

EstComp-Tarea12

Bruno Gonzalez

21/11/2019

12-Metropolis

Regresamos al ejercicio de IQ de la tarea anterior, en ésta hiciste cálculos para el caso de una sola observación. En este ejercicio consideramos el caso en que observamos una muestra $x = \{x_1, \dots, x_N\}$, y utilizaremos Metrópolis para obtener una muestra de la distribución posterior.

- a) Crea una función *prior* que reciba los parámetros μ y τ que definen tus creencias del parámetro desconocido θ y devuelva $p(\theta)$, donde $p(\theta)$ tiene distribución $N(\mu, \sigma^2)$

```
prior <- function(mu, tau){  
  function(theta){  
    dnorm(theta, mu, tau)  
  }  
}
```

- b) Utiliza la función que acabas de escribir para definir una distribución inicial con parámetros $\mu = 150$ y $\tau = 15$, llámala *mi_prior*.

```
mi_prior <- prior(150, 15)
```

Ya que tenemos la distribución inicial debemos escribir la verosimilitud, en este caso la verosimilitud es:

$$\begin{aligned} p(x|\theta, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^N (x_j - \theta)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{j=1}^N x_j^2 - 2\theta \sum_{j=1}^N x_j + N\theta^2\right)\right) \end{aligned}$$

Recuerda que estamos suponiendo varianza conocida, supongamos que la desviación estándar es $\sigma = 20$.

$$p(x|\theta) = \frac{1}{(2\pi(20^2))^{N/2}} \exp\left(-\frac{1}{2(20^2)} \left(\sum_{j=1}^N x_j^2 - 2\theta \sum_{j=1}^N x_j + N\theta^2\right)\right)$$

- c) Crea una función *likeNorm* en R que reciba la desviación estándar, la suma de los valores observados $\sum x_i$, la suma de los valores al cuadrado $\sum x_i^2$ y el número de observaciones N la función devolverá la función de verosimilitud (es decir va a regresar una función que depende únicamente de θ).

```
# S: sum x_i, S2: sum x_i^2, N: número obs.  
likeNorm <- function(S, S2, N){  
  function(theta){  
    (1/(2*pi*20^2)^(N/2))*exp((-1/(2*20^2))*(S2-2*theta*S+N*theta^2))  
  }  
}
```

- d) Supongamos que aplicamos un test de IQ a 100 alumnos y observamos que la suma de los puntajes es 13300, es decir $\sum x_i = 13,000$ y $\sum x_i^2 = 1,700,000$. Utiliza la función que acabas de escribir para definir la función de verosimilitud condicional a los datos observados, llámala *mi_like*.

```
mi_like <- likeNorm(13000,1700000,100)
```

- e) La distribución posterior no normalizada es simplemente el producto de la inicial y la posterior:

```
postRelProb <- function(theta){  
  mi_like(theta) * mi_prior(theta)  
}
```

Utiliza Metropolis para obtener una muestra de valores representativos de la distribución posterior de θ . Para proponer los saltos utiliza una Normal(0, 5).

Primero se define la función para la caminata aleatoria

```
caminata <- function(theta){  
  salto_prop <- rnorm(1,0,5)  
  theta_prop <- theta + salto_prop  
  
  u <- runif(1)  
  
  p_move <- min(postRelProb(theta_prop)/postRelProb(theta),1)  
  if(p_move > u){  
    return(theta_prop)  
  }  
  else{  
    return(theta)  
  }  
}
```

Ahora generamos la caminata

```
pasos <- 5000  
camino <- numeric(pasos)  
camino[1] <- 100  
  
for (i in 2:pasos){  
  camino[i] <- caminata(camino[i-1])  
}
```

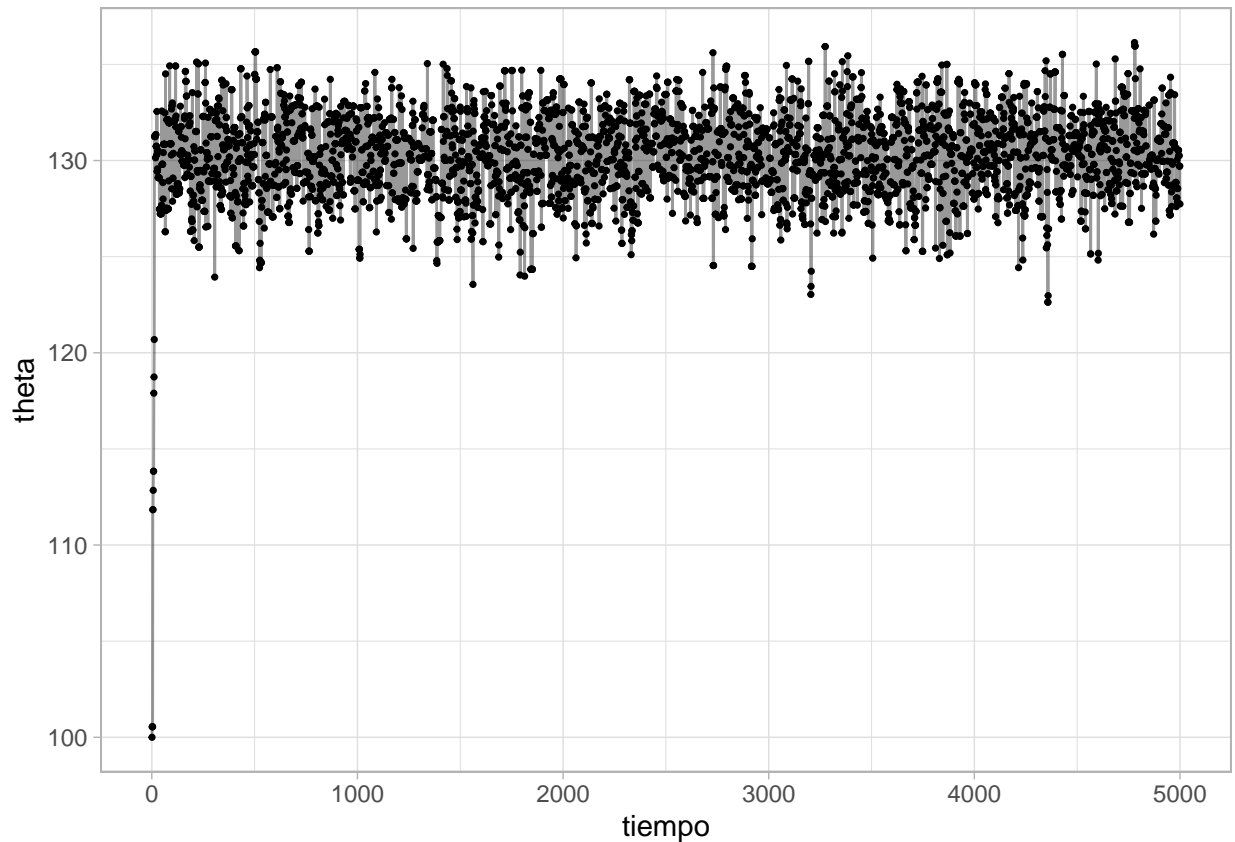
- f) Grafica los valores de la cadena para cada paso.

Primero construimos el data frame para usar *ggplot2*.

```
caminata <- data.frame(pasos = 1:pasos, theta = camino)
```

Ahora podemos construir la gráfica.

```
ggplot(caminata, aes(x=pasos, y = theta)) +
  geom_point(size = 0.6) +
  geom_path(alpha = 0.4) +
  xlab("tiempo") +
  theme_light()
```



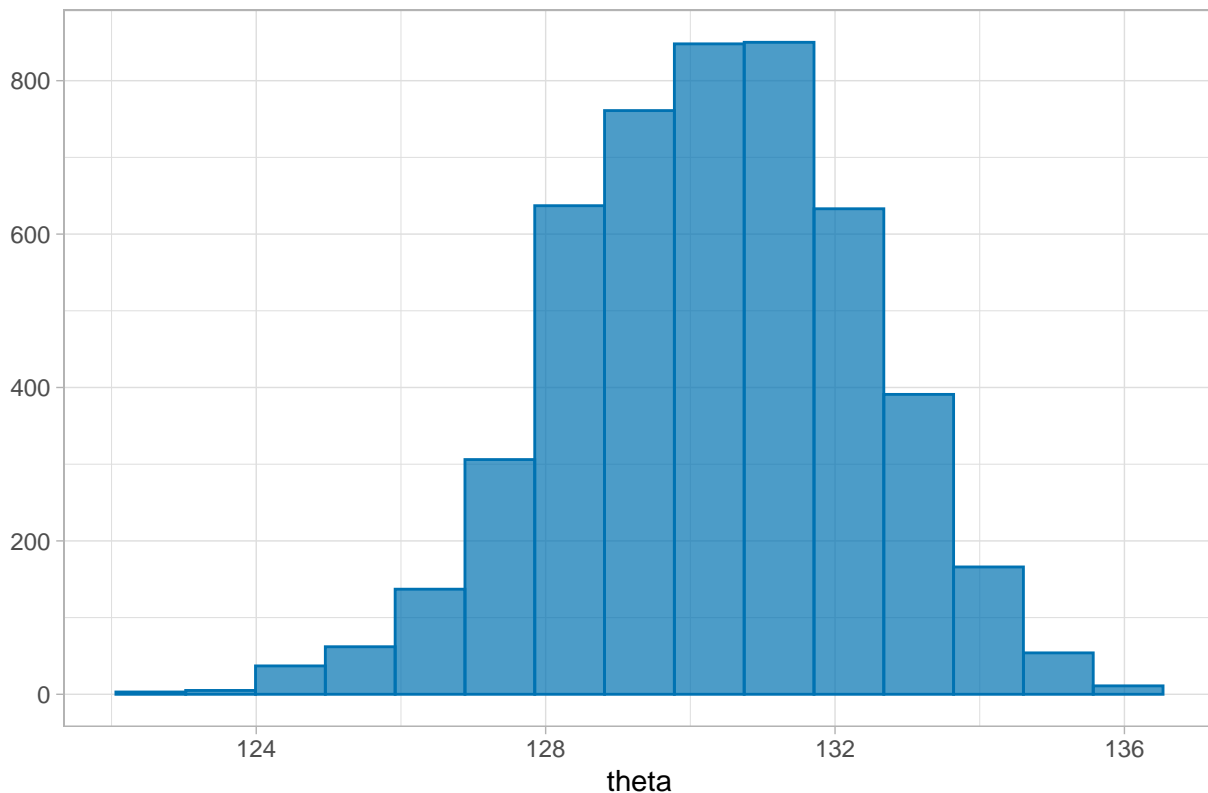
g) Elimina los valores correspondientes a la etapa de calentamiento y realiza un histograma de la distribución posterior.

En este caso la caminata converge con bastante validez, por lo que solo se eliminarán los primeros 100 valores.

```
h1 <- ggplot(caminata[100:5000,], aes(x=theta)) +
  geom_histogram(bins = 15, alpha=0.7, fill = "#0072B2", color = "#0072B2") +
  labs(title = 'Simulación Metropolis', y = element_blank()) +
  theme_light()
```

h1

Simulación Metropolis



h) Si calcularas la posterior de manera analítica obtendrías que $p(x|\theta)$ es normal con media:

$$\frac{\sigma^2}{\sigma^2 + N\tau^2}\mu + \frac{N\tau^2}{\sigma^2 + N\tau^2}\bar{x}$$

y varianza

$$\frac{\sigma^2\tau^2}{\sigma^2 + N\tau^2}$$

donde $\bar{x} = 1/N \sum_{i=1}^N x_i$ es la media de los valores observados. Realiza simulaciones de la distribución posterior calculada de manera analítica y comparalas con el histograma de los valores generados con Metropolis.

Primero definimos los parámetros

```
m_s <- (20^2/(20^2+100*15^2))*150 + ((100*15^2)/(20^2+100*15^2))*130
v_s <- ((20^2)*(15^2))/(20^2+100*15^2)
```

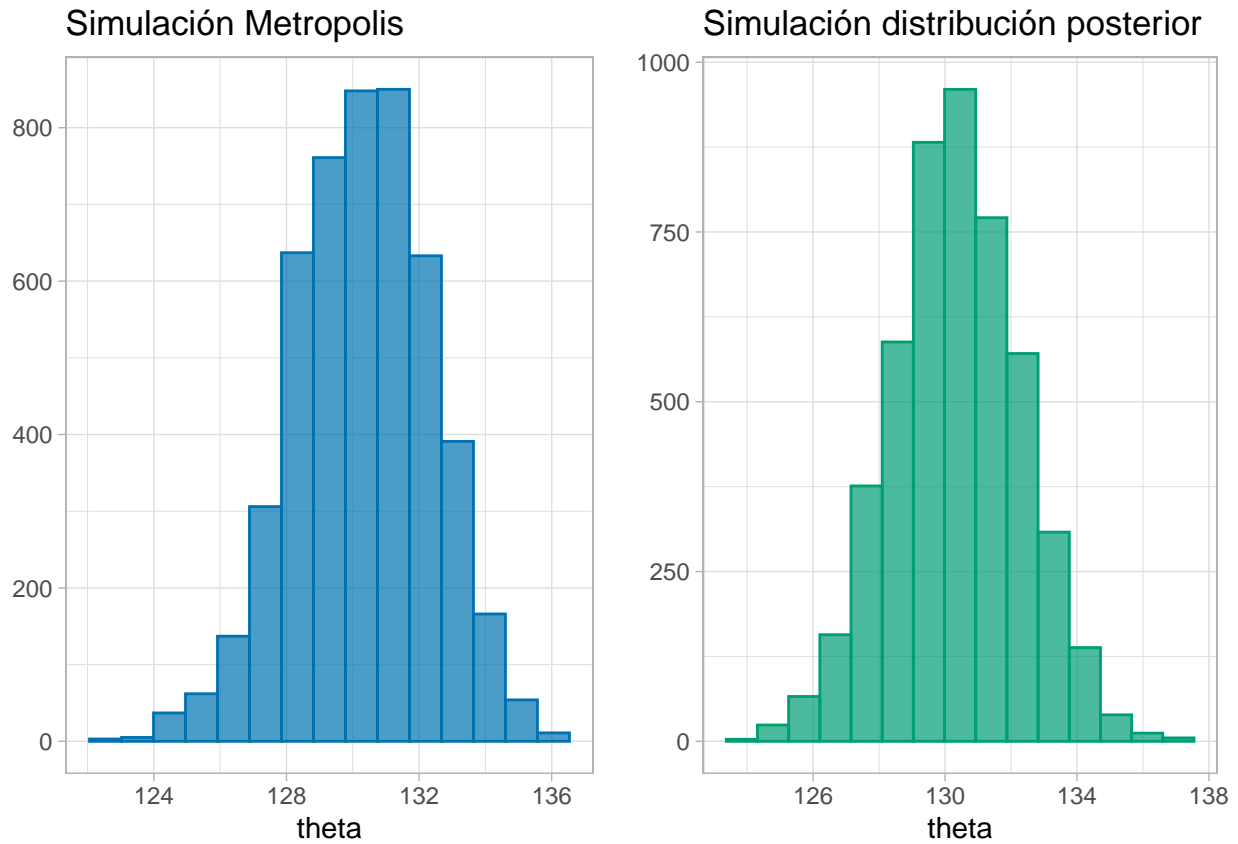
Usando estos parámetros podemos hacer la simulación

```
simulacion <- rnorm(4900, m_s, sqrt(v_s))
simulacion <- data.frame(tiempo = 1:4900, theta = simulacion)
```

Ahora podemos comparar los dos histogramas

```
h2 <- ggplot(simulacion, aes(x=theta)) +
  geom_histogram(bins = 15, alpha=0.7, fill = "#009E73", color = "#009E73") +
  labs(title = 'Simulación distribución posterior', y = element_blank()) +
  theme_light()

grid.arrange(h1,h2,ncol=2)
```



En ambos casos la distribución es muy parecida.

- i) ¿Cómo utilizarías los parámetros μ, τ^2 para describir un escenario donde sabes poco del verdadero valor de la media θ ?

Podemos realizar una serie de experimentos para obtener datos y a partir de ellos usar algún metodo de análisis bayesiano, como Metrópolis, para aproximar una distribución posterior.