Estadística Bayesiana: Clase 14

Juan Carlos Correa

20 de abril de 2021

Aproximación Bayesiana vía Simulación

Suponga tenemos una población Multinomial con k categorías. El problema lo trabajaremos usando la familia conjugada Dirichlet. Además queremos realizar inferencias acerca de θ, un parámetro que es una función de $(\pi_1,\ \pi_2, \ldots, \pi_k)$, por ejemplo

- $\theta = \frac{\pi_1}{\pi_2}$
- $\theta = \frac{\pi_1 \pi_4}{\pi_2 \pi_3}$
- $\theta = \log(\pi_1) \log(\pi_2)$
- $\theta = (\theta_1, \theta_2) = \left(\frac{\pi_2}{\pi_1},\ \frac{\pi_3}{\pi_1}\right)$
- etc.

Algoritmo

1. Genere π_i de la *Dirichlet* $(\alpha_i + \mathbf{n}_i)$ para $i = 1, 2, \ldots, k$.
2. Haga $\pi' = (\pi'_1, \pi'_2, \ldots, \pi'_k)$
3. Calcule

$$\theta = g\left(\pi'_1,\ \pi'_2, \ldots, \pi'_k\right)$$

Repita los pasos 1 a 3 muchas veces, digamos N. Al final se obtendrá

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}$$

la distribución posterior de θ se aproxima mediante un "histograma" de los valores calculados.

Tabla 2×2

Ejemplo de niños zurdos

La siguiente tabla presenta información sobre niños zurdos y el sexo:

	Zurdo	Diestro
Niño	79	202
Niña	57	138

La pregunta que nos surge es: Hay más niños zurdos que niñas?

Esquema Producto-Binomial

	Éxito	Fracaso
Pob. I	π_1	$1-\pi_1$
Pob. II	π_2	$1-\pi_2$

En esta situación se sacan dos muestras independientes con tamaños predeterminados, digamos n_1 y n_2. Aquí nos interesa comparar las distribuciones de las dos poblaciones binomiales, o sea comparar π_1 con π_2. La función de verosimilitud de los datos será

	Éxito	Fracaso
Pob. I	x_1	n_1-x_1
Pob. II	x_2	n_2-x_2

La verosimilitud será

$$L\left(\pi_1,\pi_2\right) \propto \pi_1^{x_1}\left(1-\pi_1\right)^{n_1-x_1}\pi_2^{x_2}\left(1-\pi_2\right)^{n_2-x_2}$$

Apriori

Problema: Se pueden definir varias apriori.

- Por ejemplo una apriori conjunta para (π_1, π_2), por ejemplo

$$\xi\left(\pi_1, \pi_2\right) \propto \pi_1^{\alpha_1 - 1}\left(1 - \pi_1\right)^{\beta_1 - 1} \pi_2^{\alpha_2 - 1}\left(1 - \pi_2\right)^{\beta_2 - 1}$$

  Este es el caso más sencillo ya que corresponde a la conjugada. La aposteriori será

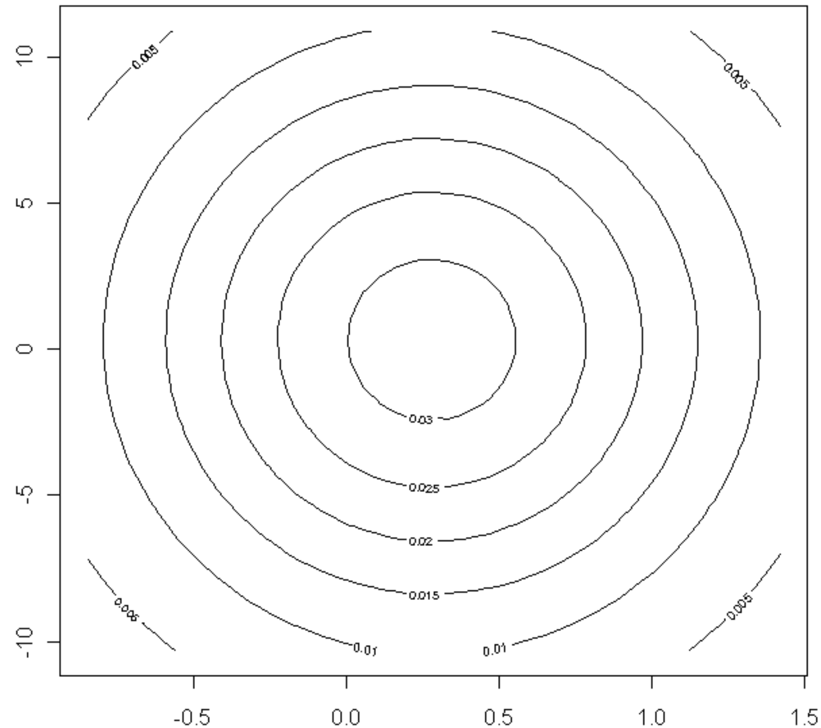
$$\xi\left(\pi_1, \pi_2 \,|\, Datos\right) \propto \pi_1^{\alpha_1 + x_1 - 1}\left(1 - \pi_1\right)^{n_1 - x_1 + \beta_1} \pi_2^{x_2 + \alpha_2 - 1}\left(1 - \pi_2\right)^{n_2 - x_2 + \beta_2 -}$$

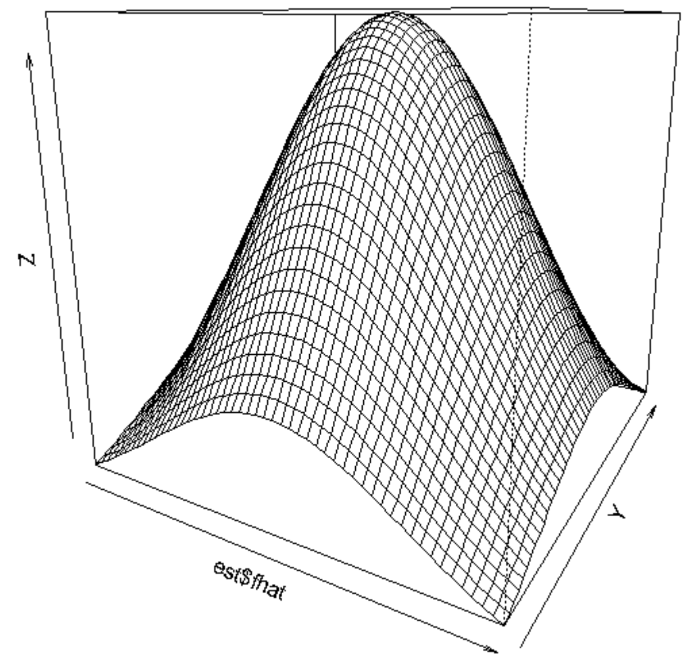
- Una apriori sobre $\pi_1 - \pi_2$
- Una apriori sobre π_1 / π_2

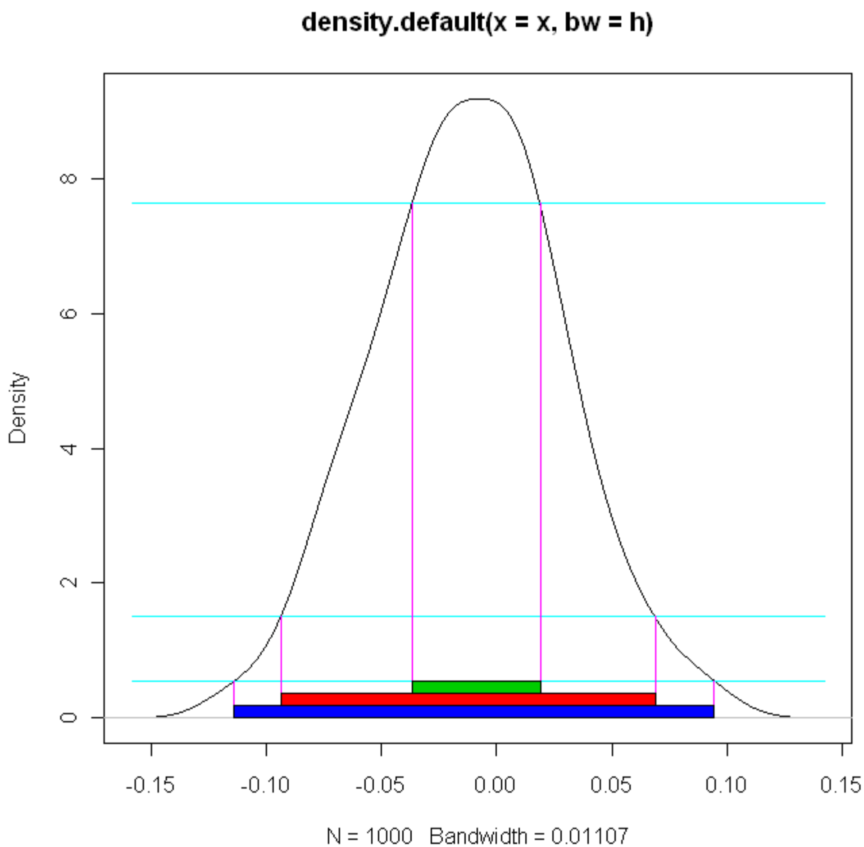
Ejemplo de los niños.

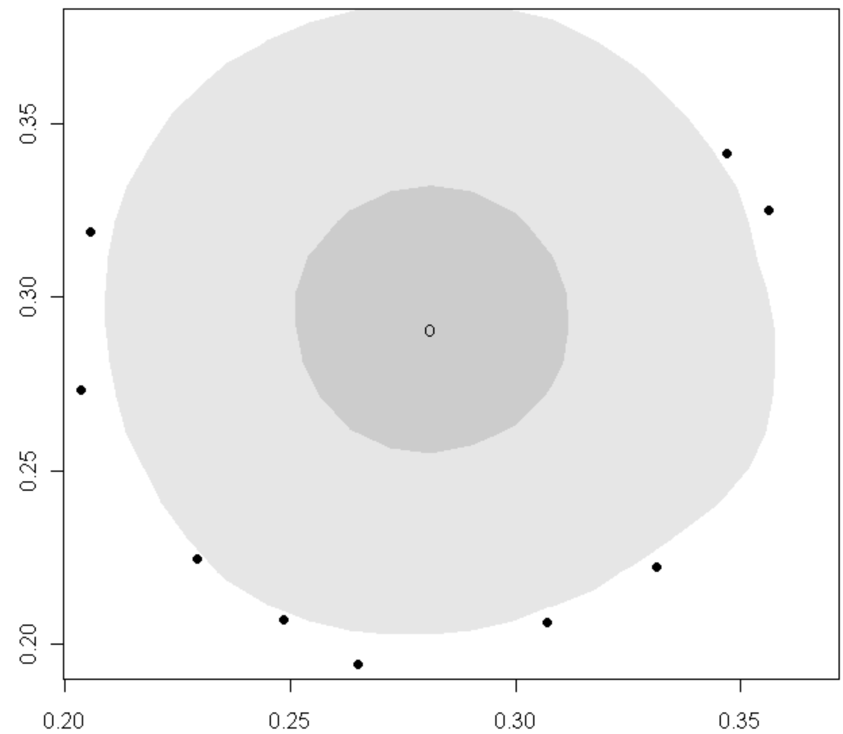
```
# Suponga apriori Beta(1,1)xBeta(1,1)
pi1<-rbeta(1000,79+1,202+1)
pi2<-rbeta(1000,57+1,138+1)

plot(pi1,pi2)
library(KernSmooth)

x <- cbind(pi1, pi2)
est <- bkde2D(x, bandwidth=c(0.7,7))
contour(est$x1, est$x2, est$fhat)
persp(est$fhat)

dif<-pi1-pi2
summary(dif)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.12510 -0.03980 -0.01140 -0.01182  0.01590  0.10900
```

```
library(hdrcde)
hdr.den(dif)
$hdr
             [,1]        [,2]
99% -0.11388344 0.09407083
95% -0.09382573 0.06901195
50% -0.03672022 0.01884642

$mode
[1] -0.00719232

$falpha
       1%         5%        50%
0.5544871 1.5005082 7.6356746

hdr.boxplot.2d(pi1, pi2)
```


0.005
0.005
0.03
0.025
0.02
0.015
0.01
0.01
0.005
0.005
10
5
0
-5
-10
-0.5
0.0
0.5
1.0
1.5


Z
est$fhat
Y


density.default(x = x, bw = h)
Density
0
2
4
6
8
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
N = 1000 Bandwidth = 0.01107


0.20
0.25
0.30
0.35
0.20
0.25
0.30
0.35
o

Esquema Multinomial

	B	B^c
A	π_{11}	π_{12}
A^c	π_{21}	π_{22}

$$\sum_i \sum_j \pi_{ij} = 1$$

En esta situación se sacan una muestra con tamaño predeterminado, digamos N. Aquí nos interesa determinar si las dos variables son dependientes o no. O sea queremos ver si $\pi_{ij} = \pi_{i+}\pi + j$. La función de verosimilitud de los datos será

	B	B^c
A	x_{11}	x_{12}
A^c	x_{21}	x_{22}

La verosimilitud será

$$L\left(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}\right) \propto \pi_{11}^{x_{11}} \pi_{12}^{x_{12}} \pi_{21}^{x_{21}} \pi_{22}^{x_{22}}$$

Bajo el esquema multinomial la distribución apriori no informativa puede seleccionarse de una amplia clase:

- **Apriori Uniforme**

$$\xi\left(\pi_1,\ \pi_2,\ \pi_3,\ \pi_4\right) \propto 1$$

- **Apriori de Jeffreys**

$$\xi\left(\pi_1,\ \pi_2,\ \pi_3,\ \pi_4\right) \propto \pi_1^{-1/2}\pi_2^{-1/2}\pi_3^{-1/2}\pi_4^{-1/2}$$

La razón de odds

La siguiente tabla presenta el modelo poblacional para una tabla 2×2, donde cada celda presenta la probabilidad de ella.

	A	A^c
B	$P(A \cap B)$	$P(A^c \cap B)$
B^c	$P(A \cap B^c)$	$P(A^c \cap B^c)$

Los odds[a] de que el evento B ocurra relativo al evento A se define como la razón de las probabilidades

$$\frac{P\left[B \mid A\right]}{P\left[B^c \mid A\right]}$$

La interpretación de la razón anteriror es directa: Asumiendo que el evento A ha ocurrido, esta razón nos dice cúantas veces ocurre el evento B por cada aparición del evento B^c. Los odds de B relativo a A^c son

$$\frac{P\left[B \mid A^c\right]}{P\left[B^c \mid A^c\right]}$$

[a]La palabra *odds* no tiene una única y precisa traducción, algunos la traducen como disparidad y otros como apuestas.

Cornfield (1951) definió la razón de odds como

$$\psi = \frac{\frac{P[B|A]}{P[B^c|A]}}{\frac{P[B|A^c]}{P[B^c|A^c]}}$$

- La interpretación es clara.
- Valores de ψ que se alejen de 1.0 en una dirección particular representa una asoción fuerte.
- Dos valores de ψ pueden representar un mismo nivel de asociación (un valor y su inverso) pero en direcciones opuestas.
- Valores menores que uno indican una asociación negativa, mientras valores mayores que 1 indican una asociación positiva.
- Para simetrizar esta medida se trabaja con el $log(\psi)$.

Tabla 2×2

Ejemplo de habilidad manual de esposos

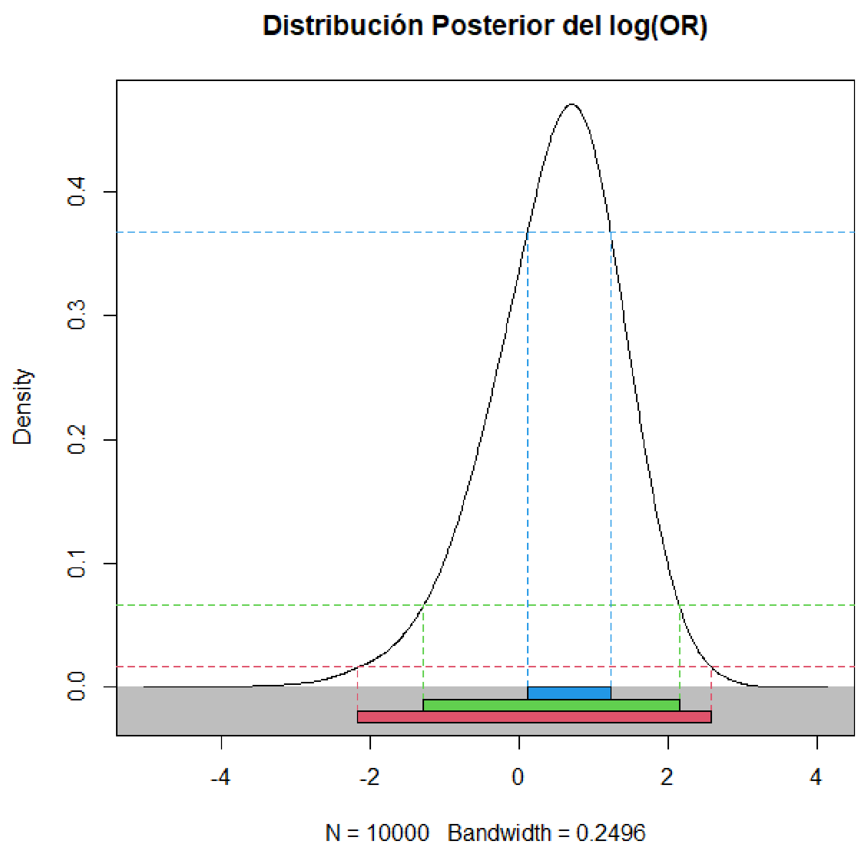
La siguiente tabla presenta información sobre la destereza manual del esposo y la esposa:

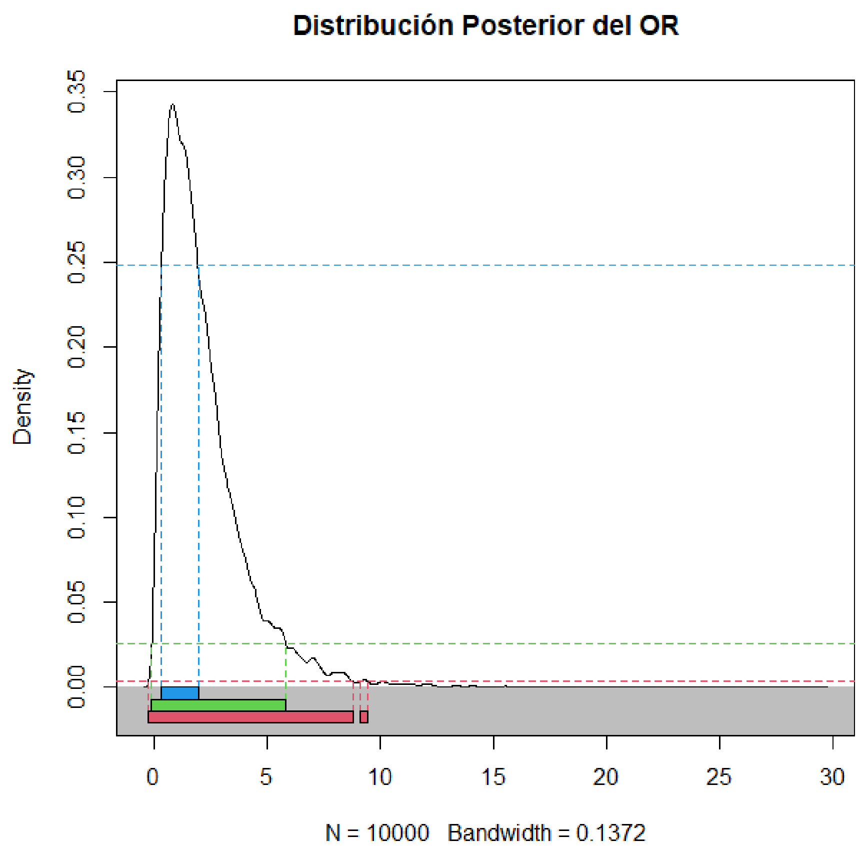
	Esposa	
Esposo	Zurda	Diestra
Zurdo	1	13
Diestro	20	284

La pregunta que nos surge es: hay independencia entre la habilidad manual del esposo y la de la esposa?

```
library(MCMCpack)
res<-rdirichlet(10000,c(1+1,13+1,20+1,284+1))

log.OR<-log(res[,1]*res[,4]/(res[,2]*res[,3]))
> summary(log.OR)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-4.28776 -0.04726  0.55672  0.47431  1.07621  3.37960
> summary(exp(log.OR))
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 0.01374  0.95384  1.74494  2.25216  2.93354 29.35900

> hdr.den(log.OR,main='Distribución Posterior del log(OR)')
$hdr
          [,1]     [,2]
99% -2.1751819 2.574805
95% -1.2844725 2.138309
50%  0.1082924 1.217970

$mode
[1] 0.7007231

$falpha
        1%          5%         50%
0.01613558 0.06673433 0.36814598
```

```
> hdr.den(exp(log.OR),main='Distribución Posterior del OR')
$hdr
           [,1]     [,2]    [,3]     [,4]
99% -0.22562515 8.800509 9.15823 9.465179
95% -0.08071155 5.863276      NA       NA
50%  0.35235364 1.960321      NA       NA

$mode
[1] 0.8342001

$falpha
         1%          5%         50%
0.003528188 0.025444277 0.247903836
```


Distribución Posterior del log(OR)
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
N = 10000 Bandwidth = 0.2496


Distribución Posterior del OR
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
5
10
15
20
25
30
N = 10000 Bandwidth = 0.1372

Página www

Página de Abertura

Contenido

Página 22 de 54

Regresar

Full Screen

Cerrar

Abandonar

MCMC: Monte Carlo por Cadenas de Markov

Los métodos MCMC son algoritmos iterativos que se utilizan cuando el muestreo directo de una distribución de interés ξ no es factible.
Una cadena de Markov es generada muestreando

$$\theta^{(t+1)} \sim p\left(\theta|\theta^{(t)}\right)$$

Este p es llamado el *kernel de transición* de la cadena de Markov. Así $\theta^{(t+1)}$ depende solo de $\theta^{(t)}$, y no de $\theta^{(0)}, \theta^{(1)}, \cdots, \theta^{(t-1)}$

Glosario de Cadenas de Markov

Irreducibilidad Una cadena de Markov $X_1, X_2, \cdots$ es irreducible si la cadena puede moverse libremente a través del espacio de estados; esto es, para dos estados cualesquiera x y x', existe un n tal que

$$P\left(X_n = x' | X_0 = x\right) > 0.$$

Recurrencia Una cadena de Markov es recurrente si el número promedio de visitas a un estado arbitrario es infinito.

Período Un estado x tiene período d si $P\left(X_{n+t} = x | X_t = x\right) = 0$ si n no es divisible por d, donde d es el mayor entero con esta propiedad.

Aperiodicidad Si un estado x tiene período $d = 1$ se dice que es aperiódico.

En una cadena irreducible todos los estados tienen el mismo período. Si ese período es $d = 1$, la cadena de Markov es aperiódica.

Convergencia a una Distribución Estacionaria

Si una cadena de Markov con espacio de estados contable $X_1, X_2, \cdots$ *es positiva, recurrente y aperiódica con distribución estacionaria* π*, entonces desde cualquier estado inicial*

$$X_n \to X \sim \pi$$

Página www

Página de Abertura

Contenido

Página 25 de 54

Regresar

Full Screen

Cerrar

Abandonar

Ergodicidad Una cadena de Markov positiva, recurrente y aperiódica es llamada ergódica.

Convergencia de Sumas (Teorema Ergódico)

Si una cadena de Markov con espacio de estados contable $X_1, X_2, \cdots$ *es ergódica con distribución esatcionaria* π, *entonces desde cualquier estado inicial*

$$\frac{1}{n}\sum_{i=1}^{n} h\left(X_i\right) \to E_\pi\left[h(X)\right]$$

Algoritmo Metropolis-Hastings

- El muestreo de importancia y el muestreo de rechazo trabajan bien si la densidad propuesta $q(\theta)$ es similar a $p(\theta)$.
- En problemas complejos puede ser difícil crear una única $q(\theta)$ que tenga esta propiedad.
- El algoritmo Metropolis utiliza una densidad propuesta q que depende del estado actual de $\theta^{(t)}$.
- La densidad $q\left(\theta' | \theta^{(t)}\right)$ puede ser tan simple como una normal localizada en $\theta^{(t)}$ y no es necesario que se parezca a $p(\theta)$.

El algoritmo se resume así:

1. Comience en cualquier lugar, y digamos que estamos en $\theta^{(t)} = \theta$.

2. Genere θ^* de $q(\theta^*|\theta)$. θ^* es llamado un *punto candidato* y q es llamada una *distribución propuesta*.

3. Calcule

$$\alpha(\theta, \theta^*) = \text{mín}\left\{1, \frac{\xi(\theta^*|Datos)\, q(\theta|\theta^*)}{\xi(\theta|Datos)\, q(\theta^*|\theta)}\right\}$$

4. Acepte $\theta^{(t+1)} = \theta^*$ con probabilidad $\alpha(\theta, \theta^*)$.

5. En otro caso $\theta^{(t+1)} = \theta$