Markov Chain MonteCarlo

Marcos Mobbili
Carolina Daza

Resumen

Markov chain Montecarlo (MCMC) es una técnica que permite hacer un muestreo efficiente de datos de funciones de probabilidad complicadas.

Aplicado a inferencia bayesiana ayuda a estimar la probabilidad a posteriori cuando no se puede calcular analíticamente su distribución.

1 Algunos conceptos

Muestreo, aceptación/rechazo
Propiedades de Cadenas de Markov
Markov Chain Montecarlo
Algoritmo de Metropolis-Hastings
Enfoque bayesiano

2 Ejemplo práctico

¿Para que muestrear?

Aproximar

Estimar esperanzas de funciones complicadas
Inferencia de posteriors

Visualización

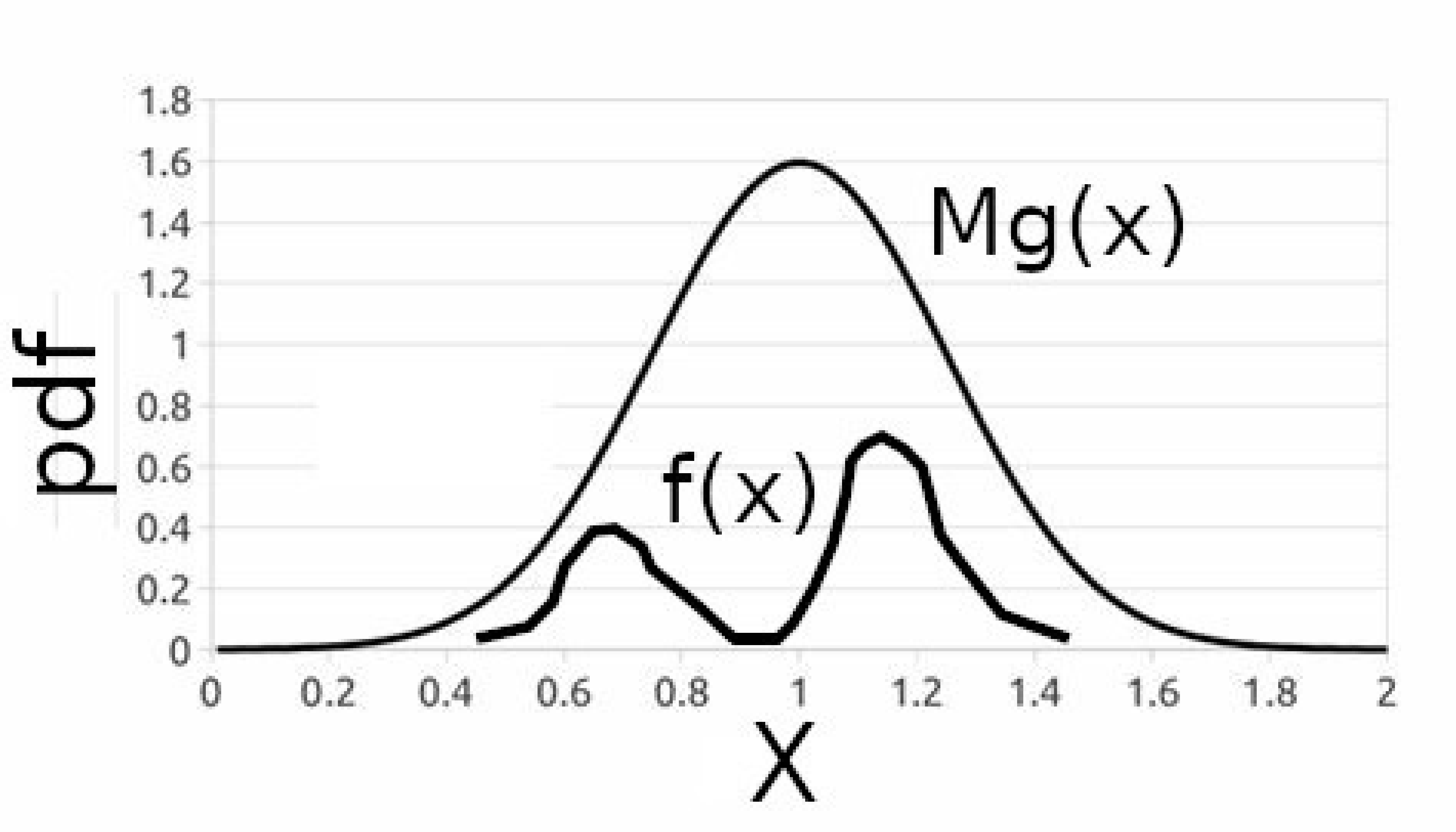
Es mas fácil la interpretación del proceso y que calcular directamente las probabilidades o aproximarlas

Es de uso general

Es aplicable a problemas de variada índole

Aceptación/Rechazo

Queremos muestrear de $p(x) \propto f(x)$, buscamos una función que cubra a f
La candidata es una normal escaleada por un número M



Proceso

Se samplea una muestra de g
Se acepta con probabilidad
f(s)/Mg(s)

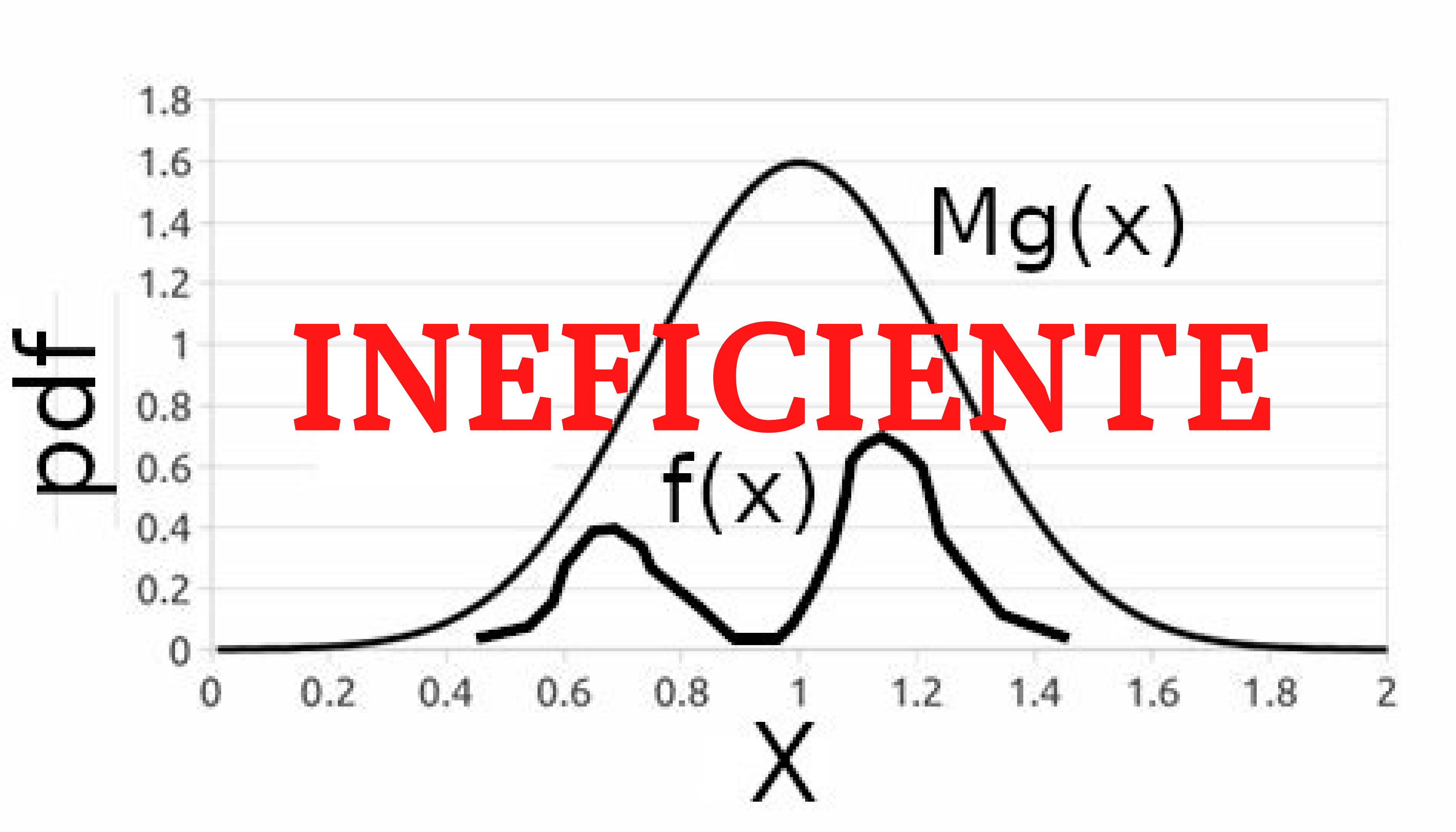
Intuición

Si el cociente es grande, la muestra es muy probable en f(x) y menos en g(x), entonces la tomo porque es de alta probabilidad y rara

Aceptación/Rechazo

Queremos muestrear de $p(x) \propto f(x)$

La candidata es una normal escaleada por un número M



Proceso

Se samplea una muestra de g

Se acepta con probabilidad

f(s)/Mg(s)

Intuición

Si el cociente es grande, la muestra es muy probable en f(x) y menos en g(x), entonces la tomo porque es de alta probabilidad y rara

Objetivo

Muestrear de una distribución de probabilidad p, que puede tener una forma muy complicada, o aproximar E[f(x)]

Intuición

Se comienza en algún punto del espacio y nos vamos moviendo hacia una región donde la probabilidad de esa distribución es alta y se trata de mantenerse en esa región

Propiedades de las cadenas de Markov

La propiedad principal es que un estado depende solamente del estado anterior:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \ldots X_n = x_n) \\ = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

Estacionaridad

Bajo supuestos de aperiodicidad e irreductibilidad, las probabilidades marginales de los estados tienden a un estacionario:

$$\pi T = \pi$$

Una cadena que cumple estas propiedades se dice *Ergódica*

Ecuación de Chapman-Kolmogorov

La probabilidad de ir en n pasos de un estado i a uno j si la cadena es homogénea es:

$$p_{i,j}(n) = T^n_{i,j}$$

y no depende del estado inicial

Reversibilidad (Balance Detallado)

Si π cumple esta propiedad, π es una distribucion estacionaria con matriz de transición T:

$$\pi_a T_{ab} = \pi_b T_{ba} \quad \forall\, a, b \in \Pi$$

Markov Chain MonteCarlo

Markov + MonteCarlo

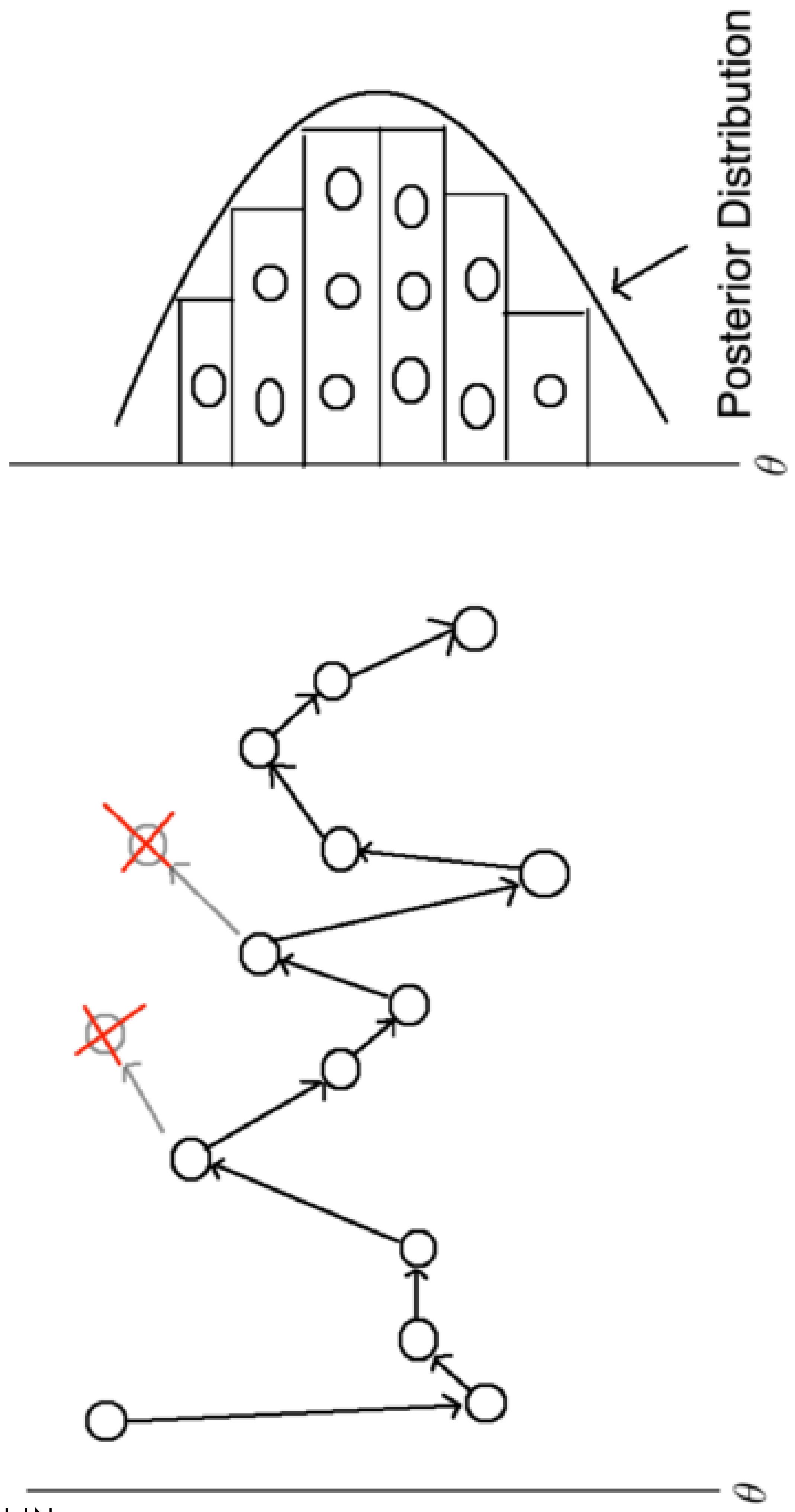
Usamos las propiedades de la cadena para estimar/aproximar empíricamente una función

Proceso

Se toma un estado inicial
Se aplica muchas veces la matriz de transición T (o una función de transición g)
Los estados obtenidos van a converger a una distribución de probabilidad en el estacionario que se aproxima al valor esperado de la función real o a la distribución real

¿Como se calcula?

Algoritmos de aceptación/rechazo como el de Metropolis/Metropolis-Hastings (1953)



CORY MAKLIN
https://towardsdatascience.com/monte-carlo-markov-chain-89cb7e844c75

Algoritmo de Metrópolis-Hastings

Paso 0

Sugerir una función de transición de los estados, por ejemplo una normal

$$g(x_{t+1}|x_t) = \mathcal{N}(x_t, \sigma^2)$$

Paso 1

Se muestrea un estado inicial con esta distribución

$$x^*_{t+1} = x_t + \mathcal{N}(x_t, \sigma^2)$$

Paso 2

Se calcula la probabilidad de aceptación A de la muestra

$$\frac{A(t \to t+1)}{A(t+1 \to t)} = \frac{f(x^*_{t+1})g(x_t|x^*_{t+1})}{f(x_t)g(x^*_{t+1}|x_t)}$$

esta es casi una

POSTERIOR

Paso 3

Si la probabilidad es mayor que 1 me quedo la muestra, si no lanzo un numero aletatorio y dependiendo de su valor, acepto si esta probabilidad es mayor que A

Enfoque Bayesiano

$$\theta : \theta_0, \theta_1 \ldots \theta_n$$

Paso 0

Sugerir una función de transición de los estados, por ejemplo una normal

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Paso 1

Se muestrea un estado inicial con esta distribución

$$A(t \rightarrow t^*) = min\left\{1, \frac{p(\theta_t^*|\mathcal{D})q(\theta_t^*, \theta_t)}{p(\theta_t|\mathcal{D})q(\theta_t, \theta_t^*)}\right\}$$

Paso 2

Se calcula la probabilidad de aceptación A de la muestra

Likelihood

$$A(t \rightarrow t^*) = min\left\{1, \frac{\boxed{p(\mathcal{D}|\theta_t^*)}\boxed{p(\theta_t^*)}q(\theta_t^*, \theta_t)}{p(\mathcal{D}|\theta_t)p(\theta_t)q(\theta_t, \theta_t^*)}\right\}$$

prior

Paso 3

Si la probabilidad es mayor que 1 me quedo la muestra, si no lanzo un numero aletatorio y dependiendo de su valor, acepto si esta probabilidad es mayor que A

$$r = p_c^*/p_0 > 1 \quad \theta_t = \theta_c$$

$$r = p_c^*/p_0 < 1$$

$$si \quad \mathcal{U}(0,1) > r \quad \theta_t = \theta_c$$

$$si \quad \mathcal{U}(0,1) < r \quad \theta_t = \theta_0$$

Referencias

Excelentes canales de matemáticos y varios

ritvikmath channel
https://www.youtube.com/channel/UCUcpVoi5KkJmnE3bvEhHR0Q

mathematicalmonk channel
https://www.youtube.com/channel/UCcAtD_VYwcYwVbTdvArsm7w

Normalized Nerd channel
https://www.youtube.com/channel/UC7Fs-Fdpe0I8GYg3IboEuXw

Introduction to Bayesian statistics, part 2: MCMC and the Metropolis–Hastings algorithm, StataCorp LLC
https://www.youtube.com/watch?v=OTO1DygELpY

Videos MCMC, Andres Farall
https://www.youtube.com/watch?v=6eCdij07uuE&t=327s

Bayesian Modeling with R and Stan (Reupload),
Salt Lake City R Users Group
https://www.youtube.com/watch?v=QqwCqPYbatA

MCMC in R
https://www.youtube.com/watch?v=AT49XR3G38w

A Markov chain Monte Carlo example in astronomy
http://personal.psu.edu/muh10/MCMCtut/MCMC.html

Penn State Astrostatistics MCMC tutorial Murali Haran. Penn State Dept. of Statistics
http://personal.psu.edu/muh10/MCMCtut/chptmodel2021.pdf