15 JULIO

jueves, 15 de julio de 2021

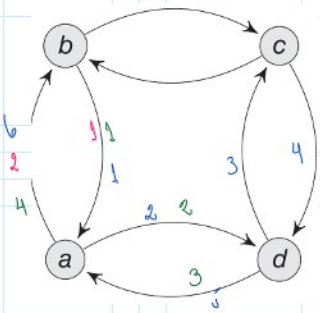
Period

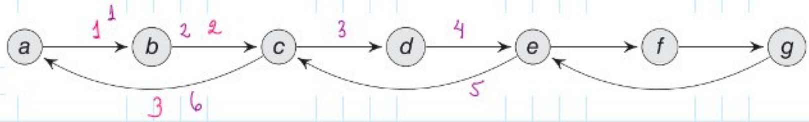
For a Markov chain with transition matrix $\boldsymbol{P}$, the *period* of state i, denoted $d(i)$, is the greatest common divisor of the set of possible return times to i. That is,

$$d(i) = \gcd\{n > 0 : P^n_{ii} > 0\}.$$

If $d(i) = 1$, state i is said to be *aperiodic*. If the set of return times is empty, set $d(i) = +\infty$.

$\gcd\{3,5\} = 1$





Periodic, Aperiodic Markov Chain

A Markov chain is called *periodic* if it is irreducible and all states have period greater than 1.

A Markov chain is called *aperiodic* if it is irreducible and all states have period equal to 1.

Note that any state i with the property that $P_{ii} > 0$ is necessarily aperiodic. Thus, a sufficient condition for an irreducible Markov chain to be aperiodic is that $P_{ii} > 0$ for some i. That is, at least one diagonal entry of the transition matrix is nonzero.

CADENA DE MARKOV ERGODICA

A Markov chain is called *ergodic* if it is irreducible, aperiodic, and all states have finite expected return times.

Fundamental Limit Theorem for Ergodic Markov Chains

Theorem 3.8. *Let $X_0, X_1, \ldots$ be an ergodic Markov chain. There exists a unique, positive, stationary distribution $\boldsymbol{\pi}$, which is the limiting distribution of the chain. That is,*

$$\pi_j = \lim_{n\to\infty} P^n_{ij}, \text{ for all } i,j.$$

$X_0, X_1, X_2, X_3, \ldots$ ergódica $\{0,1,2,..,N\}$

Existe "π" distribución estacionaria

distribución límite

$\langle = \rangle$

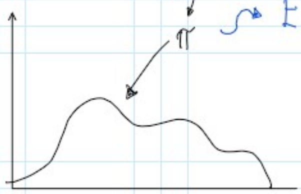
Una cadena de Markov es ergódica si y solo si la matriz de transición es regular

MCMC

MARKOV CHAIN MONTE CARLO → Simular!

Given a probability distribution $\boldsymbol{\pi}$, the goal of MCMC is to simulate a random variable X whose distribution is $\boldsymbol{\pi}$. The distribution may be continuous or discrete, although we assume at the beginning that $\boldsymbol{\pi}$ is discrete. Often, one wants to estimate an expectation or other function of a joint distribution from a high-dimensional space.

The MCMC algorithm constructs an ergodic Markov chain whose limiting distribution is the desired $\boldsymbol{\pi}$. One then runs the chain long enough for the chain to converge, or nearly converge, to its limiting distribution, and outputs the final element or elements of the Markov sequence as a sample from $\boldsymbol{\pi}$.



X π → $E[X] = \int_{-\infty}^{+\infty} x\pi(x)\,dx$ } difícil de integrar

Monte Carlo

$x_1, x_2, x_3, \ldots, x_n$ → $\widehat{E[X]} = \bar{X} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$; $n \to \infty$

→ Como generar? sean independientes

MCMC

distrib. estacionaria (distrib. de probabilidad)

π → } ergódica → distribucion estacionaria

$\pi(Q)$ → $X_0, X_1, X_2, X_3, X_4, \ldots, X_n$ } $P(X_{n+1} = j \mid X_n = i)$ } $x_1, x_2, x_4, x_5, \ldots$ "Lag"

correlación ≈ 0

$\boxed{T(Q)}$ $X_0, X_1, X_2, X_3, X_4, \dots, X_n$ } $P(X_{n+1} = j \mid X_n = i)$ | $X_1, X_2, X_4, X_5, \dots$

correlación ≈ 0

LEY DE LOS GRANDES NUMEROS

If $Y_1, Y_2, \dots$ is an i.i.d. sequence with common mean $\mu < \infty$, then the strong law of large numbers says that, with probability 1,

$$\lim_{n\to\infty} \frac{Y_1 + \cdots + Y_n}{n} = \mu.$$

Equivalently, let Y be a random variable with the same distribution as the Y_i and assume that r is a bounded, real-valued function. Then, $r(Y_1), r(Y_2), \dots$ is also an i.i.d. sequence with finite mean, and, with probability 1,

$$\lim_{n\to\infty} \frac{r(Y_1) + \cdots + r(Y_n)}{n} = E(r(Y)).$$

$\boxed{r(x) = x^2 - 2x + 3}$ ganancia

$E[r(x)] = ?$

Strong Law of Large Numbers for Markov Chains

Theorem 5.1. *Assume that $X_0, X_1, \dots$ is an ergodic Markov chain with stationary distribution $\boldsymbol{\pi}$. Let r be a bounded, real-valued function. Let X be a random variable with distribution $\boldsymbol{\pi}$. Then, with probability 1,*

$$\lim_{n\to\infty} \frac{r(X_1) + \cdots + r(X_n)}{n} = E(r(X)),$$

where $E(r(X)) = \sum_j r(j)\pi_j$.

Given an ergodic Markov chain with stationary distribution $\boldsymbol{\pi}$, let A be a nonempty subset of the state space. Write $\pi_A = \sum_{j\in A} \pi_j$. From Theorem 5.1, we can interpret π_A

$P(X \in A) = \pi_A$

as the long-term average number of visits of an ergodic Markov chain to A. Define the indicator variable

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Then, $\sum_{k=0}^{n-1} I_A(X_k)$ is the number of visits to A in the first n steps of the chain. Let X be a random variable with distribution $\boldsymbol{\pi}$. With probability 1,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(X_k) = E(I_A(X)) = P(X \in A) = \pi_A.$$

Example 5.1 Bob's daily lunch choices at the cafeteria are described by a Markov chain with transition matrix

$$\boldsymbol{P} = \begin{array}{l} \\ \text{Yogurt} \\ \text{Salad} \\ \text{Hamburger} \\ \text{Pizza} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Yogurt} & \text{Salad} & \text{Hamburger} & \text{Pizza} \end{array} \\ \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 1/4 & 1/2 \\ 1/4 & 0 & 1/4 & 1/2 \end{pmatrix} \end{array}.$$

Yogurt costs \$3.00, hamburgers cost \$7.00, and salad and pizza cost \$4.00 each. Over the long term, how much, on average, does Bob spend for lunch?

$$r(x) = \begin{cases} 3, & \text{if } x = \text{yogurt}, \\ 4, & \text{if } x = \text{salad or pizza}, \\ 7, & \text{if } x = \text{hamburger}. \end{cases}$$

The lunch chain is ergodic with stationary distribution

Yogurt	Salad	Hamburger	Pizza
7/65	2/13	18/65	6/13

Laboratorio!

$$\sum_x r(x)\pi_x = 3\left(\frac{7}{65}\right) + 4\left(\frac{2}{13} + \frac{6}{13}\right) + 7\left(\frac{18}{65}\right) = \$4.72$$

ALGORITMO METROPOLIS HASTINGS

Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots)$ be a discrete probability distribution. The algorithm constructs a reversible Markov chain $X_0, X_1, \dots$ whose stationary distribution is $\boldsymbol{\pi}$.

Let $\boldsymbol{T}$ be a transition matrix for *any* irreducible Markov chain with the same state space as $\boldsymbol{\pi}$. It is assumed that the user knows how to sample from $\boldsymbol{T}$. The $\boldsymbol{T}$ chain will be used as a *proposal* chain, generating elements of a sequence that the algorithm decides whether or not to *accept*.

To describe the transition mechanism for $X_0, X_1, \dots$, assume that at time n the chain is at state i. The next step of the chain X_{n+1} is determined by a two-step procedure.

1. Choose a new state according to $\boldsymbol{T}$. That is, choose j with probability T_{ij}. State j is called the *proposal state*.
2. Decide whether to accept j or not. Let

$$a(i,j) = \frac{\pi_j T_{ji}}{\pi_i T_{ij}}.$$

The function a is called the *acceptance function*. If $a(i,j) \geq 1$, then j is accepted

$X_0, X_1, \dots$ } muestra aleatoria

$\pi \rightarrow$ estados $\{1, 2, 3, 4, \dots, N\}$

$T = \begin{array}{c|cccc} & 1 & 2 & 3 \;\dots & N \\ 1 & T_{11} & T_{12} & T_{13} & T_{1N} \\ 2 & \vdots & \vdots & \vdots & \vdots \\ 3 & & & & \\ N & & & & \end{array}$ $\rightarrow X_1$ aceptar / no aceptar

se propone

$X_n = i \rightarrow X_{n+1} = j$ }

$P(X_{n+1} = j \mid X_n = i)$

Si $a(i,j) \geq 1 \rightarrow X_{n+1} = j$

$$a(i,j) = \frac{\pi_j T_{ji}}{\pi_i T_{ij}}.$$

The function a is called the *acceptance function*. If $a(i,j) \geq 1$, then j is accepted as the next state of the chain. If $a(i,j) < 1$, then j is accepted with probability $a(i,j)$. If j is not accepted, then i is kept as the next step of the chain.

In other words, assume that $X_n = i$. Let U be uniformly distributed on $(0, 1)$. Set

$$X_{n+1} = \begin{cases} j, & \text{if } U \leq a(i,j), \\ i, & \text{if } U > a(i,j). \end{cases}$$

Si $a(i,j) \geq 1 \rightarrow X_{n+1} = j$

Si $a(i,j) < 1$

$\hookrightarrow X_{n+1} = \begin{cases} j, & \text{si } U \leq a(ij) \\ i, & \text{si } \underbrace{U}_{\text{número aleatorio en } (0,1)} > a(i,j) \end{cases}$

Metropolis–Hastings Algorithm

The sequence $X_0, X_1, \ldots$ constructed by the Metropolis–Hastings algorithm is a reversible Markov chain whose stationary distribution is $\boldsymbol{\pi}$.

Remarks:

1. The exact form of $\boldsymbol{\pi}$ is not necessary to implement Metropolis–Hastings. The algorithm only uses ratios of the form π_j/π_i. Thus, $\boldsymbol{\pi}$ needs only to be specified up to proportionality. For instance, if $\boldsymbol{\pi}$ is uniform on a set of size c, then $\pi_j/\pi_i = 1$, and the acceptance function reduces to $a(i,j) = T_{ji}/T_{ij}$.
2. If the proposal transition matrix $\boldsymbol{T}$ is symmetric, then $a(i,j) = \pi_j/\pi_i$.
3. The algorithm works for *any* irreducible proposal chain. Thus, the user has wide latitude to find a proposal chain that is efficient in the context of their problem.
4. If the proposal chain is ergodic (irreducible and aperiodic in the finite case) then the resulting Metropolis–Hastings chain is also ergodic with limiting distribution $\boldsymbol{\pi}$.
5. The generated sequence $X_0, X_1, \ldots, X_n$ gives an approximate sample from $\boldsymbol{\pi}$. However, if the chain requires many steps to get close to stationarity, there may be initial bias. *Burn-in* refers to the practice of discarding the initial iterations and retaining $X_m, X_{m+1}, \ldots, X_n$, for some m. In that case, the strong law of large numbers for Markov chains gives

$$\lim_{n\to\infty} \frac{r(X_m) + \cdots + r(X_n)}{n - m + 1} = \sum_x r(x)\pi_x.$$

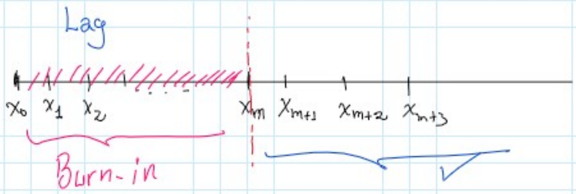
$\pi^* = (30, 20, 40, 10)$

$\hookrightarrow \pi = (0.3, 0.2, 0.40, 0.1)$

$\pi = (\pi_1, \pi_2, .., \pi_k)$

si es uniforme:

$\pi_1 = \pi_2 = \cdots = \pi_k = 1/k$



Example 5.2 Power-law distributions are positive probability distributions of the form $\pi_i \propto i^s$, for some constant S. Unlike distributions with exponentially decaying tails (e.g., Poisson, geometric, exponential, normal), power-law distributions have *fat tails*, and thus are often used to model skewed data. Let

$$\pi_i = \frac{i^{-3/2}}{\sum_{k=1}^{\infty} k^{-3/2}}, \text{ for } i = 1, 2, \ldots$$

Implement a Metropolis–Hastings algorithm to simulate from $\boldsymbol{\pi}$.

$\pi_i \propto i^s$; $i = 1, 2, 3, 4, \ldots$

$\pi_i = \frac{i^s}{\sum_{i} i^s}$; $s = -3/2$

$\pi_i \propto i^{-3/2}$

Generar v.a's "i" → ?

SOLUCION

$$T_{ij} = \begin{cases} 1/2, & \text{if } j = i \pm 1 \text{ for } i > 1, \\ 1, & \text{if } i = 1 \text{ and } j = 2, \\ 0 & \text{otherwise.} \end{cases}$$

with

$$a(i, i+1) = \left(\frac{i}{i+1}\right)^{3/2}, \text{ and } \quad a(i+1, i) = \left(\frac{i+1}{i}\right)^{3/2}, \text{ for } i \geq 2,$$

$$a(1,2) = \frac{\pi_2 T_{21}}{\pi_1 T_{12}} = \left(\frac{1}{2}\right)^{3/2} \frac{1}{2} = \left(\frac{1}{2}\right)^{5/2} \quad \text{and} \quad a(2,1) = 2^{5/2}.$$

$$a(i+1, i) \geq 1, \text{ for all } i.$$

ALGORITMO METROPOLIS HASTINGS: ESPACIO DE ESTADOS CONTINUO

Example 5.5 Using only a uniform random number generator, simulate a standard normal random variable using MCMC.