

CAPÍTULO 8. MÉTODOS NUMÉRICOS

Para leer:

Gelman et al Capítulos 9 y 10.

Dada una distribución a priori $f(\boldsymbol{\theta})$ y los datos \mathbf{x} , a menudo se quiere evaluar:

- la densidad a posteriori

$$f(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x})}{\int f(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}}$$

- dada una pérdida $L(\mathbf{T}, \boldsymbol{\theta})$, el estimador de Bayes $\min_{\mathbf{T}} E[L(\mathbf{T}, \boldsymbol{\theta})|\mathbf{x}]$
- intervalos de credibilidad, ...

Excepto las pocas situaciones en que tenemos familias conjugadas, la computación de las integrales es problemática. Se necesitan métodos numéricos.

Ejemplo 71 Sea $X|\alpha \sim \mathcal{B}(\alpha, \beta)$ donde β es conocida. Entonces existe una distribución conjugada

$$f(\alpha) \propto \left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \right)^a b^\alpha$$

pero se necesitan métodos numéricos para calcular la constante, los momentos etc.

Ejemplo 72 $X|\theta \sim \mathcal{C}(1, \theta)$. Dada una distribución a priori no-informativa, $f(\theta) \propto 1$, la distribución final es

$$f(\theta|\mathbf{x}) \propto \prod [1 + (x_i - \theta)^2]^{-1}$$

y no se puede calcular ni la constante de integración ni los momentos sin métodos aproximados.

Ejemplo 73 *Se supone que*

$$f(x|\mathbf{w}, \boldsymbol{\theta}) = \sum_{i=1}^k w_i g(x|\theta_i)$$

una mixtura de densidades.

Dados los datos, la verosimilitud será

$$l(\mathbf{w}, \boldsymbol{\theta}|\mathbf{x}) \propto \prod_{j=1}^n \sum_{i=1}^k w_i g(x_j|\theta_i)$$

que tiene k^n términos.

Para n grande, no se puede evaluar la distribución final de $(\mathbf{w}, \boldsymbol{\theta})$ en un tiempo finito.

Si Θ es discreto (e infinito), de vez en cuando hay menos dificultades en calcular la densidad. Se trunca la suma para estimar la constante de la distribución.

$$P(\theta|\mathbf{x}) \approx \frac{P(\theta)l(\theta|\mathbf{x})}{\sum_i^N P(\theta = i)l(\theta = i|\mathbf{x})}$$

donde N es un número grande. Utilizando valores diferentes de N se puede determinar una aproximación suficientemente precisa.

Ejemplo 74 Sea $X|N \sim \mathcal{BI}(N, 1/2)$ y la distribución a priori geométrica $N - 1 \sim \mathcal{GE}(1/4)$:

$$P(N = n) = \frac{1}{4} \left(\frac{3}{4}\right)^{n-1} \quad n = 1, 2, \dots$$

Se quiere estimar la constante de la distribución a posteriori de N y la media a posteriori.

Dada una observación $x = 5$, la distribución a posteriori de $N|x$ es

$$\begin{aligned}
 P(N = n|x) &\propto \frac{1}{4} \left(\frac{3}{4}\right)^{n-1} \binom{n}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{n-5} \\
 &\propto \frac{n!}{(n-5)!} \left(\frac{3}{8}\right)^n \quad \text{para } n \geq 5 \\
 &= \frac{\frac{n!}{(n-5)!} \left(\frac{3}{8}\right)^n}{\sum_{m=5}^{\infty} \frac{m!}{(m-5)!} \left(\frac{3}{8}\right)^m}
 \end{aligned}$$

Se aproxima el denominador truncado la suma. En la siguiente tabla, se muestran los valores de $p = \frac{n!}{(n-5)!} \frac{3^n}{8}$ y np para $n = 5, 6, \dots$

n	p	np
5	,8898926	4,45
6	2,002258	12,01
7	2,627964	18,40
8	2,627964	21,02
9	2,217345	19,96
10	1,663008	16,63
11	1,143318	12,58
⋮	⋮	⋮
34	,0000001	,00
35	,0000000	,00
<i>Total</i>	14,93	128,4

Entonces, la densidad a posteriori de N es aproximadamente $P(N = n|x) = \frac{1}{14,93} \frac{n!}{(n-5)!} \left(\frac{3}{8}\right)^n$ para $n > 5$ y la media a posteriori es

$$E[N|x] \approx \frac{128,4}{14,93} \approx 8,8.$$

Si se quiere una estimación más precisa, se trunca la suma en un valor más grande de n .

Integración Numérica

Hay muchos métodos. Aquí consideramos sólo el más sencillo, **la regla de Simpson**.

Queremos evaluar la integral

$$I = \int_a^b g(\theta) d\theta$$

En su versión más simple, la regla de Simpson dice que

$$I \approx \frac{b-a}{6} \left[g(a) + 4g\left(\frac{a+b}{2}\right) + g(b) \right].$$

Se mejora la aproximación dividiendo el intervalo $[a, b]$ en un número par N de subintervalos

$$[a, a+h) \cup \dots \cup [a+(N-1)h, a+Nh = b].$$

Utilizando la regla de Simpson para cada subintervalo $[a + jh, a + (j + 2)h)$ nos proporciona el estimador

$$I \approx \frac{h}{3} [g(a) + 4g(a + h) + 2g(a + 2h) + \dots + 2g(a + (N - 2)h) + 4g(a + (N - 1)h) + g(a + Nh)]$$

Ejemplo 75 Sea $\theta|\mathbf{x} \sim \mathcal{B}(7, 10)$. Entonces

$$f(\theta|\mathbf{x}) \propto \theta^6(1 - \theta)^9$$

Intentamos estimar

$$B(7, 10) = \int_0^1 \theta^6(1 - \theta)^9 d\theta$$

Usando la regla de Simpson con $h = 0,1$ obtenemos la siguiente tabla:

θ	$\theta^6(1 - \theta)^9$	$\int_0^\theta \phi^6(1 - \phi)^9 d\phi$
,0	0,00000E - 00	0,00000E - 00
,1	3,87420E - 07	
,2	8,58994E - 06	3,37987E - 07
,3	2,94178E - 05	
,4	4,12783E - 05	5,92263E - 06
,5	3,05176E - 05	
,6	1,22306E - 05	1,17753E - 05
,7	2,31568E - 06	
,8	1,34218E - 07	1,24962E - 05
,9	5,31441E - 10	
1,0	0,00000E - 00	1,25007E - 05

El valor exacto de la integral es $1,24875E - 05$. (La regla de Simpson con $h = 0,05$ nos lleva el resultado $1,24876E - 05$.)

Añadiendo más columnas a la tabla, se puede estimar la media o los otros momentos.

Ejemplo 76 *Se muestrean 5 observaciones de una distribución Cauchy con densidad*

$$f(x|\theta) = \frac{1}{\pi \{1 + (x - \theta)^2\}}.$$

Los datos son $\mathbf{x} = (11,4, 7,3, 9,8, 13,7, 10,6)^T$. Se supone una distribución a priori uniforme para θ .

Supongamos que queremos estimar la constante de integración, $P(\theta < 11,5|\mathbf{x})$ y la media a posteriori de θ .

La densidad a posteriori es

$$\begin{aligned} f(\theta|\mathbf{x}) &\propto [1 + (11,4 - \theta)^2] \dots [1 + (10,6 - \theta)^2] \\ &\propto H(\theta) \quad \text{donde} \\ H(\theta) &= 10^5 [1 + (11,4 - \theta)^2] \dots [1 + (10,6 - \theta)^2]. \end{aligned}$$

Multiplicamos por la constante (10^5) para disminuir los problemas numéricos.

Para calcular la constante, necesitamos estimar $\int_{-\infty}^{\infty} H(\theta) d\theta$. Utilizamos la regla de Simpson con $h = 0,5$.

θ	$H(\theta)$	$\theta H(\theta)$	$\int_{-\infty}^{\theta} H(t) dt$	$F(\theta \mathbf{x})$	$\int_{-\infty}^{\theta} tH(t) dt$
4,5	,0	,0	,0	,000	,0
5,0	,0	,0			
5,5	,0	,1	,0	,000	,0
6,0	,1	,4			
6,5	,2	1,4	,1	,000	0,5
7,0	,8	5,6			
7,5	2,3	16,9	1,0	,002	7,3
8,0	4,8	38,8			
8,5	10,7	90,7	6,4	,012	51,1
9,0	28,2	253,9			
9,5	82,7	785,9	40,8	,076	366,4
10,0	196,1	1961,4			
10,5	290,6	3051,2	233,8	,436	2313,5
11,0	250,1	2751,1			
11,5	129,2	1485,6	470,5	,877	4903,8
12,0	47,4	568,2			
12,5	17,3	216,3	526,4	,982	5566,3
13,0	7,4	96,3			
13,5	3,3	44,0	534,8	,997	5673,9
14,0	1,1	15,4			
14,5	,3	4,2	536,1	1,000	5692,2
15,0	,1	1,2			
15,5	,0	,4	536,2	1,000	5693,7
16,0	,0	,1			
16,5	,0	,0	536,3	1,000	5693,8

Utilizamos la aproximación

$$\int_{-\infty}^{\infty} H(\theta) d\theta \approx \int_{4,5}^{16,5} H(\theta) d\theta$$

porque la densidad es aproximadamente cero fuera del intervalo [4,5, 16,5].

$$P(\theta < 11,5|\mathbf{x}) \approx 470,5/536,3 = ,877.$$

Se estima la media a posteriori con

$$E[\theta|\mathbf{x}] \approx 5693,8/536,3 \approx 10,6$$

Existen otros métodos de integración numérica que son más precisos, por ejemplo, cuadratura de Gauss, wavelets, etc.

Métodos Monte-Carlo

Se quiere evaluar la integral

$$E[g(\theta)|\mathbf{x}] = \int g(\theta)f(\theta|\mathbf{x}) d\theta$$

Si podemos muestrear $f(\theta|\mathbf{x})$, podemos estimar $E[g(\theta)|\mathbf{x}]$ con

$$\hat{g} = \frac{1}{N} \sum_{i=1}^N g(\theta^{(i)})$$

donde $\theta^{(1)}, \dots, \theta^{(N)}$ es una muestra. Observemos que

- cuando $N \rightarrow \infty$, $\hat{g} \rightarrow E[g(\theta)|\mathbf{x}]$
- usando el teorema central del límite, un intervalo de confianza para $E[g(\theta)|\mathbf{x}]$ es aproximadamente

$$\hat{g} \pm 2\hat{se}(\hat{g})$$

Ejemplo 77 $X|\mu, \phi \sim \mathcal{N}(\mu, 1/\phi)$ y $f(\mu, \phi) \propto 1/\phi$. Entonces

$$\begin{aligned}\phi|\mathbf{x} &\sim \mathcal{G}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) \\ \mu|\phi, \mathbf{x} &\sim \mathcal{N}(\bar{x}, 1/(n\phi))\end{aligned}$$

Se puede muestrear $f(\mu, \phi|\mathbf{x})$ utilizando el algoritmo

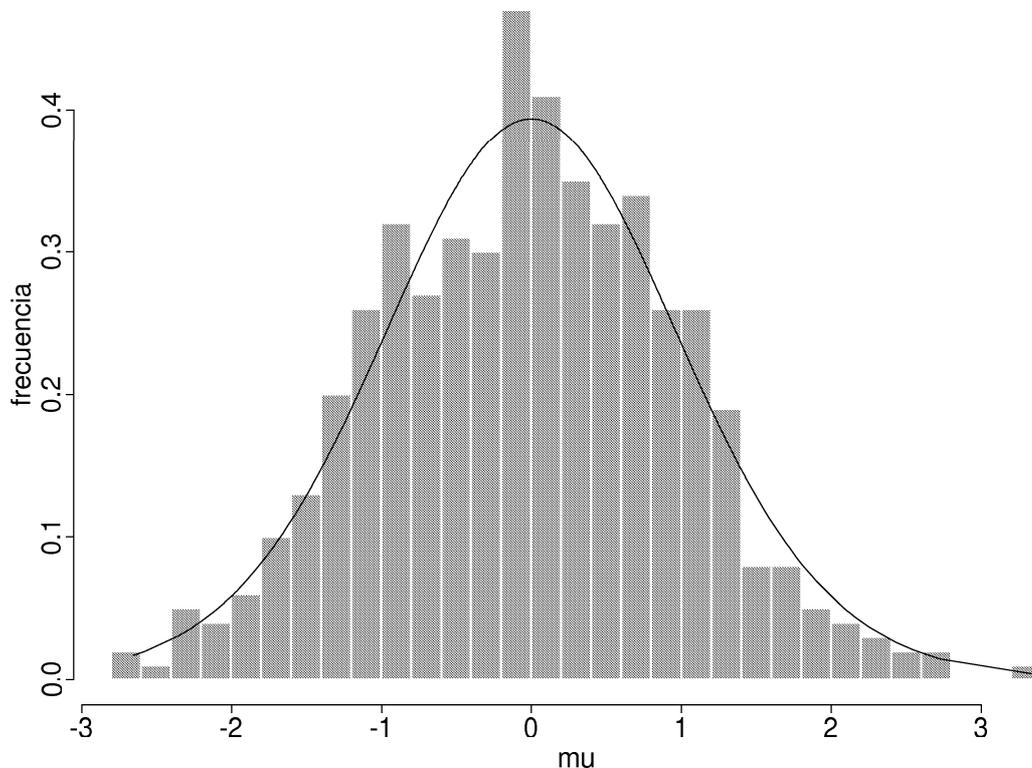
1. $i = 1$
2. Generar $\phi^{(i)}$ de $\mathcal{G}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$,
3. Generar $\mu^{(i)}$ de $\mathcal{N}(\bar{x}, 1/(\phi^{(i)}n))$
4. $i = i + 1$. Si $i \leq N$, ir a 2.

Ejemplo 78 *Supongamos que queremos calcular un intervalo de credibilidad a posteriori para μ , donde $X|\mu, \phi \sim \mathcal{N}(\mu, 1/\phi)$ y utilizamos la distribución a priori de Jeffreys. Dados los datos $n = 20$, $\bar{x} = 0$ y $s^2 = 20$ tenemos las distribuciones a posteriori*

$$\begin{aligned}\phi|\mathbf{x} &\sim \mathcal{G}\left(\frac{19}{2}, \frac{19 \times 20}{2}\right) \\ \mu|\phi, \mathbf{x} &\sim \mathcal{N}(0, 1/(20\phi))\end{aligned}$$

Hacemos una muestra Monte-Carlo de tamaño 500 de la distribución a posteriori.

El diagrama muestra un histograma de los datos. El histograma es similar a la densidad t_{19} que es la verdadera densidad a posteriori de μ .



Estimamos un intervalo de credibilidad de 90 % utilizando los cuantiles de la muestra. El 5 % cuantil es $-1,67$ y el 95 % cuantil es $1,63$. El verdadero 95 % cuantil de t_{19} es $1,73$

La media muestral es $,022$ que es bastante parecida a la verdadera media (0).

Se mejora la aproximación utilizando muestras Monte-Carlo más grandes.

La función de importancia

Son poco comunes las situaciones en que se puede muestrear la distribución a posteriori fácilmente. Normalmente, no se conoce la constante de integración. En situaciones así, se modifica el método de Monte Carlo.

Pongamos una distribución bien conocida $h(\theta)$, **la función de importancia**. Entonces tenemos

$$E[g(\theta|\mathbf{x})] = \frac{\int w(\theta)g(\theta)h(\theta) d\theta}{\int w(\theta)h(\theta) d\theta}$$

donde el peso

$$w(\theta) = \frac{f(\theta)l(\theta|x)}{h(\theta)}.$$

Dados unos datos muestreados de $h(\theta)$, podemos aproximar la esperanza con

$$E[g(\theta|\mathbf{x})] \approx \frac{\sum_{i=1}^N w_i g(\theta^{(i)})}{\sum_{i=1}^N w_i}$$

donde

$$w_i = \frac{f(\theta^{(i)})l(\theta^{(i)}|x)}{h(\theta^{(i)})}.$$

Observamos que los pesos w_i dependen de la constante desconocida.

Observación 28 *En teoría, se puede utilizar 'cualquier' función de importancia $h(\theta)$ pero si $h(\theta)$ es muy diferente a $f(\theta|\mathbf{x})$ necesitaremos muestras muy grandes. Por tanto el método funciona mejor si $h(\theta)$ está próxima a $f(\theta|\mathbf{x})$. Algunas elecciones posibles son $h(\theta) = f(\theta)$ o (mejor) $h(\theta) \propto l(\theta|\mathbf{x})$.*

Si $h(\theta) = f(\theta|\mathbf{x})$ tenemos el método simple Monte-Carlo.

Ejemplo 79 *Volvemos al Ejemplo 75. Tenemos*

$$f(\theta|\mathbf{x}) \propto \theta^6(1 - \theta)^9$$

para $0 < \theta < 1$. Supongamos que queremos estimar la constante de distribución y la media.

Usamos una muestra de importancia, muestreando una distribución uniforme.

1. Para $i = 1, \dots, N$.

a) Muestrear $\theta_i \sim U(0, 1)$.

b) Calcular $w_i = \theta_i^6(1 - \theta_i)^9$.

c) Calcular $h(\theta_i) = \theta_i w_i$

2. Estimar $\int_0^1 \theta^6(1 - \theta)^9 \approx \frac{1}{N} \sum_{i=1}^N w_i$.

3. Estimar $E[\theta|\mathbf{x}] \approx \frac{\sum_{i=1}^N \theta_i w_i}{\sum_{i=1}^N w_i}$

La siguiente tabla proporciona los valores estimados de la constante de distribución y la media para varias muestras de importancia.

N	$\int \theta^6(1 - \theta)^9$	$E[\theta \mathbf{x}]$
100	,000144	,4215
500	,000126	,4020
1000	,000121	,4059
5000	,000126	,4094
10000	,000125	,4100

La verdadera media a posteriori es $\frac{7}{17} \approx ,4118$.

El método de rechazo

La idea es muestrear una distribución $h(\theta)$ (una función envolvente) con la propiedad de que

$$f(\theta)l(\theta|\mathbf{x}) < Mh(\theta)$$

para una constante $M > 0$. El algoritmo es

1. $i = 1$,
2. Generar $\theta^{(i)}$ de $h(\cdot)$,
3. Generar $U \sim \mathcal{U}(0, 1)$,
4. Si $MUh(\theta^{(i)}) < f(\theta^{(i)})l(\theta^{(i)}|\mathbf{x})$ aceptar $\theta^{(i)}$ y si no, rechazarlo,
5. Continuar hasta que la muestra sea completa.

Observación 29 $h(\cdot)$ debe ser parecida a $f(\theta|\mathbf{x})$, pero con colas más largas. El método funciona mejor cuando M es pequeño. Por tanto si $h(\theta) = f(\theta|\mathbf{x})$ tenemos el método Monte-Carlo.

La desventaja del método es que puede ser difícil elegir $h(\cdot)$ y M .

Otros métodos

- Bootstrap
- El algoritmo EM para estimar la moda de la distribución final
- Métodos MCMC

CAPÍTULO 10. MCMC Y EL MUESTREO GIBBS

Para leer

Lee. Capítulo 9, Secciones 9.4–9.5.

Hasta ahora hemos visto métodos de Monte-Carlo para muestrear $f(\boldsymbol{\theta}|\mathbf{x})$. Ahora, consideramos otro método relacionado. Se necesita la idea de una cadena de Markov.

Definición 14 *Un proceso estocástico X_t es una cadena de Markov si $f(x_t|x_1, \dots, x_{t-1}) = f(x_t|x_{t-1})$. La distribución estacionaria es*

$$f(x) = \int f(y)f(X_t = x|X_{t-1} = y) dy$$

Supongamos que tenemos una cadena de Markov cuya distribución estacionaria es $f(\theta|\mathbf{x})$. Entonces, se muestrea la cadena y después de un tiempo suficiente, la muestra aproximará a una muestra de la distribución a posteriori.

Podemos utilizar los datos de la muestra para estimar las cantidades relevantes como se hace en cualquiera muestra Monte-Carlo.

El problema es cómo construir la cadena de Markov.

El algoritmo Metropolis-Hastings

Este es el algoritmo más general. Supongamos que sabemos la densidad a posteriori $f(\boldsymbol{\theta}|\mathbf{x})$ salvo por la constante:

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto l(\boldsymbol{\theta}|\mathbf{x})f(\boldsymbol{\theta}).$$

Se define una densidad condicional $g(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ (*proposal distribution*) que es fácil muestrear. El algoritmo Metropolis-Hastings es la siguiente:

1. Definir un valor inicial $\boldsymbol{\theta}^{(0)}$,

2. $t = 0$,

3. Generar $\phi \sim g(\phi|\boldsymbol{\theta}^{(t)})$,

4. Definir

$$\alpha(\boldsymbol{\theta}^{(t)}, \phi) = \min \left[1, \frac{f(\phi|\mathbf{x})g(\boldsymbol{\theta}^{(t)}|\phi)}{f(\boldsymbol{\theta}^{(t)}|\mathbf{x})g(\phi|\boldsymbol{\theta}^{(t)})} \right]$$

5. Tomar

$$\boldsymbol{\theta}^{(t+1)} = \begin{cases} \phi & \text{con probabilidad } \alpha \\ \boldsymbol{\theta}^{(t)} & \text{en caso contrario} \end{cases}$$

6. $t = t + 1$. Goto 3.

Se puede demostrar que este algoritmo simula una muestra de una cadena de Markov con distribución estacionaria $f(\boldsymbol{\theta}|\mathbf{x})$.

Demostración parcial del algoritmo.

Las probabilidades de transición son

$$P(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}) = g(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)})\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)}) + \\ I_{\boldsymbol{\theta}^{(t+1)}=\boldsymbol{\theta}^{(t)}} \left[1 - \int g(\boldsymbol{\phi}|\boldsymbol{\theta}^{(t)})\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}) d\boldsymbol{\phi} \right]$$

y usando el resultado que

$$\frac{f(\boldsymbol{\theta}^{(t)}|\mathbf{x})g(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)})\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)})}{f(\boldsymbol{\theta}^{(t+1)}|\mathbf{x})g(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t+1)})\alpha(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\theta}^{(t)})} = 1$$

tenemos la siguiente ecuación ("balance equation")

$$f(\boldsymbol{\theta}^{(t)}|\mathbf{x})P(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}) = f(\boldsymbol{\theta}^{(t+1)}|\mathbf{x})P(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t+1)})$$

y integrando con respecto a $\boldsymbol{\theta}^{(t)}$ nos lleva

$$\int f(\boldsymbol{\theta}^{(t)}|\mathbf{x})P(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}) d\boldsymbol{\theta}^{(t)} = f(\boldsymbol{\theta}^{(t+1)}|\mathbf{x})$$

Observación 30 *En teoría, se puede utilizar cualquier distribución $g(\phi|\theta)$. Lo más importante es que es fácil muestrearla. La convergencia del algoritmo será más rápida si $g \approx f$.*

En el caso particular de que $g(\cdot)$ guarde proporciones con la distribución a posteriori, tenemos $\alpha = 1$ y el método es igual al método de Monte-Carlo simple.

¿Como se decide si hay convergencia?

- Comparación de los resultados con valores iniciales diferentes,
- Gráficos de los valores $\theta^{(t)}$,
- Gráficos de la media estimada (running mean") de θ ,
- Diagnósticos formales de convergencia.

Habitualmente, se usa la primera parte de los datos generados por la cadena de Markov como datos de "burn-in" para que la cadena olvide su estado inicial. Entonces, sólo se utiliza la ultima parte de los datos generados para estimación.

Se estima (por ejemplo) la media a posteriori de θ usando

$$E[\theta|\mathbf{x}] \approx \frac{1}{N - M} \sum_{t=M+1}^{M+N} \theta^{(t)}$$

donde M es el número de iteraciones de burn-in y N es el número total de iteraciones.

La elección de los valores iniciales no es muy importante. La convergencia puede ser más rápida si $\theta^{(0)}$ aproxima a la moda de la distribución final.

Para estimar la forma de la distribución final, se usan métodos de estimar el núcleo de la densidad, dados los datos muestreados (suavizado del histograma, etc.)

El algoritmo con partición

Puede ser difícil muestrear el vector entero θ a la vez. En muchos casos es más fácil dividir el vector $\theta = (\theta_1, \dots, \theta_k)$ en bloques y muestrear las partes.

Se define $\theta_{-i}^t = (\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_k^{(t)})$ el vector después de que los primeros $i - 1$ componentes hayan sido modificados. Luego, si $g_i(\phi | \theta_i, \theta_{-i})$ es una distribución de proposición, se acepta el valor $\theta_i^{(t+1)} = \phi$ con probabilidad

$$\alpha = \min \left[1, \frac{f(\phi | \theta_{-i}^{(t)}, \mathbf{x}) g_i(\theta_i | \phi, \theta_{-i})}{f(\theta_i | \theta_{-i}^{(t)}, \mathbf{x}) g_i(\phi | \theta_i, \theta_{-i})} \right]$$

Los otros componentes no cambian en esta etapa.

La distribución $f(\theta_i | \theta_{-i}, \mathbf{x})$ es la distribución condicional a posteriori de θ_i .

El algoritmo tiene la ventaja que con una separación de los parámetros, se simplifica el cálculo de las probabilidades de aceptación. En muchas situaciones, es bastante fácil evaluar las distribuciones conjugadas.

El algoritmo Metropolis

Si usamos $g(\phi|\theta) = g(\theta|\phi)$, una distribución simétrica, tenemos

$$\alpha = \min \left[1, \frac{f(\phi|\mathbf{x})}{f(\theta|\mathbf{x})} \right]$$

Ejemplo 80 Supongamos $X|\theta \sim \mathcal{C}(\theta, 1)$ y una distribución a priori no informativa $f(\theta) \propto 1$. Queremos simular una muestra de $f(\theta|\mathbf{x})$. Podemos usar el algoritmo Metropolis con, por ejemplo, $g(\phi|\theta) = \mathcal{N}(\theta, s^2)$.

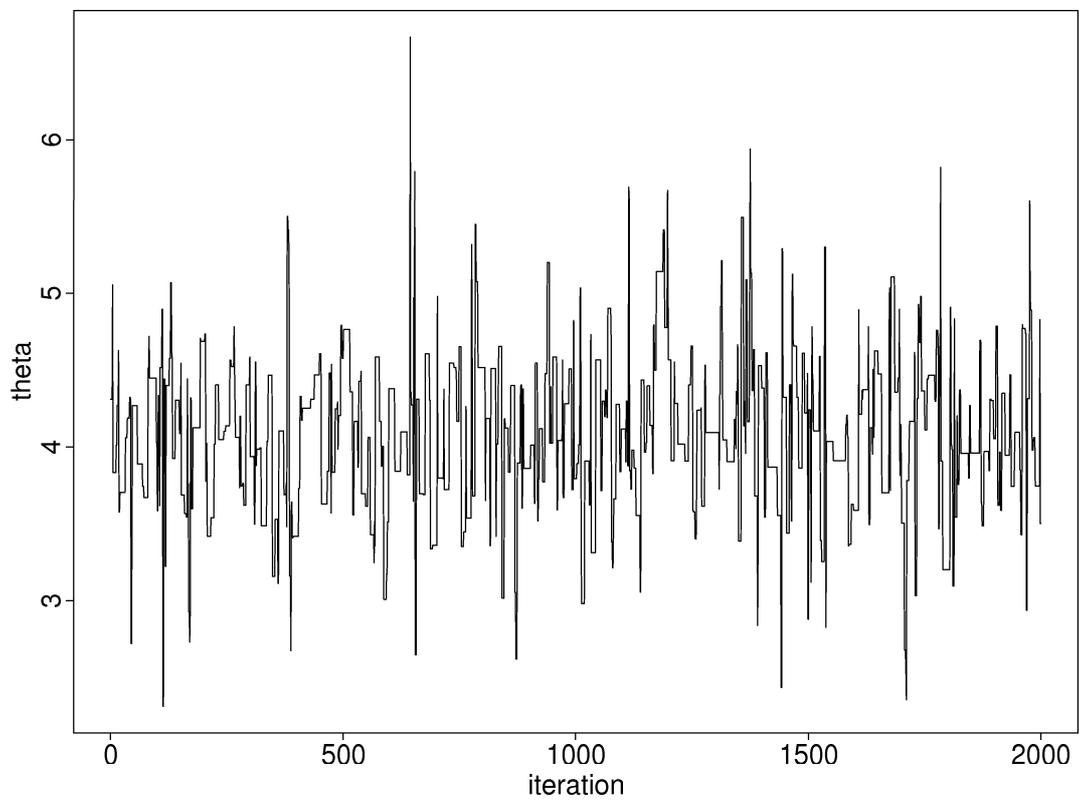
La probabilidad de aceptar un valor ϕ dado el valor corriente $\theta^{(t)}$ es

$$\alpha = \prod_{i=1}^n \frac{1 + (x_i - \theta^{(t)})^2}{1 + (x_i - \phi)^2}$$

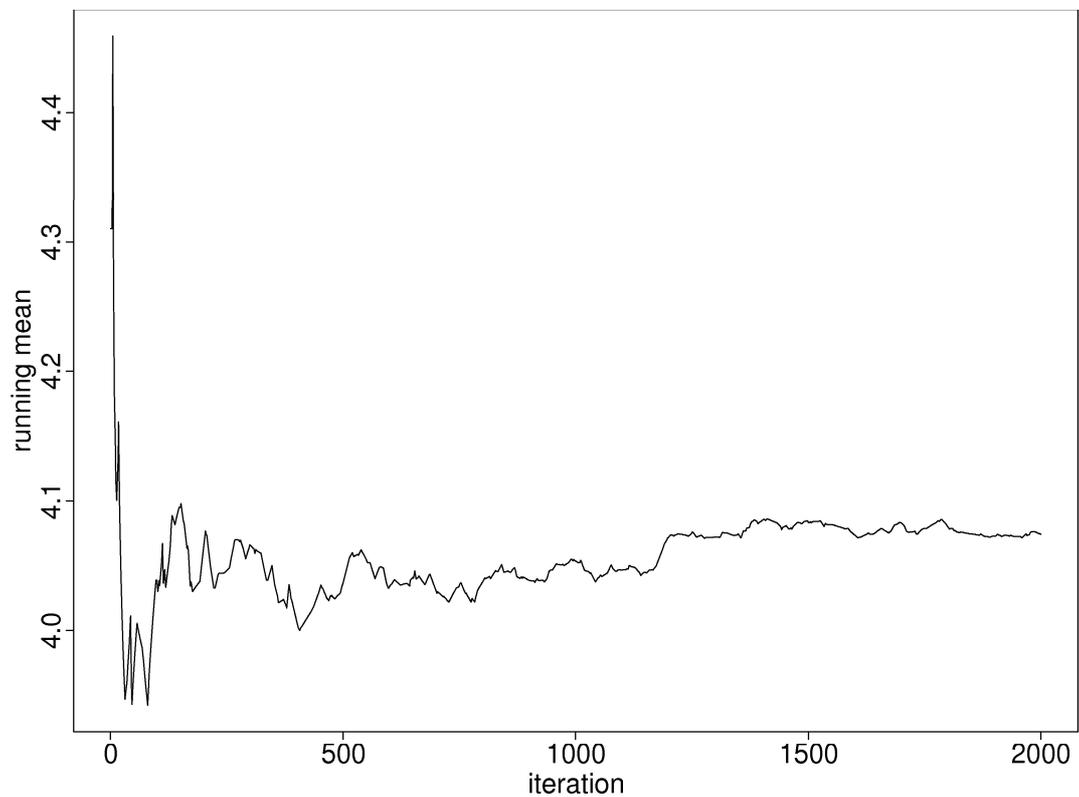
El diagrama ilustra los valores de θ generados en una muestra de 2000 iteraciones, dados los datos

$\mathbf{x} = (4, 3, 2, 2, 3, 1, 8, 4, -1, 2, 6, 7, 4, 4, 7, 3, 4, 1, 3, 8)$

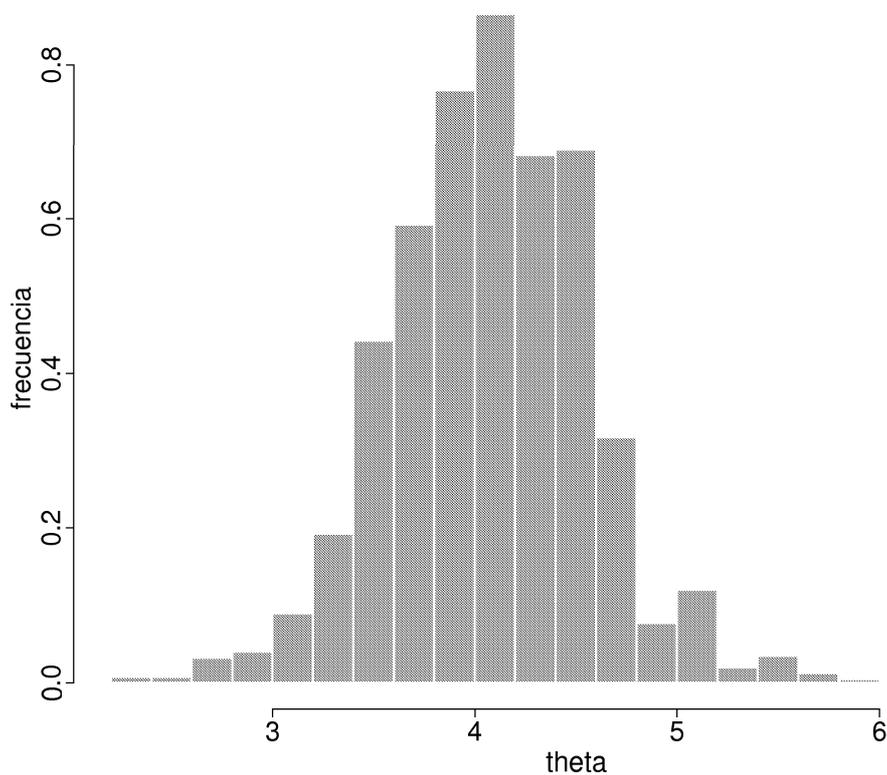
y usando el valor $\theta^{(0)} = \bar{x}$ para inicializar el algoritmo.



El segundo diagrama muestra la media estimada de θ frente el número de iteraciones. Podemos ver que la media se ha estabilizado después de (aproximadamente) 1200 iteraciones.



El tercer diagrama ilustra un histograma de los datos generados por el algoritmo.



El estimador de la media a posteriori es $E[\theta|\mathbf{x}] \approx 4,09$ y la tasa de aceptación es aproximadamente 23 %.

El muestreo de independencia

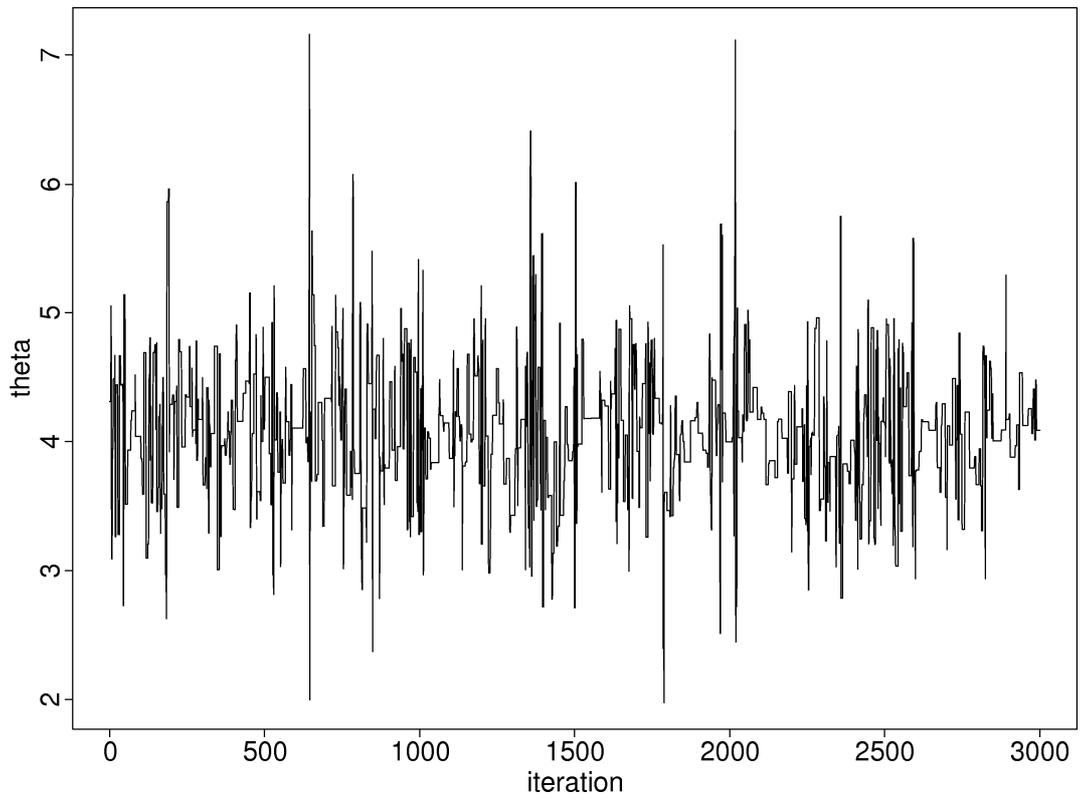
Supongamos que $g(\phi|\theta) = g(\phi)$ independiente de θ .

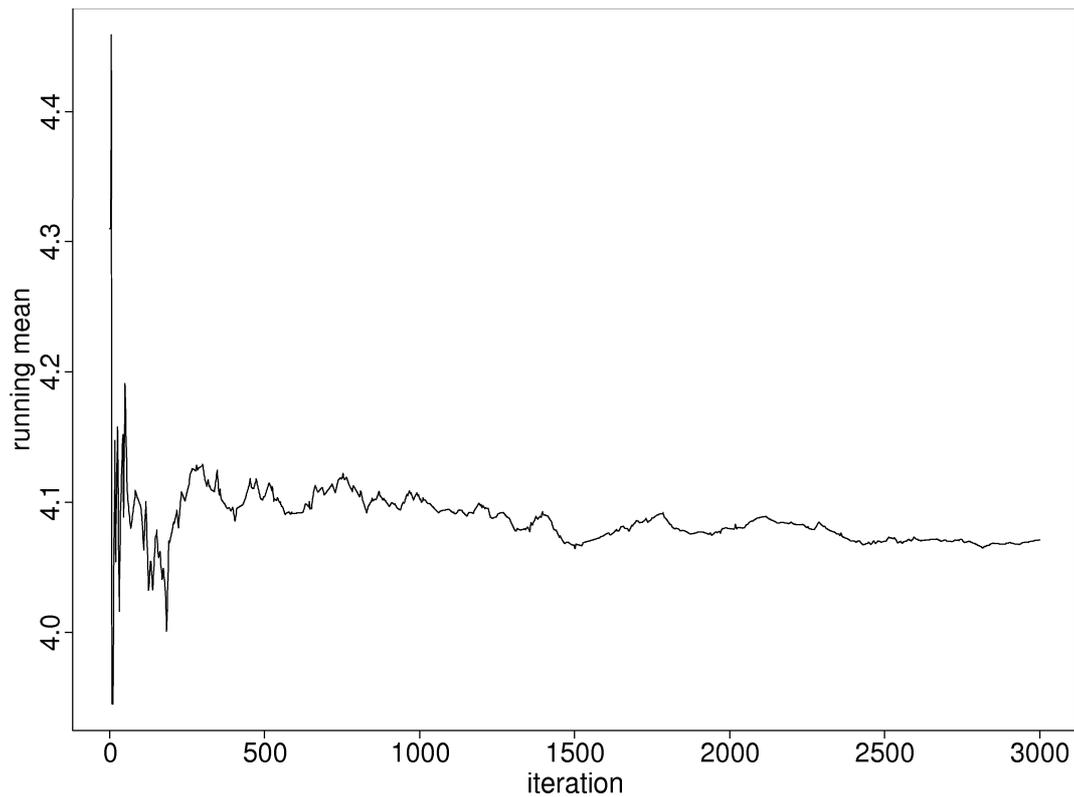
Ejemplo 81 Volviendo a Ejemplo 80, supongamos ahora que $g(\phi) = \mathcal{N}(\bar{x}, s^2)$.

Entonces, la probabilidad de aceptar ϕ dado $\theta^{(t)}$ será

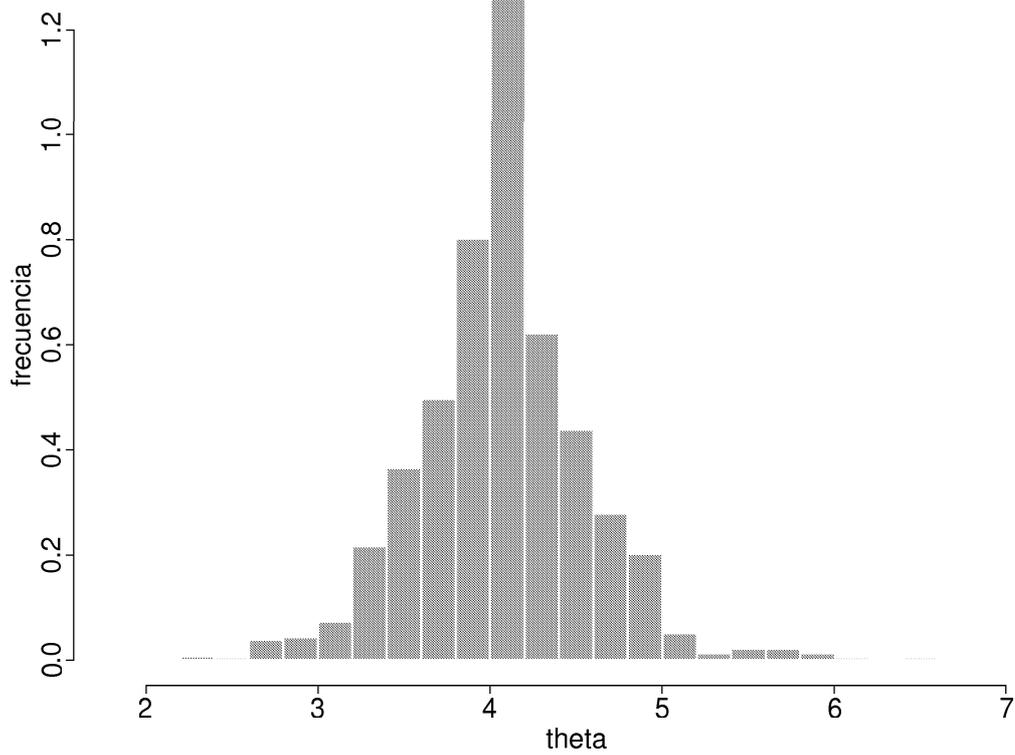
$$\alpha = e^{-\frac{1}{2s^2}[(\theta^{(t)} - \bar{x})^2 - (\phi - \bar{x})^2]} \prod_{i=1}^n \frac{1 + (x_i - \theta^{(t)})^2}{1 + (x_i - \phi)^2}$$

Empezando el algoritmo con los mismos datos que antes, podemos ilustrar los resultados.





Se ve que la convergencia es menos rápida que para el método Metropolis.



El histograma aparece similar al otro.

La estimación de la media a posteriori es $E[\theta|\mathbf{x}] \approx 4,06$ con tasa de aceptación de 24 %.

El muestreo de Gibbs

El algoritmo de Gibbs es una versión del método Metropolis Hastings con partición del espacio con

$$g_i(\phi|\theta_i, \theta_{-i}) = f(\phi|\theta_{-i}, \mathbf{x})$$

la distribución condicional a posteriori.

Dada esta distribución, se puede demostrar que la probabilidad de aceptar ϕ es 1; luego los valores propuestos son siempre aceptados.

El algoritmo para el muestreo de Gibbs es el siguiente:

1. Comenzar con valores iniciales arbitrarios $\theta^{(0)}$
2. Generar $\theta_1^{(t+1)} \sim f(\theta_1 | \mathbf{x}, \theta_2^{(t)}, \dots, \theta_k^{(t)})$,
3. Generar $\theta_2^{(t+1)} \sim f(\theta_2 | \mathbf{x}, \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)})$,
4. \vdots
5. Generar $\theta_k^{(t+1)} \sim f(\theta_k | \mathbf{x}, \theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)})$,
6. Ir a 2.

El algoritmo converge a una tasa geométrica en N , el tamaño de la muestra de Gibbs (usando la norma de supremo).

Un ejemplo fácil

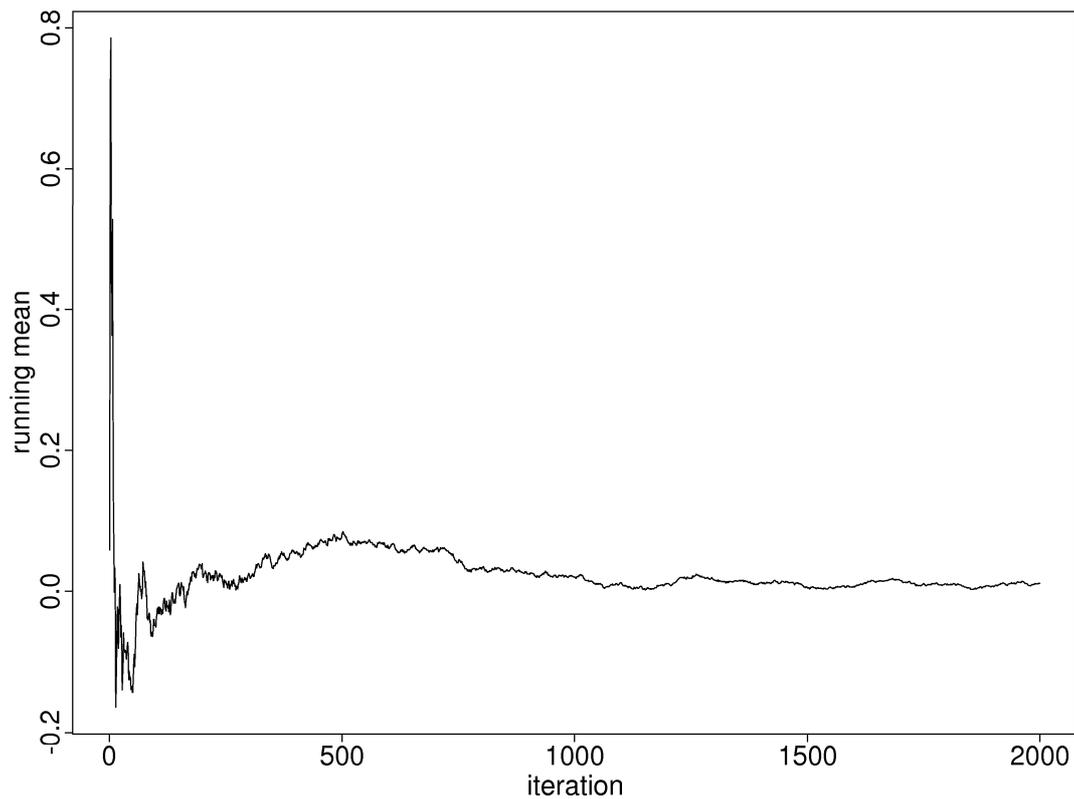
Ejemplo 82 *Volvamos al Ejemplo 78. También se puede utilizar una muestra de Gibbs para muestrear la distribución a posteriori.*

Las distribuciones condicionales en este caso son

$$\begin{aligned}\mu|\mathbf{x}, \phi &\sim \mathcal{N}\left(\bar{x}, \frac{1}{n\phi}\right) \\ \phi|\mathbf{x}, \mu &\sim \mathcal{G}\left(\frac{n-1}{2}, \frac{(n-1)s^2 + n(\mu - \bar{x})^2}{2}\right)\end{aligned}$$

y es muy fácil muestrear las dos.

El diagrama muestra el valor estimado de $E[\mu|\mathbf{x}]$ frente al número de iteraciones. Se necesitan aproximadamente 1000 iteraciones para la convergencia.



Observación 31 *Es menos eficaz utilizar el método de Gibbs cuando se puede utilizar un método Monte-Carlo simple.*

Ejemplo 83 Modelos con datos censurados.

Supongamos que Z_i es la vida de una máquina i en un estudio. Pongamos una densidad $f(Z_i|\boldsymbol{\theta})$. Habitualmente, la duración del test es corta y sólo observamos unas pocas vidas enteras y las otras observaciones están truncadas a un tiempo T .

Entonces, los datos observados son

$$X_i = \begin{cases} Z_i & \text{si } X_i < T \\ T & \text{si } X_i > T \end{cases}$$

y la forma de la verosimilitud es

$$l(\boldsymbol{\theta}|\mathbf{x}) \propto \prod_{i=1}^{n_1} f(z_i|\boldsymbol{\theta})(1 - F(T|\boldsymbol{\theta}))^{n_2}$$

donde se supone que se observan n_1 tiempos de vida enteros y n_2 tiempos truncados a T .

A menudo la forma de la verosimilitud es complicada y no se pueden utilizar métodos simples para hacer la inferencia.

Pero se pueden simular las variables Z_i en un algoritmo Gibbs. Dado X_i, θ tenemos $Z_i = X_i$ si $X_i < T$ y $f(Z_i|\theta, X_i = T) \propto f(Z_i|\theta)$ truncado a $[T, \infty)$ si $X_i = T$. Es bastante fácil simular la distribución truncada usando por ejemplo el método de rechazo.

Observación 32 *Para simplificar el problema se han introducido los tiempos de vida no observados. Estas variables son variables latentes.*

Un algoritmo de Gibbs para evaluar la distribución a posteriori de θ dados los datos x es la siguiente:

1. $t = 0$
2. Fijar valores iniciales $\theta^{(0)}$.
3. Generar $z_{n_1+1}^{(t+1)}, \dots, z_{n_1+n_2}^{(t+1)}$ de $f(z_i|x_i, \theta^{(t)})$
4. Generar $\theta^{(t+1)}$ de $f(\theta|\mathbf{x}, \mathbf{z}^{(t+1)}) = f(\theta|\mathbf{z}^{(t+1)})$.
5. $t = t + 1$, Ir a 3.

Ejemplo 84 Podemos ilustrar el algoritmo del Ejemplo previo suponiendo que $Z_i \sim \mathcal{E}(\theta)$. En este caso, se puede evaluar la verosimilitud exactamente.

Supongamos que $\mathbf{x} = (x_1, \dots, x_5, 1, 1)$ con 5 vidas no truncadas y 2 vidas truncadas a 1. Supongamos que la suma de las vidas no truncadas es $\sum_{i=1}^5 x_i = 3$.

Entonces, la verosimilitud será

$$\begin{aligned}l(\theta|\mathbf{x}) &\propto \theta^5 \exp\left(-\theta \sum_{i=1}^5 x_i\right) \exp(-\theta)^2 \\ &\propto \theta^5 \exp(-5\theta)\end{aligned}$$

y dada una distribución a priori de Jeffreys, $f(\theta) \propto \frac{1}{\theta}$, la distribución a posteriori es $\theta|\mathbf{x} \sim \mathcal{G}(5, 5)$.

Vamos a comparar los resultados con los del algoritmo de Gibbs. Tenemos

$$\begin{aligned}f(z|\theta, z > 1) &= \frac{\theta e^{-\theta z}}{e^{-\theta}} \\ &= \theta e^{-\theta(z-1)}\end{aligned}$$

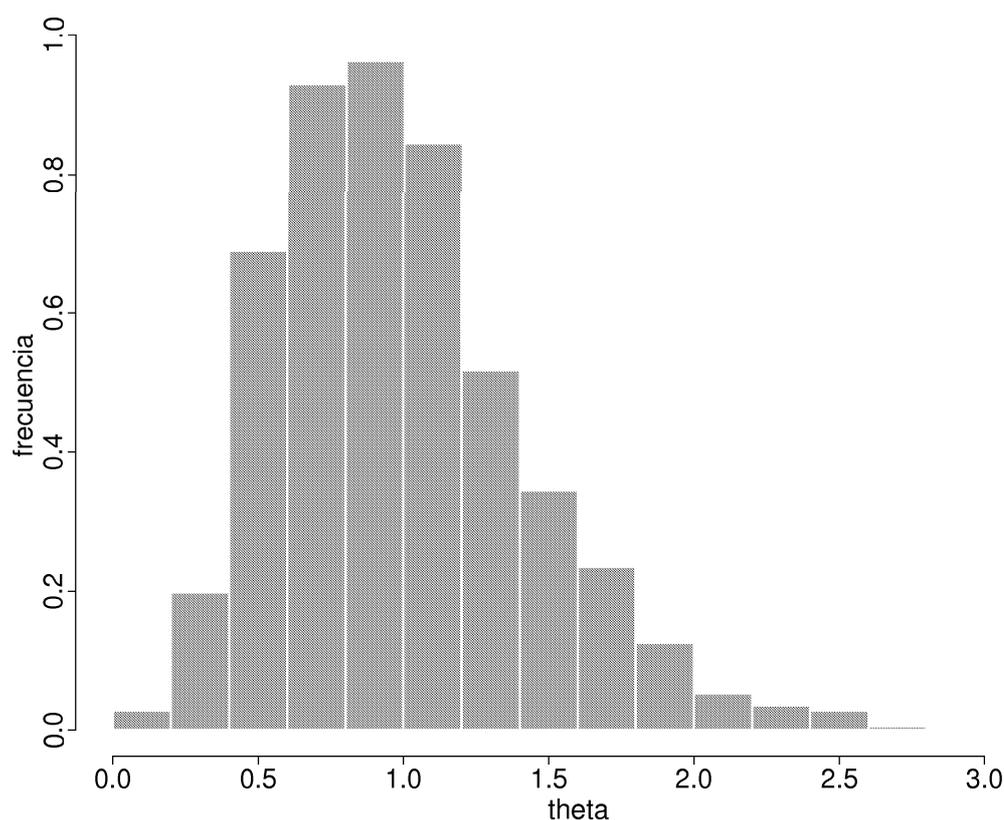
para $z > 1$ que implica que

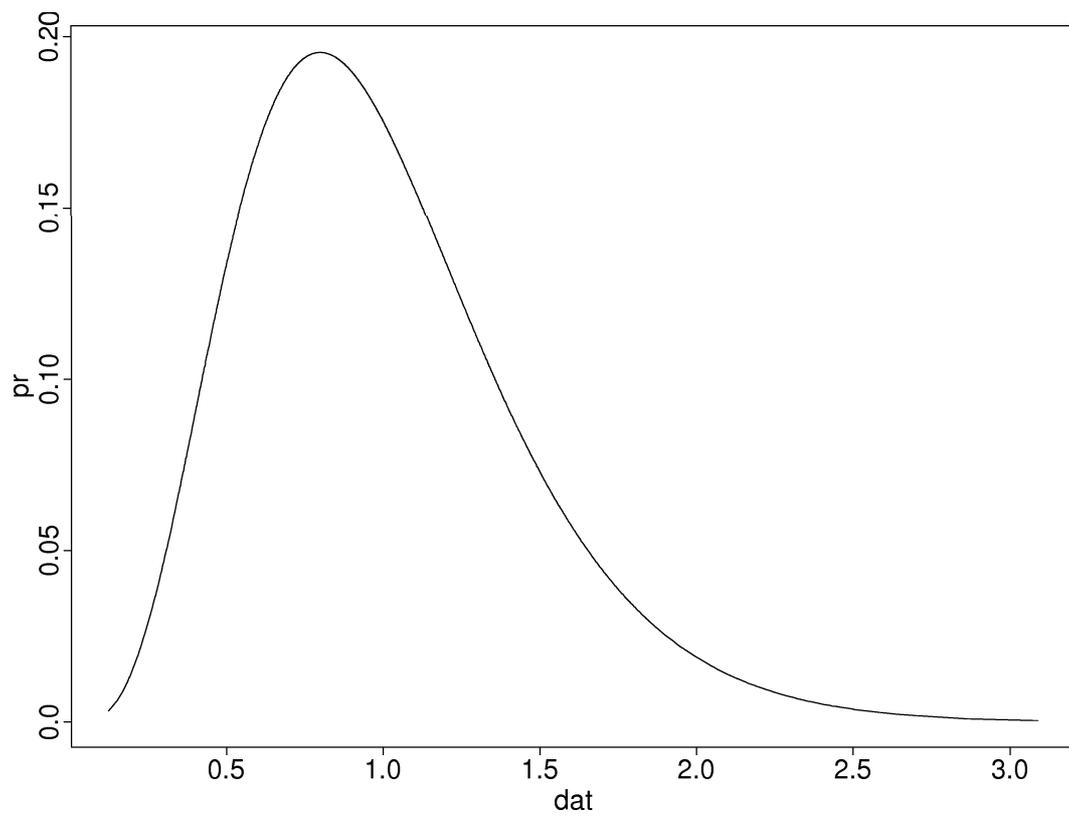
$$Z - 1|\theta, Z > 1 \sim \mathcal{E}(\theta)$$

y se puede muestrear esta distribución.

También tenemos $\theta|\mathbf{z} \sim \mathcal{G}(7, 3 + z_6 + z_7)$ que es fácil de muestrear.

Una muestra de Gibbs usando 1000 iteraciones en equilibrio generó los estimadores $E[\theta|\mathbf{x}] \approx 1,001$ y $V[\theta|\mathbf{x}] \approx ,198$. Los valores precisos son $E[\theta|\mathbf{x}] = 1$ y $V[\theta|\mathbf{x}] = 0,2$. El diagrama ilustra un histograma de los datos generados. Es parecido a la distribución $\mathcal{G}(5, 5)$.





Los últimos descubrimientos

Un problema de los algoritmos de MCMC es que es difícil decidir cuando ha convergido la cadena. Pero en algunos artículos recientes, se ha introducido el muestreo perfecto: una versión de MCMC para que se pueda estar seguro que se tiene una muestra de la distribución de equilibrio de la cadena.

El algoritmo tipo usa la idea de "coupling from the past" para determinar la llegada en equilibrio.

Ver la pagina del web
dimacs.rutgers.edu/~dbwilson/exact.html/