónicos

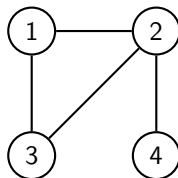
Sea G un grafo no dirigido.

Suponemos dado un orden lineal para el conjunto de los caminos y ciclos simples en G , y suponemos que cada uno de ellos tiene un vértice inicial.

- ▶ Si el camino no es un ciclo, el vértice inicial tiene que ser un extremo
- ▶ Si el camino es un ciclo, el vértice inicial es un nodo arbitrario

Ejemplo de un orden sobre los caminos y ciclos simples

Considere el siguiente grafo:



Ejemplo de un orden sobre los caminos y ciclos simples

Un orden lineal para los caminos y ciclos simples:

- ▶ $Q_1 = (1, 2)$
- ▶ $Q_2 = (1, 3, 2)$
- ▶ $Q_3 = (1, 3)$
- ▶ $Q_4 = (1, 2, 3)$
- ▶ $Q_5 = (1, 2, 4)$
- ▶ $Q_6 = (1, 3, 2, 4)$
- ▶ $Q_7 = (1, 2, 3, 1)$
- ▶ $Q_8 = (2, 3)$
- ▶ $Q_9 = (2, 1, 3)$
- ▶ $Q_{10} = (2, 4)$
- ▶ $Q_{11} = (3, 1, 2, 4)$
- ▶ $Q_{12} = (3, 2, 4)$

Ejemplo de un orden sobre los caminos y ciclos simples

Y un vértice inicial para cada camino:

- ▶ $Q_1 = (1, 2)$
- ▶ $Q_2 = (1, 3, 2)$
- ▶ $Q_3 = (1, 3)$
- ▶ $Q_4 = (1, 2, 3)$
- ▶ $Q_5 = (1, 2, 4)$
- ▶ $Q_6 = (1, 3, 2, 4)$
- ▶ $Q_7 = (1, 2, 3, 1)$ (ciclo)
- ▶ $Q_8 = (2, 3)$
- ▶ $Q_9 = (2, 1, 3)$
- ▶ $Q_{10} = (2, 4)$
- ▶ $Q_{11} = (3, 1, 2, 4)$
- ▶ $Q_{12} = (3, 2, 4)$

Definición de los caminos canónicos

Sean $I, F \in \Omega$

Podemos describir la diferencia simétrica $(I \setminus F) \cup (F \setminus I)$ como una secuencia $Q_1, \dots, Q_r$ de caminos y ciclos simples disjuntos que respetan el orden lineal.

- ▶ Construimos el camino canónico de I a F secuencialmente a partir de estos caminos y ciclos simples
 - ▶ Tenemos que tratar a los caminos y ciclos simples de distinta forma

Caso 1: Q_i no es un ciclo

Sea $Q_i = (v_0, v_1, \dots, v_\ell)$, con $v_i \in \Omega$ para todo $i \in \{0, \dots, \ell\}$ y v_0 el vértice inicial

En primer lugar, suponemos que $\{v_0, v_1\} \in F$

- ▶ Realizamos una serie de transiciones de tipo 3, reemplazando $\{v_{2j+1}, v_{2j+2}\}$ por $\{v_{2j}, v_{2j+1}\}$ para $j = 0, 1, \dots$ tal que $2j + 2 \leq \ell$
 - ▶ En cada paso quitamos un arco $\{v_{2j+1}, v_{2j+2}\} \in I$ y agregamos el arco anterior $\{v_{2j}, v_{2j+1}\} \in F$
- ▶ Finalizamos con una transición de tipo 2 si ℓ es impar
 - ▶ Si ℓ es impar, falta considerar el arco $\{v_{\ell-1}, v_\ell\}$, que agregamos a través de una transición de tipo 2

Caso 1: Q_i no es un ciclo

Sea $Q_i = (v_0, v_1, \dots, v_\ell)$, con $v_i \in \Omega$ para todo $i \in \{0, \dots, \ell\}$ y v_0 el vértice inicial

En segundo lugar, suponemos que $\{v_0, v_1\} \in I$

- ▶ Partimos con una transición de tipo 1, removiendo $\{v_0, v_1\}$
- ▶ Procedemos con el camino reducido $(v_1, \dots, v_\ell)$
 - ▶ Nótese que si el camino $(v_1, \dots, v_\ell)$ tiene a v_ℓ como vértice inicial y $\{v_{\ell-1}, v_\ell\} \in I$ entraremos nuevamente en este caso

Caso 2: Q_i es un ciclo

Sea $Q_i = (v_0, v_1, \dots, v_{2\ell+1}, v_0)$, con $v_i \in \Omega$ para todo $i \in \{0, \dots, 2\ell+1\}$, v_0 el vértice inicial y $\{v_{2j}, v_{2j+1}\} \in I$ para todo $j \in \{0, \dots, \ell\}$

- ▶ El ciclo debe tener un número par de aristas. ¿Por qué?
- ▶ Nótese que el orden del ciclo no es relevante, por lo que si $\{v_{2j}, v_{2j+1}\} \in F$ para todo $j \in \{0, \dots, \ell\}$, podemos considerar el ciclo en el orden invertido y cumplirá con las condiciones pedidas

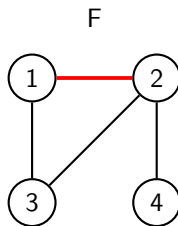
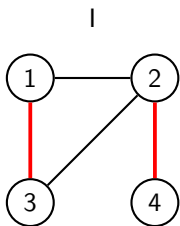
Caso 2: Q_i es un ciclo

Sea $Q_i = (v_0, v_1, \dots, v_{2\ell+1}, v_0)$, con $v_i \in \Omega$ para todo $i \in \{0, \dots, 2\ell + 1\}$, v_0 el vértice inicial y $\{v_{2j}, v_{2j+1}\} \in I$ para todo $j \in \{0, \dots, \ell\}$

- ▶ Partimos con una transición de tipo 1 removiendo $\{v_0, v_1\}$
- ▶ Denotamos al camino simple resultante como $O = (v_1, \dots, v_{2j+1}, v_0)$, con extremos v_1, v_0
- ▶ Tratamos a O igual que cualquier camino simple utilizando v_0 o v_1 como el vértice inicial
 - ▶ Se utiliza el orden inverso para O para diferenciar a los ciclos simples de los caminos simples durante el análisis del mixing time de la cadena de Markov

Definición de los caminos canónicos: un ejemplo

Consideremos el ejemplo introducido anteriormente. Supongamos que queremos encontrar el camino canónico entre los siguientes matchings:



Definición de los caminos canónicos: un ejemplo

La diferencia simétrica $(I \setminus F) \cup (F \setminus I) = \{\{1, 3\}, \{2, 4\}, \{1, 2\}\}$, lo que corresponde al camino simple $Q_{11} = (3, 1, 2, 4)$

- ▶ Este camino no es un ciclo y su primera arista es $\{3, 1\} \in I$

Luego, el camino canónico entre estos estados corresponde a:

- ▶ Utilizar una transición de tipo 1 para remover $\{3, 1\}$
- ▶ Utilizar una transición de tipo 3 para reemplazar $\{2, 4\}$ por $\{1, 2\}$

Definición de los caminos canónicos: un ejemplo

En un ejemplo más extenso, simplemente realizamos las transiciones correspondientes a cada camino y ciclo simple en el orden establecido entre ellos.

- ▶ Nótese que, dado que los caminos y ciclos simples utilizados tienen que ser disjuntos, solo habrá una forma de crear cada camino canónico

Matchings de un grafo: mixing time

A partir del conjunto de caminos canónicos descritos, podemos establecer la siguiente cota para el mixing time de la cadena.

Suponiendo que el grafo (no dirigido) de entrada es $G = (V, E)$, $M \in \Omega$ es un matching arbitrario de G y $\varepsilon \in (0, 1)$, se tiene que:

$$\begin{aligned}\tau_M(\varepsilon) &\leq 128 \cdot |E|^2 \cdot \left(\ln |\Omega| + \ln \frac{1}{\varepsilon} \right) \\ &\leq 128 \cdot |E|^2 \cdot \left(\ln 2^{|E|} + \ln \frac{1}{\varepsilon} \right) \\ &= 128 \cdot |E|^2 \cdot \left(|E| \cdot \ln 2 + \ln \frac{1}{\varepsilon} \right)\end{aligned}$$