(MCMC, Markov Chain Monte Carlo)

MONTECARLO POR CADENA DE MARKOV

↳ problema: estimar $\mu = E_\pi[g(x)]$

Generar $\vec{x}_1, \dots \vec{x}_n \sim \pi$ y calcular $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$

Como aplicar el metodo montecarlo?

CONSTRUIR U CADENA DE MARKOV TAL QUE $\lim_{t\to\infty} \pi_t = \pi$

↳ SI SE CUMPLEN LAS CONDICIONES NECESARIAS, SE TENDRÁ QUE π SERÁ ESTACIONARIA PARA LA CADENA DE MARKOV

$$\begin{array}{cccc} x_0 \sim x & \dots & \dots & x_t \\ ? & ? & \dots & ? \\ \pi_1 \simeq \pi & \pi_2 \simeq \pi & & \pi_t \simeq \pi \end{array}$$

Entonces, para aplicar el metodo:

1. Obtener realizaciones de la cadena de Markov hasta un cierto t suficientemente grande (CON VALORES INICIALES INDEPENDIENTES)

2. Recuperar los valores $x_1, \dots x_n$. de la variable X_t para cada una de esas realizaciones.

3. $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$

$$\left.\begin{array}{l} \xrightarrow{t \text{ pasos}} x_1 \\ \longrightarrow x_2 \\ \longrightarrow x_3 \\ \vdots \\ \longrightarrow x_n \end{array}\right\} \overset{iid}{\sim} \pi$$

En general, los valores $x_1, \dots x_n$ se toman de una unica realizacion de la cadena de Markov.

W8 - 09/11

Definición 1 *Un proceso estocástico es una colección de variables aleatorias $\{X_t\}_{t\in\mathcal{T}}$ sobre un mismo espacio de probabilidad, donde $\mathcal{T}$ es cualquier conjunto de índices.*

El conjunto E de posibles valores de cada X_t (que asumimos que es independiente de t) se llama el espacio de estados del proceso.

El conjunto de índices $\mathcal{T}$ se toma a menudo como un subconjunto numerable o continuo de $\mathbb{R}$, considerando de esa forma un proceso aleatorio como una variable aleatoria que evoluciona a lo largo del tiempo.

Definición 2 *Una cadena de Markov de tiempo discreto es un proceso estocástico $\boldsymbol{X} = \{X_t\}_{t\in\mathbb{Z}_{\geq 0}}$ que cumple la propiedad de Markov:*

$$\boxed{\mathbb{P}(X_{t+1} \mid X_0, \dots, X_t) = \mathbb{P}(X_{t+1} \mid X_t), \quad \forall t \in \mathbb{Z}_{\geq 0}}$$

Una cadena de Markov es homogénea en el tiempo si las probabilidades anteriores no dependen del índice t:

$$\mathbb{P}(X_{t+1} \mid X_t) = \mathbb{P}(X_1 \mid X_0), \quad \forall t \in \mathbb{Z}_{\geq 0}$$

En lo que sigue solo consideraremos cadenas de Markov de tiempo discreto homogéneas.

La distribución de una cadena de Markov homogénea viene determinada por

1. La distribución marginal de X_0, llamada *distribución inicial*.
2. La distribución condicional $X_{t+1} \mid X_t$, llamada *distribución de transición*.

Si el espacio de estados de la cadena de Markov es finito, $E = \{x_1, \dots, x_d\}$, entonces la distribución inicial se puede representar como un vector $(p_1, \dots, p_d) \in \mathbb{R}^d_{\geq 0}$ tal que

$$\sum_{i=1}^{d} p_i = 1$$
$$\mathbb{P}(X_0 = x_i) = p_i, \quad \forall i = 1, \dots, d$$

y la distribución de transición como una matriz $P = (p_{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{d\times d}_{\geq 0}$ tal que

$$\sum_{j=1}^{d} p_{ij} = 1, \quad \forall i = 1, \dots, d$$
$$\mathbb{P}(X_{t+1} = x_j \mid X_t = x_i) = p_{ij}, \quad \forall i, j = 1, \dots, d$$

Si el espacio de estados es infinito numerable, entonces las distribuciones inicial y de transición se pueden representar como un vector y una matriz infinitas, respectivamente.

Si el espacio de estados es continuo, entonces debemos pensar en la distribución inicial como una distribución no condicional y en la distribución de transición como una distribución condicional.

E finito o numerable

Proposición 3 *Denotemos por $p_{ij}^{(t)}$ la probabilidad de que una cadena de Markov $\{X_t\}_{t\in\mathbb{Z}_{\geq 0}}$ pase del estado i al estado j en t unidades de tiempo, esto es*

$$p_{ij}^{(t)} = \mathbb{P}(X_t = j \mid X_0 = i)$$

Las ecuaciones de Chapman-Kolmogorov establecen que

$$p_{ij}^{(t+t')} = \sum_{k\in E} p_{ik}^{(t)} p_{kj}^{(t')}, \quad \forall i,j \in E, \forall t,t' \geq 0$$

Para las cadenas de Markov con espacio de estado finito o infinito numerable, denotando por $P^{(t)}$ la matriz de probabilidades de transición en t pasos, las ecuaciones anteriores se traducen en que $P^{(t)}$ se puede obtener multiplicando P consigo misma t veces.

Definición 4 *Sea $\boldsymbol{X}$ una cadena de Markov con espacio de estados E y matriz de transición P. Sea $\boldsymbol{\pi} = (\pi_i)_{i\in E}$ una distribución de probabilidad sobre E, es decir,*

$$0 \leq \pi_i \leq 1, \forall i \in E \qquad \text{y} \qquad \sum_{i\in E} \pi_i = 1$$

- *π es* invariante *o* estacionaria *para $\boldsymbol{X}$ si*

  $$X_0 \sim \boldsymbol{\pi} \quad \Rightarrow \quad X_t \sim \boldsymbol{\pi}, \;\; \forall t \in \mathbb{Z}_{\geq 0}$$

  Es fácil comprobar que esto se verifica si y solo si $\boxed{\boldsymbol{\pi} = \boldsymbol{\pi} P.}$

- *π es una* distribución límite *de $\boldsymbol{X}$ si*

  $$\boxed{\pi_j = \lim_{t\to+\infty} \mathbb{P}(X_t = j \mid X_0 = i), \quad \forall i,j \in E}$$

Definición 5

1. *Un estado j se dice que es* accesible *desde un estado i si $p_{ij}^{(t)} > 0$, para algún $t \geq 0$.*
2. *Dos estados i y j se dice que* se comunican*, $i \leftrightarrow j$, si cada uno es accesible desde el otro.*

Proposición 6 *La relación de comunicación es una relación de equivalencia sobre el espacio de estados de la cadena de Markov. Es decir,*

1. *Para todo estado $i \in E$, $i \leftrightarrow i$.*
2. *Para cualesquiera estados $i,j \in E$, si $i \leftrightarrow j$, entonces $j \leftrightarrow i$.*
3. *Para cualesquiera estados $i,j,k \in E$, si $i \leftrightarrow j$ y $j \leftrightarrow k$, entonces $i \leftrightarrow k$.*

Dos estados que se comunican se dice que están en la misma clase *(de equivalencia).*

Definición 7 *Una cadena de Markov se dice que es* irreducible *si posee una única clase. Es decir, si todos sus estados se comunican entre sí.*

Definición 8 *Un estado i se dice que tiene periodo n si*

$$n = \text{mcd}\{t > 0 \mid p_{ii}^{(t)} > 0\}$$

Si $n > 1$ el estado se dice que es periódico *y si $n = 1$ se dice que es* aperiódico.

Proposición 9

1. *Todos los estados de una misma clase comunicante tienen el mismo periodo.*
2. *En una cadena de Markov irreducible todos los estados tienen el mismo periodo.*

Definición 10 *La variable aleatoria $\tau(i)$ que dada una cadena de Markov $\{X_t\}_{t\in\mathbb{Z}_{\geq 0}}$ cuenta el mínimo número de transiciones necesarias para alcanzar el estado i se define como*

$$\tau(i) = \text{ínf}\{t \geq 1 \mid X_t = i\}$$

donde $\tau(i) = +\infty$ si el conjunto anterior es el conjunto vacío.

Teorema 11 *Sea $\boldsymbol{X}$ una cadena de Markov irreducible y aperiódica con espacio de estados finito $\boldsymbol{E}$. Entonces,*

1. *$\boldsymbol{X}$ tiene una única distribución invariante $\boldsymbol{\pi} = (\pi_i)_{i\in E}$.*
2. *π es distribución límite de $\boldsymbol{X}$.*
3. *$\pi_i = \frac{1}{m_i}$ para todo $i \in E$, donde $m_i = \mathbb{E}[\tau(i) \mid X_0 = i]$.*

Definición 12

1. *Un estado i se dice que es* recurrente *si $\mathbb{P}\big(\tau(i) < +\infty \mid X_0 = i\big) = 1$, en caso contrario se dice que es* transitorio.
2. *Un estado recurrente i se dice que es* positivo recurrente *si $\mathbb{E}[\tau(i) \mid X_0 = i] < +\infty$, en caso contrario se dice que es* cero recurrente.

Proposición 13

1. *La recurrencia (positiva) y la transitoriedad son propiedades de clase.*
2. *En una cadena de Markov con espacio de estados finito todo estado recurrente es positivo recurrente.*
3. *Todos los estados de una cadena de Markov irreducible con espacio de estados finito son positivo recurrentes.*

Teorema 14 *Sea* $\boldsymbol{X}$ *una cadena de Markov irreducible y aperiódica con espacio de estados infinito numerable* $\boldsymbol{E}$*. Entonces ocurre uno de los siguientes casos:*

- *Todos los estados son transitorios y*

$$\lim_{t \to +\infty} \mathbb{P}(X_t = j \mid X_0 = i) = 0, \quad \forall i, j \in E$$

- *Todos los estados son cero recurrentes y*

$$\lim_{t \to +\infty} \mathbb{P}(X_t = j \mid X_0 = i) = 0, \quad \forall i, j \in E$$

- *Todo los estados son positivo recurrentes y*
  1. $\boldsymbol{X}$ *tiene una única distribución invariante* $\boldsymbol{\pi} = (\pi_i)_{i \in E}$.
  2. $\boldsymbol{\pi}$ *es distribución límite de* $\boldsymbol{X}$.
  3. $\pi_i = \frac{1}{m_i} > 0$ *para todo* $i \in E$*, donde* $m_i = \mathbb{E}[\tau(i) \mid X_0 = i]$.

Definición 15 *Una cadena de Markov* $\{X_t\}_{t \in \mathbb{Z}_{\geq 0}}$ *estacionaria se dice que es* reversible *si*

$$\mathbb{P}(X_{t+1} \mid X_{t+2} = i) = \mathbb{P}(X_{t+1} \mid X_t = i)$$

para todo $i \in E$.

Definición 16 *La matriz de transición* P *de una cadena de Markov se dice que está en* balance detallado *con una distribución de probabilidad* $\boldsymbol{\pi}$ *si*

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad \forall i, j \in E$$

Teorema 17 *Sea* $\boldsymbol{X}$ *una cadena de Markov cuya matriz de transición está en balance detallado con una distribución de probabilidad* $\boldsymbol{\pi}$*. Entonces,*

1. $\boldsymbol{\pi}$ *es la distribución invariante de* $\boldsymbol{X}$.
2. *Tomando* $\boldsymbol{\pi}$ *como distribución inicial,* $\boldsymbol{X}$ *es reversible.*

Teorema 18 *Sean* $\{X_t\}_{t \in \mathbb{Z}_{\geq 0}}$ *una cadena de Markov irreducible con distribución estacionaria* $\boldsymbol{\pi}$ *y* $g \colon E \to \mathbb{R}$ *una función acotada. Entonces,*

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{t=0}^{n} g(x_t) = \sum_{i \in E} g(i)\pi_i = \mathbb{E}_\pi[g(X)]$$

con probabilidad 1*, para cualquier realización* $(x_t)_{t=0,\ldots,n}$ *de la cadena de Markov.*

E continuo

Asumimos $E = \mathbb{R}^d$ y que la distribución de transición de la cadena de Markov se expresa mediante un *núcleo de transición* $K\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, de manera que

$$\mathbb{P}(X_{t+1} \in A \mid X_t = x) = \int_A K(x, y)\, dy$$

para todo conjunto de Borel $A \subseteq \mathbb{R}^d$. Es decir, $K(x, \cdot)$ es justamente la función de densidad condicionada de X_{t+1} dado $X_t = x$.

Proposición 19 *La transición en t unidades de tiempo de la cadena de Markov viene dada por*

$$\mathbb{P}(X_t \in A \mid X_0 = x_0) = \int_A K^{(t)}(x_0, x_t)\, dx_t$$

donde

$$K^{(t)}(x_0, x_t) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} K(x_0, x_1) \cdots K(x_{t-1}, x_t)\, dx_{t-1} \cdots dx_1$$

Las ecuaciones de Chapman-Kolmogorov *establecen que*

$$K^{(t+t')}(x, y) = \int_{\mathbb{R}^d} K^{(t)}(x, z) K^{(t')}(z, y)\, dz$$

Definición 20 *Una distribución de probabilidad π con función de densidad f_π es* invariante *o* estacionaria *para el núcleo de transición K si*

$$f_\pi(y) = \int_{\mathbb{R}^d} f_\pi(x) K(x, y)\, dx, \quad \forall y \in \mathbb{R}^d$$

Si $\{X_t\}_{t \in \mathbb{Z}_{\geq 0}}$ es una cadena de Markov con distribución invariante π y tomamos π como su distribución inicial, entonces $X_t \sim \pi$ para todo $\mathrm{t} \in \mathbb{Z}_{\geq 0}$.

Definición 21 *Dada una medida φ sobre $\mathbb{R}^d$, una cadena de Markov se dice que es φ-irreducible si para todo conjunto de Borel $A \subseteq \mathbb{R}^d$ tal que $\varphi(A) > 0$ y para todo $x \in \mathbb{R}^d$ existe un $t > 0$ tal que $\mathbb{P}(X_t \in A \mid X_0 = x) > 0$.*

Definición 22 *Una cadena de Markov con distribución invariante π se dice que es* aperiódica *si no existen $n > 1$ y subconjuntos de Borel disjuntos $A_1, \dots, A_n \subseteq \mathbb{R}^d$ tales que:*

1. *$\pi(A_i) > 0$, para todo $i = 1, \dots, n$.*
2. *$\mathbb{P}(X_{t+1} \in A_{i+1} \mid X_t \in A_i) = 1$, para todo $i = 1, \dots, n-1$ y $\mathbb{P}(X_{t+1} \in A_1 \mid X_t \in A_n) = 1$.*

Definición 23 *El* número de visitas *de una cadena de Markov a un conjunto $A \subseteq \mathbb{R}^d$ viene dado por*

$$V(A) = \sum_{t=0}^{+\infty} \mathbb{1}(X_t \in A)$$

Definición 24

1. *Un conjunto de Borel $A \subseteq \mathbb{R}^d$ se dice que es* recurrente *si para todo $x \in A$*

$$\mathbb{E}[V(A) \mid X_0 = x] = +\infty$$

2. *Una cadena de Markov se dice que es* recurrente *si*
   a) *La cadena es φ-irreducible para alguna medida φ.*
   b) *Todo conjunto de Borel $A \subseteq \mathbb{R}^d$ con $\varphi(A) > 0$ es recurrente.*

Definición 25

1. *Un conjunto de Borel $A \subseteq \mathbb{R}^d$ se dice que es* Harris-recurrente *si para todo $x \in A$*

$$\mathbb{P}\big(V(A) = +\infty \mid X_0 = x\big) = 1$$

2. *Una cadena de Markov se dice que es* Harris-recurrente *si*
   a) *La cadena es φ-irreducible para alguna medida φ.*
   b) *Todo conjunto de Borel $A \subseteq \mathbb{R}^d$ con $\varphi(A) > 0$ es Harris-recurrente.*

Proposición 26 *Toda cadena de Markov φ-irreducible que tenga una distribución invariante es recurrente.*

Definición 27 *La* distancia de variación total *entre dos distribuciones de probabilidad π_1 y π_2 se define como*

$$\|\pi_1 - \pi_2\| = \sup_{A \subseteq \mathbb{R}^d} |\pi_1(A) - \pi_2(A)|$$

Teorema 28 *Si una cadena de Markov con función núcleo K es φ-irreducible y aperiódica y tiene distribución invariante π, entonces para π-casi todo $x \in \mathbb{R}^d$ se tiene que*

$$\lim_{t \to +\infty} \left\| K^{(t)}(x, \cdot) - \pi \right\| = 0$$

En particular, $\lim_{t \to +\infty} \mathbb{P}(X_t \in A \mid X_0 = x) = \pi(A)$, para todo conjunto de Borel $A \subseteq \mathbb{R}^d$.

Si la cadena es Harris-recurrente, entonces el resultado se da para todo $x \in \mathbb{R}^d$.

Definición 29 *Un núcleo de transición K se dice que está en* balance detallado *con una distribución π con función de densidad f_π si*

$$f_\pi(x)K(x,y) = f_\pi(y)K(y,x)$$

para todo $x, y \in \mathbb{R}^d$.

Teorema 30 *Sea $\boldsymbol{X}$ una cadena de Markov cuyo núcleo de transición está en balance detallado con una distribución de probabilidad π. Entonces,*

1. *π es la única distribución invariante de $\boldsymbol{X}$.*
2. *Tomando π como distribución inicial, $\boldsymbol{X}$ es reversible.*

Teorema 31 *Sean $\{X_t\}_{t\in\mathbb{Z}_{\geq 0}}$ una cadena de Markov φ-irreducible con distribución estacionaria π y $g\colon \mathbb{R}^d \to \mathbb{R}$ una función π-integrable. Entonces, para π-casi todo $x_0 \in \mathbb{R}^d$*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{t=0}^{n} g(x_t) = \int_{\mathbb{R}^d} g(x) f_\pi(x)\, dx = \mathbb{E}_\pi[g(X)]$$

con probabilidad 1 para cualquier realización $\{x_t\}_{t=0,\ldots,n}$ de la cadena de Markov.

Si la cadena es Harris-recurrente, entonces el resultado se verifica para todo $x_0 \in \mathbb{R}^d$.