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.

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

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 $\vec{\pi}$

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 $\vec{\pi}$

Markov Chain Monte Carlo

Ejemplo

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

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

- Si $\sigma_1(\varphi) = \sigma_2(\varphi) = 1$, entonces $\vec{\pi}[\sigma_1] = \vec{\pi}[\sigma_2]$
- $\displaystyle\sum_{\sigma \,:\, \sigma(\varphi)=0} \vec{\pi}[\sigma] \ \leq \ \frac{1}{2}$

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

Markov Chain Monte Carlo

Ejemplo

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

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

- Si $\sigma_1(\varphi) = \sigma_2(\varphi) = 1$, entonces $\vec{\pi}[\sigma_1] = \vec{\pi}[\sigma_2]$
- $\displaystyle\sum_{\sigma \,:\, \sigma(\varphi)=0} \vec{\pi}[\sigma] \ \leq \ \frac{1}{2}$

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

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

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 $\vec{\pi}$ a partir de una cadena de Markov que no tenga a $\vec{\pi}$ como distribución estacionaria?

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 $\vec{\pi}$ a partir de una cadena de Markov que no tenga a $\vec{\pi}$ 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?

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 $\vec{\pi}$ a partir de una cadena de Markov que no tenga a $\vec{\pi}$ 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)

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:

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

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

Una demostración de la proposición

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

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

Una demostración de la proposición

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

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

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**

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 nos indica que para que la cadena de Markov tenga a $\vec{\pi}$ 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 $\vec{\pi}$

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 $\vec{\pi}$?

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:

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

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:

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

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 transición 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

Considere 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

Considere 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.

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.

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

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 $\displaystyle\sum_{a \in \Omega} Q[a, b] = 1$ para todo $b \in \Omega$

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

- ¿Por qué?

Metropolis-Hastings: la idea del algoritmo

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

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

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

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:

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

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

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 $\alpha : \Omega \times \Omega \to [0, 1]$ 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 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 $\alpha : \Omega \times \Omega \to [0, 1]$ 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 $\{a_1, a_2, \ldots, a_N\}$